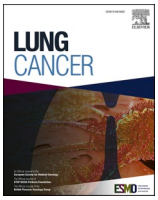
Lung Cancer 205 (2025) 108613

Contents lists available at ScienceDirect

Lung Cancer

journal homepage: www.elsevier.com/locate/lungcan



Development and clinical validation of a prognostic algorithm for stroma-tumor ratio quantification in non-small cell lung cancer

Waleed K.M. Ahmad [a,e], Tillmann Bedau [a], Yuan Wang [a], Sebastian Michels [b], Anna Rasokat [b], Jürgen Wolf [b], Matthias Heldwein [c], Simon Schallenberg [d], Alexander Quaas [a], Reinhard Büttner [a], Yuri Tolkach [a,*]

[a] Institute of Pathology, University Hospital Cologne, Cologne, Germany
[b] Clinic I of Internal Medicine, Center for Integrated Oncology, Germany
[c] Department of Cardiothoracic Surgery, University Hospital Cologne, Germany
[d] Institute of Pathology, Charité University Hospital, Berlin, Germany
[e] Medical Faculty, University of Cologne, Cologne, Germany

ARTICLE INFO

Keywords:
Stroma-tumor ratio (STR)
non-small cell lung cancer (NSCLC)
Segmentation algorithm
Digital pathology

ABSTRACT

Background and Aim: Lung cancer is the leading cause of cancer-related mortality worldwide, highlighting the importance of refining diagnostic modalities. This study's main focus is the development of a digital pathology, prognostic algorithm for fully automatized quantification of stroma-tumor ratio (STR) in patients with resectable non-small cell lung cancer (NSCLC).

Materials and Methods: The developed STR algorithm is built upon a powerful multi-class tissue segmentation algorithm that generates precise maps of the full tumor region. One retrospective exploration cohort of NSCLC patients (n = 902) and three validation cohorts (n = 784) of patients with lung adenocarcinoma (LUAD) and squamous cell carcinoma (LUSC) were included to identify and validate optimal prognostic cut-offs and different risk stratification methods with regard to different clinical endpoints: overall survival (OS), cancer-specific survival (CSS) and progression-free survival (PFS).

Results: For LUAD, we show that the minimal STR value for the whole case is decisive for prognostic evaluation. Different approaches (single STR cut-off, multiple STR cut-offs, using STR as a continuous parameter) allow for robust stratification of patients into prognostic risk groups, independent of the classical clinicopathological variables and conventional histological grading. For LUSC, STR may assist in identifying a small subset of patients with unfavorable prognosis (based on the maximum STR for the whole case), however, its prognostic impact varies between cohorts.

Conclusion: STR quantification in LUAD NSCLC subtype shows a promising role as a prognostic biomarker. It can be easily implemented in routine diagnostics and could be considered as a component of advanced prognostic systems in LUAD. Our results in LUSC cohorts suggest that STR quantification in its current implementation is of limited value in this subtype.

1. Introduction

Lung cancer remains the most significant cause of mortality from cancer worldwide [1]. NSCLC, the most common subtype of lung cancer, represents approximately 81 % of these cases [2]. There is still a critical need for the development of improved diagnostic, prognostic, and predictive instruments to enhance NSCLC patient survival. Computational pathology methods and AI algorithms allow for extraction of clinically relevant features directly from hematoxylin and eosin (H&E)-stained histological slides without the necessity of further costly or laborious analysis, however, their routine clinical implementation remains to be established [3,4].

STR is a promising candidate for a universal biomarker in epithelial malignant tumors [5]. It is especially well-established for resectable colorectal cancer [5,6]. The STR evaluation was initially developed for pathologists using classical microscopy methodology [5]. Pathologists

* Corresponding author.
E-mail address: yuri.tolkach@gmail.com (Y. Tolkach).

https://doi.org/10.1016/j.lungcan.2025.108613
Received 31 August 2024; Received in revised form 9 May 2025; Accepted 31 May 2025
Available online 1 June 2025

are expected to identify the most relevant part of the tumor based on visual estimation. Naturally, selecting relevant regions but also STR quantification itself are very labor intensive and prone to interobserver variability [5]. Few studies implemented the classical or slightly modified approach in patients with resectable NSCLC [7–10], with STR quantifications carried out by pathologists. These single studies show the feasibility and prognostic value of STR quantification in both LUAD and LUSC patients. Several studies showed that AI-based image analysis on H&E-stained histological sections can be implemented to automate and objectivize STR analysis, e.g., in colorectal cancer [11]. For NSCLC, two studies implemented image analysis for STR quantification [12,13], however, both utilizing pan-cytokeratin staining instead of H&E-stained sections.

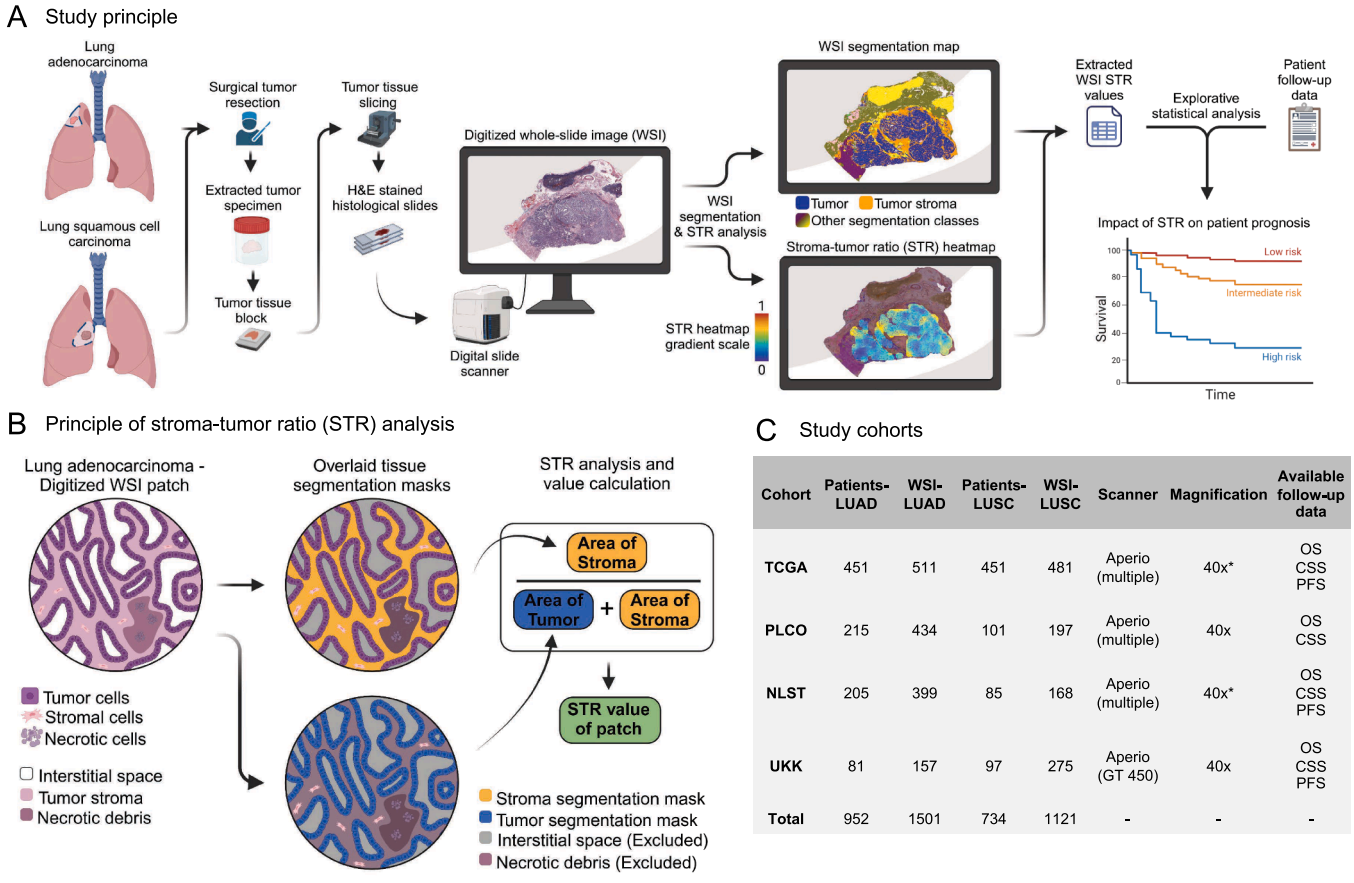
In this study, we develop a powerful, fully automatized, AI-based pipeline for systematic STR quantification in resectable NSCLC cases (Fig. 1A,B). We refine the methodology of STR analysis in lung cancer including the optimal region-of-interest size, analysis approach and risk stratification thresholds. Using four large, well-characterized, multi-institutional, international patient cohorts, we demonstrate the independent prognostic value for STR in NSCLC patients for multiple clinical endpoints. The developed pipeline might be a valuable instrument for patient risk stratification in clinical routine and within clinical trials.

2. Materials and methods

2.1. Patient cohorts and digitization

Four cohorts of resectable NSCLC patient cases containing LUAD and LUSC cases were included (Fig. 1C, detailed clinicopathological characteristics in Table 1). One cohort (The Cancer Genome Atlas, TCGA) was primarily utilized for the exploration phase. Three other cohorts: Prostate, Lung, Colorectal and Ovarian cancer screen trial (PLCO) cohort, National Lung Screen Trial (NLST) cohort, and NSCLC cohort of the University Hospital Cologne (UKK) were used for validation purposes. Inclusion criteria were presence of digitized whole-slide images of tumor region and available follow-up data. Follow-up data was available for multiple prognostic endpoints (Fig. 1C).

TCGA, PLCO, and NLST were digitized by Leica Aperio (Leica Biosystems, Wetzlar, Germany) scanner family, mostly model AT2 or similar (but not GT450DX). The majority of the slides were scanned at 40x magnification, with a limited number of cases from the TCGA and NLST cohorts scanned at 20x magnification. The UKK cohort was digitized by Leica Aperio GT450 scanning system and SVSUtil software [14] (https://github.com/ghsmith/SVSUtil) was used to apply the native scanner ICC profile of UKK cases to all images after scanning.



Cohort	Patients-LUAD	WSI-LUAD	Patients-LUSC	WSI-LUSC	Scanner	Magnification	Available follow-up data
TCGA	451	511	451	481	Aperio (multiple)	40x*	OS CSS PFS
PLCO	215	434	101	197	Aperio (multiple)	40x	OS CSS
NLST	205	399	85	168	Aperio (multiple)	40x*	OS CSS PFS
UKK	81	157	97	275	Aperio (GT 450)	40x	OS CSS PFS
Total	952	1501	734	1121	-	-	-

Fig. 1. Overview of the automatic stroma-tumor ratio (STR) quantification pipeline, principle of STR analysis, and study cohorts. A: Study principle. Surgically extracted resections of lung adenocarcinoma and lung squamous cell carcinoma cases are processed into histological slides which are then digitized into whole-slide images (WSIs). These WSIs are then further analyzed by two algorithms (multi-class tissue segmentation algorithm for lung specimens and the STR quantification algorithm). STR analysis is performed on the regional level to identify the most aggressive part of the tumor. These measurements are then used for prognostic analysis and patient stratification according to risk. **B:** Principle of STR quantification. The STR values for single regions range between 0 (stroma-low) and 1 (stroma-high). **C:** Characteristics of study cohorts, aspect of digitization, and available follow-up information. *A limited number of cases from the TCGA and NLST cohorts were scanned at 20x magnification.

Table 1
Clinico-pathological characteristics of study cohorts.

Parameter	TCGA Cohort (n = 902)		PLCO Cohort (n = 316)		NLST Cohort (n = 290)		UKK Cohort (n = 178)	
	LUAD (%)	LUSC (%)	LUAD (%)	LUSC (%)	LUAD (%)	LUSC (%)	LUAD (%)	LUSC (%)
Patients	451	451	215	101	205	85	81	97
Whole slide images	511	481	434	197	399	168	157	275
Average number of WSIs per case (range)	1 (1–10)	1 (1–9)	2 (1–3)	2 (1–3)	2 (1–4)	2 (1–5)	2 (1–3)	2 (1–3)
Sex								
Female	244 (54.10 %)	114 (25.28 %)	112 (52.09 %)	29 (28.71 %)	98 (47.80 %)	20 (23.53 %)	41 (50.62 %)	16 (16.49 %)
Male	207 (45.90 %)	337 (74.72 %)	103 (47.91 %)	72 (71.29 %)	107 (52.20 %)	65 (76.47 %)	40 (49.38 %)	81 (83.51 %)
Age								
Minimum	33	39	52	55	55	56	44	52
Maximum	88	90	74	74	74	74	87	90
Mean	65.19	67.37	63.55	64.33	63.84	64.62	65.46	69
Median	66	68	64	64	64	65	66	71
pT stage								
pT1	156 (34.59 %)	99 (21.94 %)	111 (51.62 %)	36 (35.64 %)	145 (70.73 %)	51 (60.00 %)	39 (48.15 %)	28 (28.86 %)
pT2	236 (52.32 %)	269 (59.65 %)	86 (40 %)	52 (51.49 %)	40 (19.51 %)	26 (30.59 %)	17 (20.99 %)	30 (30.93 %)
pT3	40 (8.87 %)	62 (13.75 %)	5 (2.33 %)	7 (6.93 %)	13 (6.34 %)	8 (9.41 %)	12 (14.81 %)	27 (27.84 %)
pT4	16 (3.55 %)	21 (4.66 %)	13 (6.05 %)	6 (5.94 %)	6 (2.93 %)	0	13 (16.05 %)	12 (12.37 %)
Unavailable	3 (0.67 %)	0 (0)	0 (0)	0	1 (0.49 %)	0	0	0
pN stage								
pN0	295 (65.41 %)	285 (63.19 %)	169 (78.60 %)	73 (72.28 %)	147 (0.7171)	67 (78.82 %)	48 (59.26 %)	62 (63.92 %)
pN1	88 (19.51 %)	121 (26.83 %)	23 (10.70 %)	16 (15.84 %)	25 (12.20 %)	11 (12.94 %)	15 (18.52 %)	21 (21.65 %)
pN2-pN3	57 (12.64 %)	41 (9.09 %)	22 (10.23 %)	12 (11.88 %)	18 (0.0877)	2 (2.35 %)	18 (22.22 %)	11 (11.34 %)
Unavailable	11 (2.44 %)	4 (0.89 %)	1 (0.47 %)	0	15 (7.32 %)	5 (5.89 %)	0	3 (3.09 %)
M status								
M0	303 (67.18 %)	376 (83.37 %)	193 (89.77 %)	92 (91.09 %)	191 (93.17 %)	82 (96.47 %)	76 (93.83 %)	96 (98.97 %)
M1	23 (5.10 %)	6 (1.33 %)	16 (7.44 %)	7 (6.93 %)	8 (3.90 %)	1 (1.18 %)	5 (6.17 %)	1 (1.03 %)
Unavailable	125 (27.72 %)	69 (15.30 %)	6 (2.79 %)	2 (1.98 %)	6 (2.93 %)	2 (2.35 %)	0	0
UICC stage								
I	247 (54.77 %)	219 (48.56 %)	147 (68.38 %)	65 (64.36 %)	132 (64.39 %)	63 (74.11 %)	28 (34.57 %)	36 (37.11 %)
II	110 (24.39 %)	146 (32.37 %)	22 (10.23 %)	18 (17.82 %)	24 (11.71 %)	11 (12.94 %)	21 (25.93 %)	33 (34.02 %)
III	63 (13.97 %)	76 (16.85 %)	30 (13.95 %)	11 (10.89 %)	24 (11.71 %)	4 (4.71 %)	27 (33.33 %)	27 (27.84 %)
IV	24 (5.32 %)	6 (1.33 %)	16 (7.44 %)	7 (6.93 %)	8 (3.90 %)	0	5 (6.17 %)	1 (1.03 %)
Unavailable	7 (1.55 %)	4 (0.89 %)	0 (0)	0	17 (8.29 %)	7 (8.24 %)	0	0
OS								
Alive	284 (62.97 %)	255 (56.54 %)	52 (24.19 %)	21 (20.79 %)	95 (46.34 %)	35 (41.18 %)	41 (50.62 %)	60 (61.86 %)
Dead	155 (34.37 %)	196 (43.46 %)	163 (75.81 %)	80 (79.21 %)	110 (53.66 %)	50 (58.82 %)	40 (49.38 %)	37 (38.14 %)
Unavailable	12 (2.66 %)	0	0	0	0	0	0	0
CSS								
Non-cancer deaths/Alive	310 (68.74 %)	310 (68.74 %)	110 (51.16 %)	52 (51.49 %)	36 (17.56 %)	20 (23.53 %)	61 (75.31 %)	79 (81.44 %)
Cancer death	88 (19.51 %)	99 (21.95 %)	105 (48.84 %)	49 (48.51 %)	74 (36.10 %)	30 (35.29 %)	20 (24.69 %)	18 (18.56 %)
Unavailable	53 (11.75 %)	42 (9.31 %)	0	0	95 (46.34 %)	35 (41.18 %)	0	0
PFS								
No progression	250 (55.43 %)	266 (58.98 %)	–	–	130 (63.41 %)	49 (57.65 %)	49 (60.49 %)	68 (70.11 %)
Progression	115 (25.50 %)	108 (23.95 %)	–	–	52 (25.37 %)	20 (23.53 %)	20 (24.69 %)	24 (24.74 %)
Unavailable	86 (19.07 %)	77 (17.07 %)	–	–	23 (11.22 %)	16 (18.82 %)	12 (14.82 %)	5 (5.15 %)
Follow-up duration								
Range in months	1–242	1–177	1–278	1–229	1–159	1–159	2–108	1–108
Mean	31.45	34.24	92.68	77.64	94.73	88.27	40.1	33.33
Smoke pack years								
Range in months	0.15–154	1–240	1–230	0–172	32–224	30–221.4	7–80	2–104
Mean	41.88	52.99	47.25	63.27	64.02	75	46.53	41.8

2.2. STR evaluation pipeline

The central part of the analytical pipeline (Fig. 1A) is a previously developed and independently, extensively clinically validated deep learning-based multi-tissue segmentation algorithm for H&E-stained NSCLC WSIs [15]). Shortly, this model is a UNet++ architecture-based decoder with an EfficientNetB0 encoder completely retrained on a large, high-quality, pixel-level annotated dataset. The algorithm analyzes WSIs in patches of 512 x 512 pixels (10x magnification) and builds a precise segmentation map of the WSI and of the tumor region particularly. The algorithm achieved a final average Dice-score (pixel-wise segmentation accuracy) of 0.893 across all segmented tissue types which is an outstanding segmentation quality. The algorithm is implemented in PyTorch 1.10 and Python 3.6.15 and requires a typical consumer-grade GPU card for implementation.

Within tumor regions, the algorithm detects all tissue classes (epithelial tumor component, tumor stroma, necrosis, mucin, tertiary lymphoid structures) that are relevant for STR assessment. The generated multi-class tissue segmentation mask is post-processed to prepare it for STR analysis (tumor stroma without adjacent tumor tissue is removed, all classes other than tumor and tumor stroma are masked).

The STR evaluation algorithm systematically processes the tumor region using small circular regions-of-interest (ROI) of a pre-defined size. This is a principle adopted from classical STR evaluation using a microscope [7,8]. Several eligibility tests are performed on each ROI to decide if the STR measurement should be performed. First, tumoral tissue should be present in all four quadrants of the ROI. Second, necrotic tissue and mucin content should be less than 20 % and 30 %, respectively. These thresholds were defined as optimal based on experimental testing which evaluated how patient risk group classification changed when single parameters were adjusted. The results/heatmaps were reviewed by experienced pathologists in accordance with classical guidelines (Suppl. Fig. 6). Third, the tumor and all tumor-associated classes (stroma, necrosis, mucin) must represent > 95 % of the ROI content to exclude all other non-relevant classes. During analysis, there is an overlap between adjacent ROIs of 70 % on each side. This ensures that all eligible ROIs are included, which depends on the position of each ROI frame and local tissue constellation.

In each eligible ROI, STR is quantified as Area Stroma/(Area Tumor + Area Stroma) (Fig. 1B). While this manuscript uses the term "Stroma-tumor ratio (STR)" for mathematical clarity, it is equivalent to the traditionally used "Tumor-stroma ratio (TSR)" in other literature, ensuring consistency in comparisons. Single measurements are recorded and aggregated if several WSIs are available for a case (Fig. 3A,B), and final metrics (minimal, maximal STR value, alongside the mean of median STR values for WSIs/case) are provided at case level.

2.3. Analysis visualization

Visualization is provided for each WSI to assist pathologists in double-checking the correctness, presented in the form of a heatmap allowing for a fully explainable evaluation process (Fig. 2A,B).

2.4. Statistics

All tests for statistical analyses were performed using R version 4.4.0 (R Foundation for Statistical Computing, Vienna, Austria). Pearson's correlation analysis was utilized to investigate associations between STR and common clinicopathological variables. For the prognostic significance assessment of STR, we utilized Kaplan-Meier survival estimates with log-rank tests. Univariate and multivariate Cox proportional hazards regression analysis was carried out to investigate the prognostic significance of STR and its independent significance in relation to the typical clinicopathological variables. P values lower than 0.05 were considered significant.

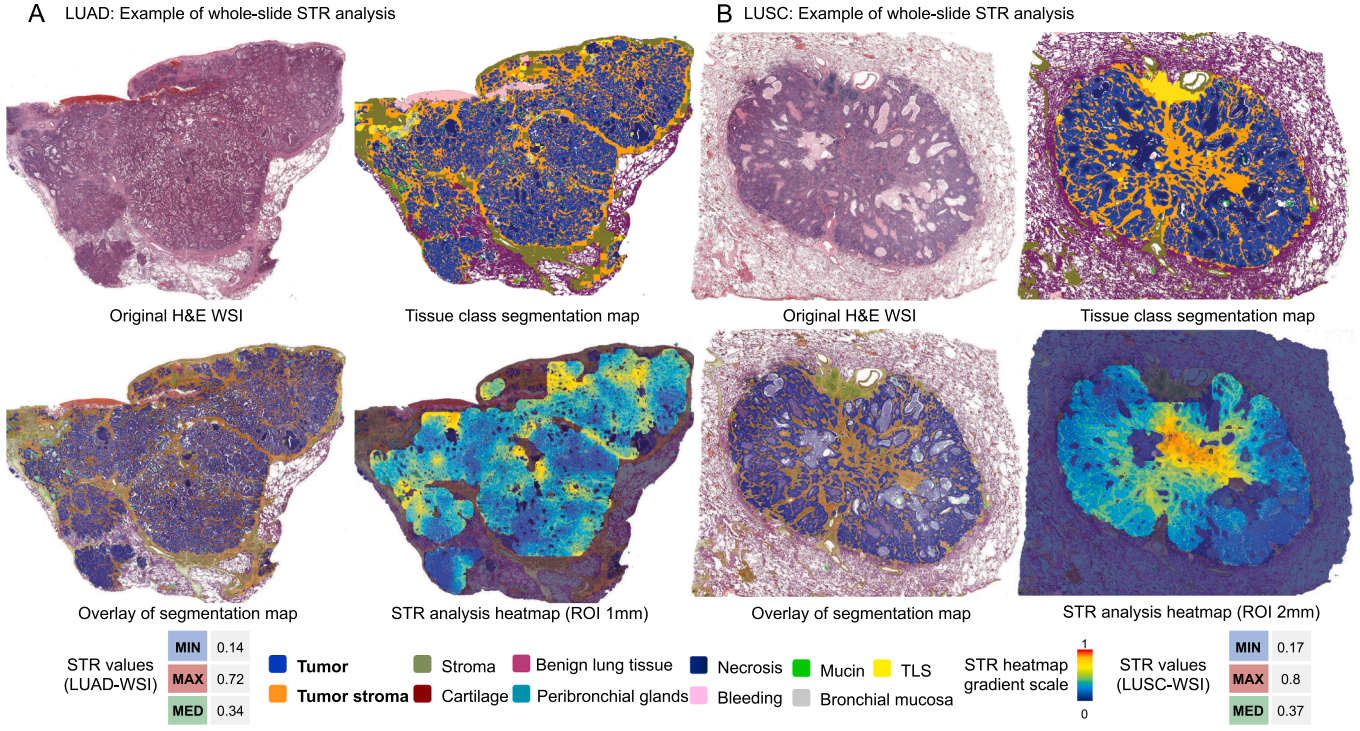


Fig. 2. Examples of whole-slide image analysis by developed STR pipeline for (A) Lung adenocarcinoma case and **(B)** lung squamous cell carcinoma case. Original H&E stained whole-slide image (WSI), multi-class tissue segmentation map of the WSI, overlay of the segmentation map on the original WSI, and a heatmap of the calculated STR values for visual analysis are shown. Legend outlines the colors corresponding to each segmented tissue class as well as a heatmap for STR values. Derived STR values (MIN: minimal, MED: median, and MAX: maximal for WSI). Abbreviations: WSI – whole-slide image, STR – stroma-tumor ratio, ROI – region-of-interest, TLS – tertiary lymphoid structures, LUAD – lung adenocarcinoma, LUSC – lung squamous cell carcinoma.

2.5. Ethical considerations

All study steps were performed in accordance with the Declaration of Helsinki. This study was approved by the Ethical committee of the University of Cologne (20–1583). Given the retrospective/archive nature of used data, the necessity of obtaining patient permission was waived by the ethical committee. Permissions for using PLCO and NLST cohorts were applied for and approved via https://cdas.cancer.gov/.

3. Results

3.1. STR quantification pipeline

First, we establish a fully automated STR quantification pipeline from H&E-stained WSIs containing tumors in patients with NSCLC (Fig. 1A,B). This pipeline includes two algorithms: 1) a powerful, deep-learning-based algorithm for multi-class (n = 11) tissue segmentation in WSIs developed earlier and extensively validated using numerous independent cohorts [15] (for details see Methods), 2) a statistical algorithm for the systematic analysis of STR in ROIs of a predefined size using segmentation masks of tumor regions from the first step. The principle and representative examples are provided in Fig. 2A,B, Suppl. Fig. 8–9. The STR algorithm implements rigorous quality control to assess the eligibility of single ROIs (tissue content) and aggregates the results from several WSIs, if available, in single cases (Fig. 3A,B). The main metrics produced by the pipeline are the lowest STR value for a case (MIN-STR), the highest STR value (MAX-STR), and a median value (MED-STR) of all STR measurements per case. These metrics are further evaluated for their prognostic significance in four large patient cohorts with resectable NSCLC (Table 1).

3.2. Establishing optimal STR parameters for analysis in LUAD and LUSC

The exploration cohort was used to determine an optimal approach for STR-based extraction of prognostic information. First, an optimal ROI size has never been experimentally determined for NSCLC. The typical ROI diameter, approx. 2 mm, is known from classical methodology used by pathologists with a microscope. We performed an analysis using different ROI sizes (1.0 mm, 1.5 mm, 2.0 mm; Suppl. Fig. 1) and evaluated the changes in the resulting metrics per case with different ROI sizes. We found that ROI diameters of 1.0 mm for LUAD and 2.0 mm for LUSC provided the best prognostic stratification of patients in the exploration cohort among the three different sizes tested (s. Suppl. Fig. 4 for details).

Second, we determine that MIN-STR primarily correlates with survival in LUAD (lower MIN-STR levels = unfavorable prognosis), while MAX-STR analysis is the best approach for LUSC (higher MAX-STR levels = unfavorable prognosis) and implement them in all further analyses (Suppl. Fig. 2 and 3 for details to disqualified evaluation approaches). The bidirectionality of STR in different tumor entities is known [9,10,16] (see Discussion). The distributions of single MIN-STR, MAX-STR, and MED-STR values for all cohorts are shown in Fig. 4A,B. In general, STR parameters showed similar distribution in cohorts (Suppl. Table 1), however, for both LUAD and LUSC, MAX-STR values were slightly lower in TCGA cohort. This might correspond to a selection bias of the TCGA slides towards tumor regions with a higher tumor cell content (necessary for sequencing).

In LUAD, the NLST cohort stood out with significantly higher MIN values compared to all other cohorts, and with higher MED values when compared with TCGA and UKK ($p < 0.05$ for all). The TCGA cohort demonstrated MAX values that were significantly lower than all other cohorts ($p < 0.001$).

For LUSC, the TCGA cohort had significantly lower MIN, MAX and

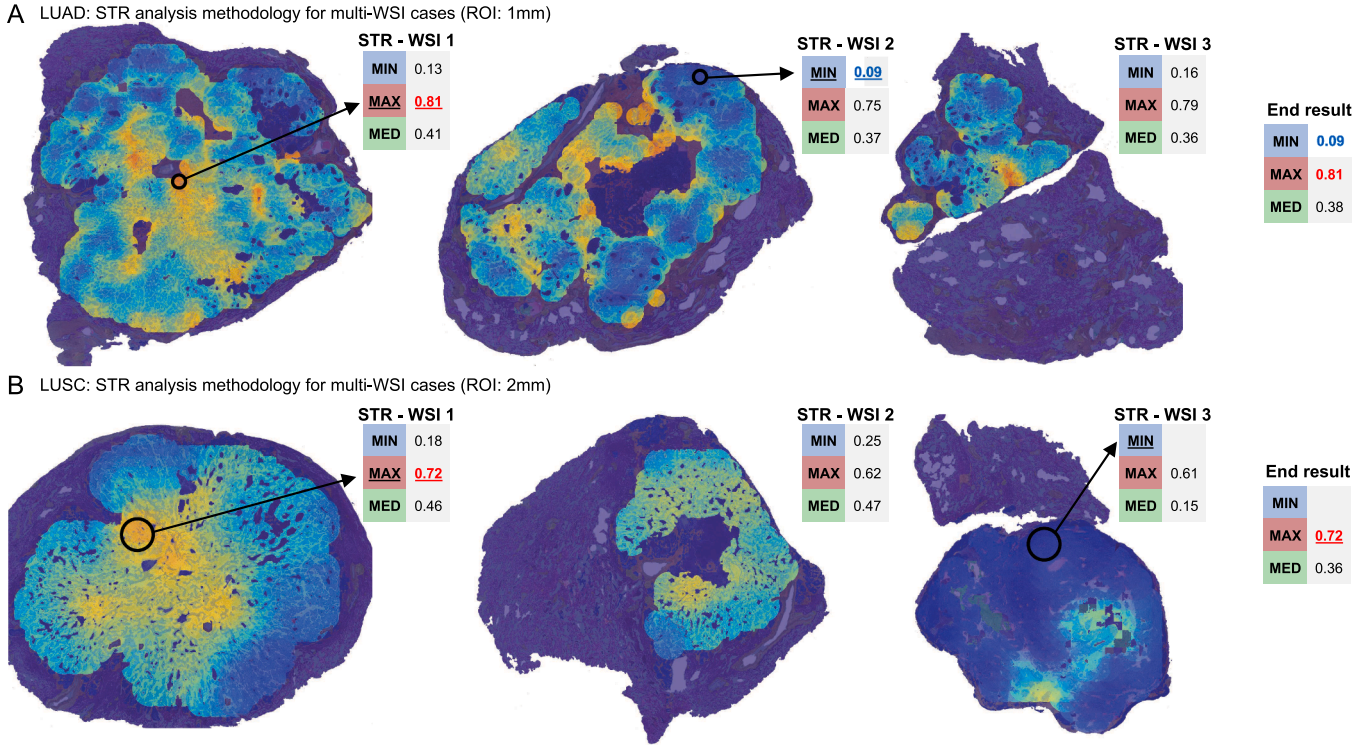


Fig. 3. Aggregation of STR values from multiple whole-slide images in single patient cases. Heatmaps for multiple slides of one patient (WSI 1, WSI 2, WSI 3) are shown for **(A)** a lung adenocarcinoma (LUAD) case and **(B)** a lung squamous cell carcinoma (LUSC) case. MIN: minimal, MED: median, and MAX: maximal values are provided for each WSI. For multi-slide cases, the highest STR value found across all analyzed slides is designated as the final case-level MAX-STR value of the patient. Similarly, the lowest STR value is chosen as the final case-level MIN-STR value and the final MED-STR value is calculated by averaging the generated MED values of each WSI.

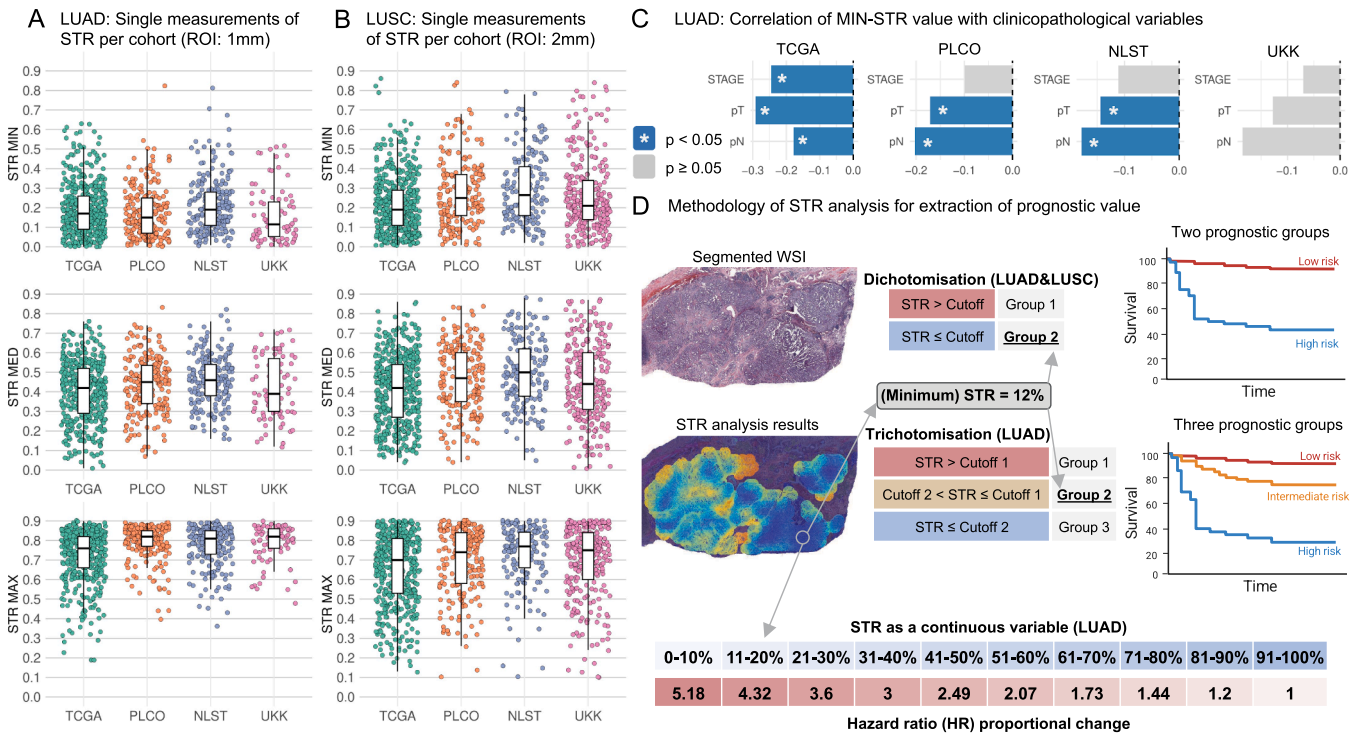


Fig. 4. STR value distribution in study cohorts, STR correlation with clinicopathological variables, and methodological approaches to STR analysis. Distribution of MIN: minimal, MED: median, and MAX: maximal WSI-level values in (A) lung adenocarcinoma (LUAD) and (B) lung squamous cell carcinoma (LUSC) cohorts. For details of statistical significance analysis of differences see Suppl. Table 1. C: Pearson's correlation analysis of clinicopathological variables and case-level MIN-STR values in LUAD cohorts (p-values < 0.05 considered as significant). D: Methodology of STR analysis for extraction of prognostic value. For LUAD, the smallest calculated STR values (case-level MIN-STR) using the ROI size of 1.0 mm are utilized to stratify patients in two or three prognostic groups (one or two cut-offs, respectively). Alternatively, STR can be analyzed as a continuous variable with increments of 10 %. For LUSC, the case-level MAX-STR calculated using the ROI size of 2.0 mm with a single cut-off is used to stratify the patients into two groups, low- and high risk.

MED values than all of the three validation cohorts ($p < 0.001$), with NLST additionally showing higher MIN and MED values than UKK and higher MAX values than both UKK and PLCO ($p < 0.05$).

Regarding correlations with clinicopathological variables, for LUAD, higher MIN-STR values were statistically significantly negatively correlated with pT and pN categories in three of four cohorts (Fig. 4C), with a similar negative correlation to the grade of the tumor apparent for the PLCO and NLST cohorts (Suppl. Figs. 13–14, Suppl. Table 2; grading was performed before the IASLC grading principles were released [17]). However, the effect size was rather small (Pearson's $r < 0.3$). For LUSC, no statistically significant correlations with clinicopathological parameters were evident (tumor budding was not included in the evaluated parameters).

Third, using the determined optimal ROI size and analysis principle (MIN-STR for LUAD, MAX-STR for LUSC), we evaluate the capability of STR for extracting prognostic information using three methodologies (Fig. 4D): 1) Single prognostic cut-off (two risk groups), 2) Two prognostic cut-offs (three risk groups), and 3) using STR measurements as a continuous variable.

3.3. Evaluation of STR prognostic role in LUAD

During the exploration phase (TCGA cohort), a MIN-STR of 0.2 was found to be the optimal prognostic cut-off for dichotomization (two risk groups) (Fig. 5A), while cut-offs of 0.07 and 0.2 provided best results for stratification of patients into three risk groups (Fig. 5B). For dichotomization and multivariate analysis including common clinicopathological variables (AGE, pT, pN), HR for lower MIN-STR values were 1.96 (95 %CI 1.34–2.87) and 2.17 (1.29–3.63) for OS and CSS, respectively (all $p < 0.003$; Fig. 5A). Trichotomization allowed for even finer risk stratification (Fig. 5B; all $p < 0.014$). For OS and CSS, lower MIN-STR values, when used as a continuous parameter, were associated with increased risk of negative prognostic outcomes, with a HR of 1.20 (1.04–1.38) and 1.33 (1.10–1.62) for each 0.1 decrease in parameter value, respectively (multivariate Cox-analysis, all $p < 0.024$; Fig. 5C). For the PFS endpoint, neither of the three methodological approaches were independently associated with prognostic endpoints in the exploration TCGA cohort (Fig. 5A,B,C).

During independent validation in the merged PLCO and NLST cohorts (two screening trial cohorts with similar patient populations and long-term follow-up), MIN-STR showed independent prognostic value across all three methodologies for survival prediction in multi-variate analysis with common clinicopathological variables (Fig. 5D,E,F). Additionally, incorporating tumor grade as an additional covariate in the multivariate Cox analysis maintained this prognostic significance of MIN-STR (Suppl. Fig. 15). Interestingly, also for the PFS endpoint, MIN-STR was independently associated with prognosis, with lower STR values showing HR of 2.47 (1.28–4.78) for dichotomized risk groups, and HR of 2.42 (1.11–4.86) and 2.60 (1.11–6.06) for intermediate and high-risk groups (two prognostic cut-offs), all $p < 0.027$ (Fig. 5D,E). Next, in the full study cohort (all four cohorts together: TCGA, PLCO, NLST, UKK; $n > 930$) MIN-STR could be validated as having independent prognostic value for all three endpoints (OS, CSS, PFS), independent of evaluation methodology used (Fig. 6A,B,C; separate results for validation cohorts in Suppl. Figs. 10, 12).

3.4. Evaluation of STR prognostic role in LUSC

In the exploration cohort (TCGA), using the determined best approach (MAX-STR, ROI size 2.0 mm), we found that only very high cut-offs (MAX-STR = 0.86) allow for meaningful stratification of patients leading to identification of a small sub-group of patients (41/439

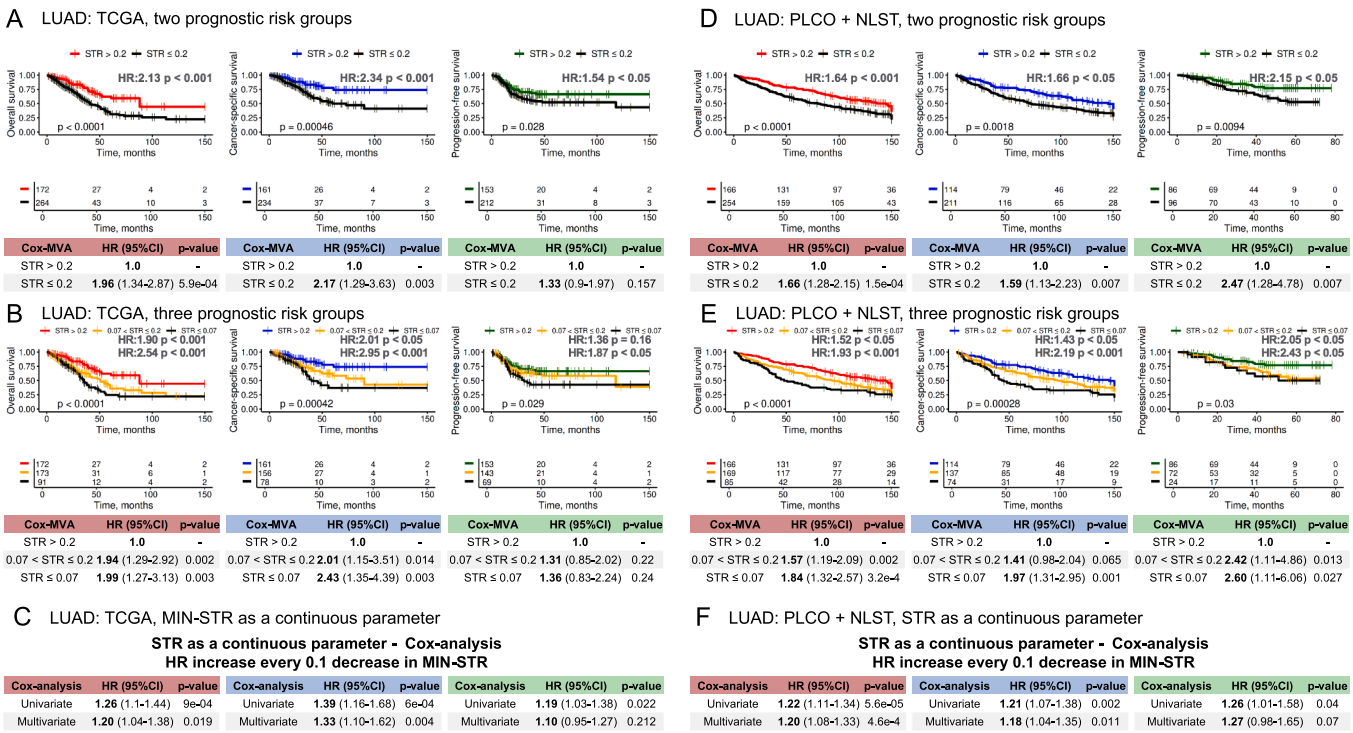


Fig. 5. STR prognostic value in patients with lung adenocarcinoma (LUAD) using different analysis approaches in exploration and validation cohorts. A-C: Exploration cohort (TCGA) using two risk groups (A), three risk groups (B), and STR as continuous variable (C). **D-F:** Merged PLCO/NLST independent validation cohort using two risk groups (D), three risk groups (E), and STR as continuous variable (F). The analysis is presented for three different clinical endpoints (overall survival, cancer-specific survival, and progression-free survival). In all analyses, the minimal case-level STR value is used (MIN-STR). The hazard ratio (HR)/p-level in the upper right part of Kaplan-Meier curves stems from univariate Cox-regression analysis. The p-level in the lower left part of Kaplan-Meier curves is for log-rank test. The table below each Kaplan-Meier plot is the result of multi-variate Cox analysis with inclusion of pT, pN, and patient age (latter – for overall survival). A cut-off of 150 months for follow-up data was used for all cohorts after which only few patients had available follow-up. Abbreviations: STR – stroma-tumor ratio, MVA – Cox multivariate analysis.

in OS analysis, 35/397 in CSS, and 33/374 patients in PFS analysis) with an unfavorable prognosis (Fig. 6D). Applying this cut-off conferred independent prognostic value in multivariate Cox-analysis with HR 2.01 (1.34–3.01), 2.52 (1.47–4.32), and 2.25 (1.28–3.98) for OS, CSS, and PFS endpoints, respectively (all $p < 0.005$; Fig. 6D).

However, discrepant results were evident when validating the determined stratification cut-off in independent cohorts. In the smaller UKK cohort, underpowered for multivariate analysis, MAX-STR allowed the patient stratification in the same prognostic vector as in TCGA cohort (Fig. 6F). Notably, in combined PLCO/NLST cohorts, the same cut-off stratified the patients in the opposite way, with higher MAX-STR values correlated with better prognosis (Fig. 6E). The outlier nature of the identified cut-off and the discrepant validation results indicate an uncertain role of STR evaluation in LUSC, especially when compared to LUAD.

4. Discussion

In this study, we establish a digital pathology pipeline for the quantification of the STR in patients with resectable NSCLC (both LUAD and LUSC), Fig. 1A, B. Our pipeline has several advantages. It allows for high precision tissue analysis directly from H&E-stained sections, thanks to a highly accurate multi-class tissue segmentation algorithm that robustly identifies tumor regions and segments the tumor into different compartments (Fig. 2A,B). It is fully automatic and requires no intervention from pathologists (e.g., for defining tumor region); it effectively aggregates information, should multiple slides with tumor be available for analysis (Fig. 3). The STR analysis pipeline systematically evaluates each available tumor region and implements a rigorous eligibility control step. Importantly, the STR analysis approach was initially developed by humans for humans (e.g., ROI size: typically 1 power field under 10x magnification, approx. 2 mm; content cut-off: 50 %) [5]. In our study, we systematically approach and refine the methodology of STR analysis, demonstrating the optimal ROI size and stroma-tumor cutoffs for patient risk stratification for both LUAD and LUSC (Suppl. Fig. 4).

In LUAD patients, we show that lower case-level STR values (less stroma, more epithelial tumor component) are associated with negative outcomes, while for LUSC, stroma-high regions are linked to an unfavorable prognosis in a portion of the tested cohorts (Fig. 4D,5,6, Suppl. Figs. 2-3). Accordingly, identifying the tumor area with minimal STR (MIN-STR) and maximal STR value (MAX-STR) is of importance for LUAD and LUSC cases. For LUAD, we show that STR is an independent prognostic parameter, regardless of the analysis approach and principle of patient risk stratification (two prognostic groups, three prognostic groups, or using MIN-STR as a continuous variable, Fig. 4D), also when considering typical clinicopathological variables (pT, pN, patient age). Its value at the established analysis methodology and analysis cut-offs were validated using large (LUSC/LUAD: slides n = 1501/1121, cases n = 952/734), well characterized, multi-institutional, international independent cohorts of patients (Fig. 5A-F, Fig. 6A-C) for all prognostic endpoints (OS, CSS, PFS).

Regarding LUSC tumors, we detected discrepant results across study cohorts. In the exploratory cohort (TCGA) as well as in one additional independent cohort (UKK), we were able to identify a small subgroup of patients with very high MAX-STR values that showed independent prognostic value for negative outcome for several clinical endpoints (Fig. 6D,F). However, the prognostic trends were inverse in the merged

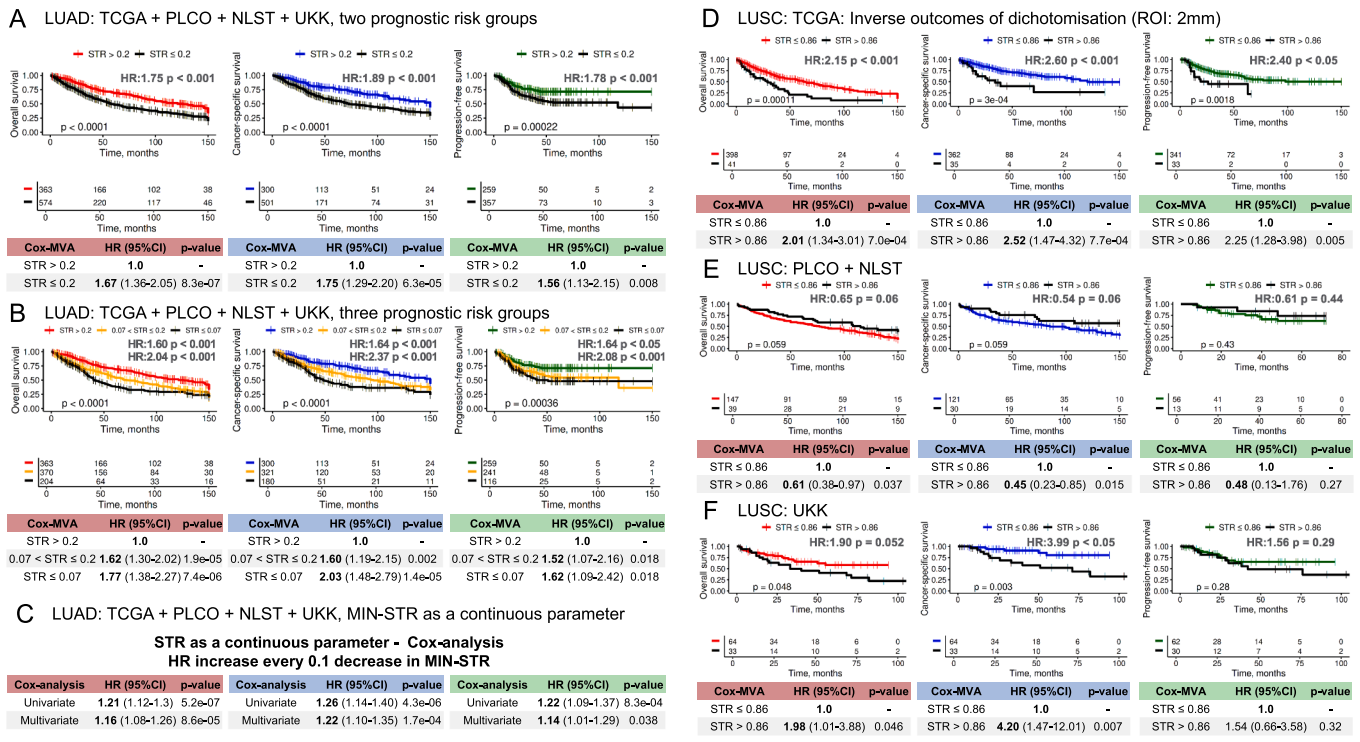


Fig. 6. STR prognostic value in patients with lung adenocarcinoma (LUAD) and squamous cell carcinoma (LUSC) using different analysis approaches. A-C: Full merged study cohort (TCGA/PLCO/NLST/UKK) using two risk groups (A), three risk groups (B), and STR as continuous variable (C). **D-F:** Analysis of prognostic value in LUSC cohorts: (D) Exploratory TCGA cohort, (E) Merged PLCO/NLST independent validation cohort, (F) UKK independent validation cohort. The analysis is presented for three different clinical endpoints (overall survival, cancer-specific survival, and progression-free survival). In all analyses, the minimal case-level STR value is used (MIN-STR). The hazard ratio (HR)/p-level in the upper right part of Kaplan-Meier curves stems from univariate Cox-regression analysis. The p-level in the lower left part of Kaplan-Meier curves is for log-rank test. The table below each Kaplan-Meier plot is the result of multi-variate Cox analysis with inclusion of pT, pN, and patient age (latter – for overall survival). A cut-off of 150 months for follow-up data was used for all cohorts after which only few patients had available follow-up. Abbreviations: STR – stroma/tumor ratio, MVA – Cox multivariate analysis.

NLST/PLCO cohort. Together with an outlier cut-off position, this implies that the prognostic value of STR in LUSC might not be as clear-cut as in LUAD. Further investigations in other patient cohorts are necessary to address this issue.

Machine learning-based quantification of the STR in NSCLC has been attempted in few studies [12,13]. Micke et al. [12] used a segmentation method that relies on pan-cytokeratin staining and autofluorescence patterns. Moreover, they used tissue-microarray (TMA) for their analysis, which is prone to significant selection biases and does not represent typical diagnostic practice. Our pipeline/algorithm is fully automatic and works directly with routine, diagnostic H&E-stained histological slides, without the need for additional investigations such as immunohistochemistry. We effectively address prominent STR intratumoral heterogeneity through fully systematic approach. Interestingly, Micke et al. [12] did not find any prognostic significance for STR in LUAD patients (OS) which might be related to above mentioned methodological issues or the small number of patients included (n = 171). In patients with LUSC (n = 90), the authors found similar trends to those we observed in the NLST/PLCO cohort. Similarly, Koike et al. [13] also relied on immunohistochemical pan-cytokeratin staining for the partially automated differentiation between tumor and stroma regions in a cohort (n = 135) of LUSC resection cases (necrotic regions were annotated manually). Using a non-deep learning method, they showed that stroma-high cases present with unfavorable disease outcomes (PFS), similar to our results in the TCGA and UKK cohorts. These two studies underscore the ambiguous role of STR for LUSC and warrant additional studies.

Several studies investigated STR in NSCLC without applying advanced image analysis methods. Ichikawa et al. [9] studied a cohort of 127 LUAD patients and demonstrated that stroma-high case tumor cases (<10 % epithelial component) showed significantly better prognosis (OS, PFS), validating our findings in LUAD. Ichikawa et al. estimated STR only in invasive regions, deviating from classical methodology, which might be methodologically challenging in lung cancer due to a lacking definition of the invasive margin and the often lepidic growth in such areas. Interestingly, in our hands, using a modified approach with analysis of median STR case value instead of regional analysis failed to demonstrate any correlation with the patient survival endpoints in both LUAD and LUSC cohorts (Suppl. Fig. 2A, 3A). This might imply that only the worst regions in the tumor are decisive, which aligns with the concept of tumor molecular-genetic evolution toward an aggressive clone. Notably, by focusing on the minimum STR value, we observed that even a single nest of patches with a stroma-tumor ratio of 20 % or lower is associated with poorer prognosis. These localized regions of high tumor cellularity likely represent aggressive nests of clones where rapid proliferation displaces the surrounding stroma. This notion is further supported by the finding that even lower STR values correlate with worse outcomes, indicating faster tumor growth and increased tumor burden. Also, Smit et al. [10] assessed the prognostic value of STR in LUSC patients (n = 174); STR was evaluated visually by pathologists using a classical cut-off of 50 %. Their findings are consistent with our results in the TCGA/UKK cohorts, with stroma-high tumors showing negative outcomes (OS, CSS). The STR analysis in patients with squamous cell carcinoma of other localizations (esophagus, n = 95 [18] and oropharyngeal, n = 182 [19]) is in concordance with Smit et al. [10] and our TCGA/UKK analysis in LUSC. In the study of Xi et al. [8], LUAD and LUSC patients (n = 182 and 79, respectively) were evaluated together as one cohort, which might be suboptimal given our results and those of the

aforementioned studies.

Our study does not involve direct comparison between pathologists and AI tools due to several important reasons. First, our STR algorithm relies on tissue segmentation masks produced by a backbone multi-class tissue segmentation model for lung cancer. As long as the segmentation masks are precise, the STR algorithm provides objective information to STR status as it systematically assesses all eligible regions within the tumor. The backbone algorithm was independently validated to have an excellent segmentation accuracy across all relevant classes [15]. Second, a typical NSCLC resection specimen contains 3–5 whole slide images. Both finding the most relevant region in a single slide and cognitive reconstruction from multiple slides are difficult tasks for human pathologists prone to high levels of interobserver variability [5,20]. Third, there is clear evidence for different levels of precision in STR quantification by AI tools (pixel-wise mathematic precision) and human beings ("eyeballing") resulting in significant threshold shifts towards higher/lower values, indicating that AI pixel-wise and pathologists' measurements are not directly comparable [21]. Altogether, this indicates that direct comparison between fully automatic analysis by an AI tool and pathologists' evaluation might not be very meaningful.

Our study is not devoid of limitations. Although we use large patient cohorts representative of different countries and practices, the analysis performed is retrospective, and prospective validation might be necessary, especially in patients with LUSC. Moreover, it is likely that not every WSI from each tumor was available from each case, which may impact case-level STR results and limit the full capture of all tumor regions. In the TCGA cohort, most patient cases contained only one WSI and thus this cohort has the highest probability of this bias. Based on this consideration, the TCGA cohort was selected for exploration purposes only, while other cohorts containing multiple WSIs per case in most patients were chosen for independent validation.

5. Conclusions

We have developed a robust and fully automated, AI-based pipeline for effective, objective, and systematic quantification of the STR in NSCLC. Our findings indicate that STR is a reliable, independent prognostic parameter, especially in patients with resectable lung adenocarcinoma. The STR might be useful in assessing tumor aggressiveness and identifying high-risk patients post-resection who might benefit from postoperative adjuvant therapy.

Funding sources

This project was partially funded by Federal Ministry of Education and Research of Germany: Project FED-PATH (YT, RB) and North Rhine-Westphalia state (European Fond for Regional Development (EFRE), 2014–2020; REACT-EU): Project DIGI-PATH (YT, RB). YT and WA had access to all the data.

CRediT authorship contribution statement

Waleed K.M. Ahmad: Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Tillmann Bedau:** Writing – review & editing, Formal analysis, Data curation. **Yuan Wang:** Writing – review & editing, Data curation. **Sebastian Michels:** Writing – review & editing, Validation, Data curation. **Anna Rasokat:** Writing – review & editing, Validation, Data curation. **Jürgen Wolf:** Writing – review & editing, Validation, Data curation. **Matthias Heldwein:** Writing – review & editing, Data curation. **Simon Schallenberg:** Writing – review & editing, Resources, Data curation. **Alexander Quaas:** Writing – review & editing, Validation, Resources, Data curation. **Reinhard Büttner:** Writing – review & editing, Validation, Supervision, Data curation. **Yuri Tolkach:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank the Regional Computing Center of the University of Cologne (RRZK) for providing computing time on the DFG-funded (Funding number: INST 216/512/1FUGG) High Performance Computing (HPC) system CHEOPS as well as technical support.

Data and code availability

TCGA cohort is publicly available at https://portal.gdc.cancer.gov/. PLCO cohort is publicly available at https://cdas.cancer.gov/plco/. NLST cohort is publicly available at https://cdas.cancer.gov/nlst/. UKK cohort is available on reasonable request from corresponding author (YT). The algorithm code will be made available at https://github.com/cpath-ukk after publication.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.lungcan.2025.108613.

References

[1] F. Bray, et al., Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA Cancer J. Clin.* 74 (2024) 229–263.
[2] R.L. Siegel, K.D. Miller, N.S. Wagle, A. Jemal, Cancer statistics, 2023, *CA Cancer J. Clin.* 73 (2023) 17–48.
[3] K. Bera, K.A. Schalper, D.L. Rimm, V. Velcheti, A. Madabhushi, Artificial intelligence in digital pathology — new tools for diagnosis and precision oncology, *Nat. Rev. Clin. Oncol.* 16 (2019) 703–715.
[4] A. Echle, et al., Deep learning in cancer pathology: a new generation of clinical biomarkers, *Br. J. Cancer* 124 (2021) 686–696, https://doi.org/10.1038/s41416-020-01122-x.
[5] G.W. van Pelt, et al., Scoring the tumor-stroma ratio in colon cancer: procedure and recommendations, *Virchows Arch.* 473 (2018) 405–412.
[6] M. Polack, et al., Results from the UNITED study: a multicenter study validating the prognostic effect of the tumorestroma ratio in colon cancer, *ESMO Open* 9 (2024) 102988.
[7] T. Zhang, et al., Tumor-stroma ratio is an independent predictor for survival in NSCLC, *Int. J. Clin. Exp. Pathol.* 8 (2015) 11348.
[8] K.-X. Xi, et al., Tumor-stroma ratio (TSR) in non-small cell lung cancer (NSCLC) patients after lung resection is a prognostic factor for survival, *J. Thorac. Dis.* 9 (2017) 4017–4026.
[9] T. Ichikawa, et al., The ratio of cancer cells to stroma within the invasive area is a histologic prognostic parameter of lung adenocarcinoma, *Lung Cancer* 118 (2018) 30–35.
[10] M.A. Smit, et al., The prognostic value of the tumor-stroma ratio in squamous cell lung cancer, a cohort study, *Cancer Treat. Res. Commun.* 25 (2020) 100247.
[11] K. Zhao, et al., Artificial intelligence quantified tumour-stroma ratio is an independent predictor for overall survival in resectable colorectal cancer, *EBioMedicine* 61 (2020) 103054.
[12] P. Micke, et al., The prognostic impact of the tumour stroma fraction: a machine learning-based analysis in 16 human solid tumour types, *EBioMedicine* 65 (2021) 103269.
[13] Y. Koike, et al., Machine learning-based histological classification that predicts recurrence of peripheral lung squamous cell carcinoma, *Lung Cancer* 147 (2020) 252–258.
[14] Smith G. GitHub - ghsmith/SVSUtil: Apply 3D LUT color transform to an SVS, updating the image tiles in-situ. https://github.com/ghsmith/SVSUtil.
[15] C. Kludt, et al., Next-generation lung cancer pathology: development and validation of diagnostic and prognostic algorithms, *Cell Rep. Med.* 5 (2024) 101697.
[16] J. Wu, C. Liang, M. Chen, W. Su, Association between tumor-stroma ratio and prognosis in solid tumor patients: a systematic review and meta-analysis, *Oncotarget* 7 (2016) 68954–68965.

[17] A.L. Moreira, et al., A grading system for invasive pulmonary adenocarcinoma: a proposal from the international association for the study of lung cancer pathology committee, *J. Thorac. Oncol.* 15 (2020) 1599–1610.
[18] K. Wang, et al., Tumor-stroma ratio is an independent predictor for survival in esophageal squamous cell carcinoma, *J. Thorac. Oncol.* 7 (2012) 1457–1461.
[19] A. Almangush, et al., Tumor-stroma ratio is a promising prognostic classifier in oropharyngeal cancer, *Hum. Pathol.* 136 (2023) 16–24.
[20] Y. Tolkach, T. Dohmgörgen, M. Toma, G. Kristiansen, High-accuracy prostate cancer pathology using deep learning, *Nat. Mach. Intell.* 7 (2) (2020) 411–418.
[21] O.G.F. Geessink, et al., Computer aided quantification of intratumoral stroma yields an independent prognosticator in rectal cancer, *Cell. Oncol. (Dordrecht, Netherlands)* 42 (2019) 331–341.