Mod Pathol 36 (2023) 100055

MODERN PATHOLOGY

USCAP UNITED STATES AND CANADIAN ACADEMY OF PATHOLOGY
Creating a Better Pathologist

Journal homepage: https://modernpathology.org/

Research Article

# Digital Quantification of Tumor Cellularity as a Novel Prognostic Feature of Non–Small Cell Lung Carcinoma

Sherman Lin[a], Joshua P. Samsoondar[b], Ela Bandari[b], Samantha Keow[a], Binit Bikash[a], Djarren Tan[a], Jacobo Martinez-Acevedo[a], John Loggie[c], Michelle Pham[a], Nina J. Wu[a], Tanya Misra[a], Victor H.K. Lam[a], Irene Sansano[d], Matthew J. Cecchini[a, *]

[a] *Department of Pathology and Laboratory Medicine, University of Western Ontario, London, Ontario, Canada;* [b] *Schulich School of Medicine and Dentistry, University of Western Ontario, London, Ontario, Canada;* [c] *Department of Pathology, Dalhousie University, Halifax, Nova Scotia, Canada;* [d] *Department of Pathology, Hospital Universitari Vall d'Hebron, Barcelona, Catalunya, Spain*

ARTICLE INFO

*Article history:*
Received 1 November 2022
Accepted 19 November 2022
Available online 10 January 2023

*Keywords:*
cancer
image analysis
digital pathology
thoracic pathology

ABSTRACT

Non–small cell lung carcinoma is currently staged based on the size and involvement of other structures. Tumor size may be a surrogate measure of the total number of tumor cells. A recently revised reporting system for adenocarcinoma incorporates high-risk histologic patterns, which may have increased cellular density. Modern digital image analysis tools can be utilized to automate the quantification of cells. In this study, we tested the hypothesis that tumor cellularity can be used as a novel prognostic tool for lung cancer. Digital slides from The Cancer Genome Atlas lung adenocarcinoma (ADC) data set (n = 213) and lung squamous cell carcinoma (SCC) data set (n = 90) were obtained and analyzed using QuPath. The number of tumor cells was normalized with the surface area of the tumor to provide a measure of tumor cell density. Tumor cellularity was calculated by multiplying the size of the tumor with the cell density. Major histologic patterns and grade were compared with the tumor density of the lung ADC and lung SCC cases. The overall and progression-free survival were compared between groups of high and low tumor cellularity. High-grade histologic patterns in the ADC and SCC cases were associated with greater tumor densities compared with low-grade patterns. Cases with lower tumor cellularity had improved overall and progression-free survival compared with cases with higher cellularity. These results support tumor cellularity as a novel prognostic tool for non–small cell lung carcinoma that considers tumor stage and grade elements.



## Introduction

Lung cancer is a leading cause of cancer-related deaths.[1] Treatment decisions are driven by numerous features of tumors such as stage, grade, and molecular drivers. However, the risk of recurrence or progression is not completely captured by these features, and there is a need for better prognostic tools to identify patients who may benefit from personalized treatments.

The current staging system for non–small cell lung carcinoma (NSCLC) is defined by the size of the tumor and its involvement in other structures.[2] The grading of NSCLC was recently updated in the fifth edition of the World Health Organization (WHO)

* Corresponding author.
E-mail address: matthew.cecchini@lhsc.on.ca (M.J. Cecchini).


https://doi.org/10.1016/j.modpat.2022.100055

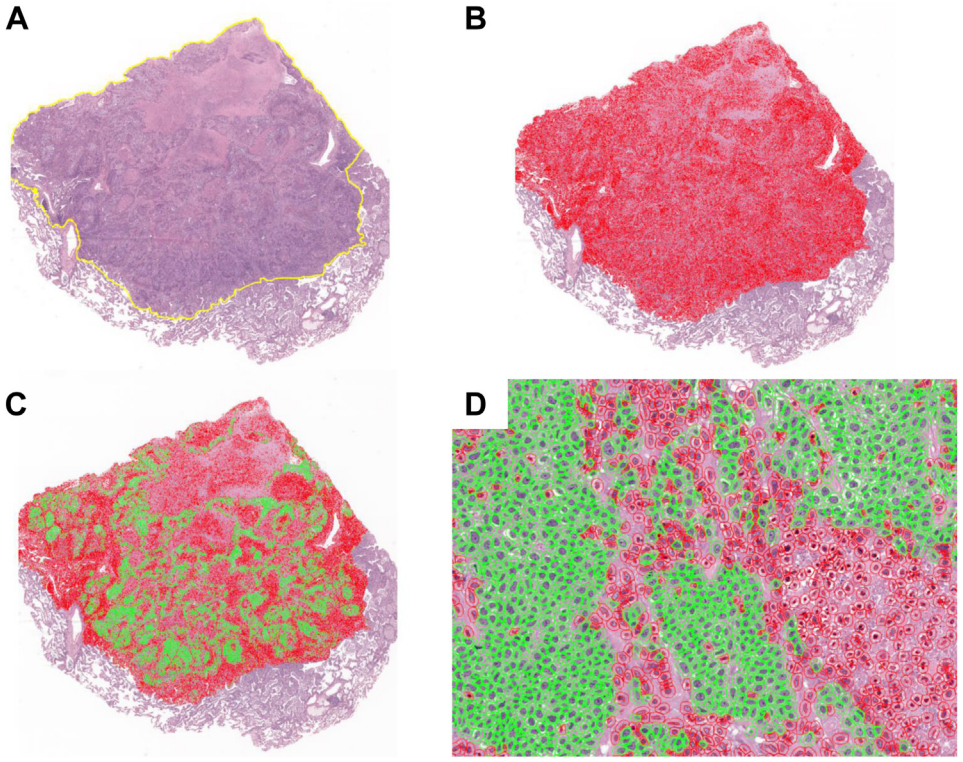


**Figure 1.**
Semiautomated calculation of tumor density. (A) The outline of the tumor was contoured (yellow line). (B) A total cell detection image analysis was performed to detect both tumor and nontumor cells. (C) Tumor areas were annotated, and using features from cell detection, a random trees-based object classifier in QuPath was used to classify tumor cells from background stroma. (D) A high-power image showing cell detection following object classification is shown, with tumor cells in green and stromal cells in red.

classification of thoracic tumors. It now includes the recognition of high-risk patterns of tumor growth that are associated with a poor prognosis.[3,4] The high-risk patterns of lung adenocarcinoma (ADC) include micropapillary, solid, and complex glandular patterns. In general, these high-grade patterns tend to visually have a greater cellular density than lower-grade patterns of lepidic, acinar, and papillary. However, this has not yet been formally measured or quantified.[5] Multiple studies have shown the identification and quantification of these morphologic patterns to have a high degree of interobserver variability and subjectivity.[6-9] Squamous cell carcinoma (SCC) continues to be descriptively graded into well-differentiated, moderately differentiated, and poorly differentiated tumors. Currently, the WHO does not recognize an established grading system for SCC lung cancer.[3]

The current grading and staging systems are typically assessed by pathologists using conventional analog microscopes. However, the increasing digitization of histologic slides presents a novel opportunity to further quantify and characterize tumors. The standard T staging and grading systems for lung cancer consider tumor size and the distribution of histologic patterns.[2] However, considerable variation in survival of patients with the same T stage suggests that tumor size, a major component of the TNM staging system, can be further improved with advancements in digital pathology tools.[10] Tumor size may serve as a surrogate measure for the total number of tumor cells. Furthermore, given the high architectural density of high-grade patterns of tumor growth, the measurement of tumor cellularity could provide a means to incorporate both tumor grade and T stage.

Counting all cells in large tumors is not feasible using analog microscopes. However, by leveraging the power of digital pathology using image analysis and machine learning, it is possible to effectively and efficiently quantify the number of tumor cells on histopathology slides to generate a measure of tumor density. The density can then be multiplied by tumor size to generate the overall tumor cellularity value. Here, we applied this novel system to a cohort of NSCLC cases and showed that tumor cellularity is a novel predictor of outcomes in patients with NSCLC.

## Methods

### *Patient and Clinical Samples*

The retrospective study population was drawn from digitized slide data sets of NSCLC from The Cancer Genome Atlas (TCGA). Cases from the lung ADC data set (n = 213) and lung SCC data set (n = 90) were used. One representative slide was available for each case. Patients with incomplete clinical information needed to stage the tumor were excluded from this study. Of this subset of patients, eligible cases were selected randomly based on the numerical order of TCGA case numbers to be included in the data set. Information in the available pathology report was utilized to stage all the cases using the current eighth edition of the AJCC staging system for lung carcinoma.[2]

### *Semiautomated Tumor Cell Counting*

Using QuPath,[11] a thoracic pathologist (M.J.C.) annotated and reviewed representative areas of the tumors. The annotation process consisted of contouring the tumor edge from the surrounding uninvolved lung parenchyma (Fig. 1A). The overall

**Table 1**
Multivariable analysis of overall survival

| Characteristic | HR | 95% CI | *P* value |
|---|---|---|---|
| Cellularity | | | |
| Low | — | — | |
| High | 2.04 | 1.14-3.63 | **.016** |
| Sex | | | |
| Male | — | — | |
| Female | 1.12 | 0.73-1.71 | .61 |
| Age | 1.00 | 0.98-1.02 | .76 |
| Tumor size (cm) | 0.83 | 0.65-1.06 | .14 |
| N stage | | | |
| 0 | — | — | |
| 1 | 1.87 | 1.20-2.92 | **.005** |
| >2 | 0.82 | 0.42-1.60 | .57 |
| T stage | | | |
| *<pT1b* | — | — | |
| *pT1c* | 0.76 | 0.32-1.80 | .53 |
| *pT2a* | 1.14 | 0.54-2.39 | .73 |
| *pT2b* | 1.82 | 0.66-5.01 | .25 |
| *pT3* | 2.21 | 0.68-7.16 | .19 |
| *pT4* | 3.67 | 0.63-21.5 | .15 |
| Diagnosis | | | |
| Adenocarcinoma | — | — | |
| Squamous cell carcinoma | 0.80 | 0.50-1.29 | .37 |
| Grade | | | |
| Well or moderate | — | — | |
| Poor | 1.10 | 0.72-1.68 | .65 |

Bolded values indicate $P < .05$.

extent of the tumor on the representative slide was annotated and reviewed for all the cases. Within the tumors, the cell detection algorithm in QuPath[11] was utilized to count cells and extract standard nuclear and cellular features, including size, density, and staining intensity (Fig. 1B). The cells were detected using the standard cell detection function based on the optical density using hematoxylin staining. Given the staining variability among the slides, the individual parameters were varied to optimize cell detection. The optimized parameters were chosen to reduce nuclear fragmentation by varying the sigma function and excluding smaller immune cells in tumors with large numbers of inflammatory cells by increasing the minimum nuclear area. The default parameters were utilized in the majority of the cases, and only a small subset required modification of these parameters to optimize cell detection. The cell detection step of the study was performed blinded to outcome data, and the optimization of cell detection for each case was reviewed by a thoracic pathologist (M.J.C.). The detected cell features were used to generate a cell classifier that utilized a random trees-based object classifier that was trained on manually annotated areas of the tumors as training data (Fig. 1C, D). Given the heterogeneity of the tumors and stain intensity among the cases, we trained the object classifier on each case independently and reviewed all the cases manually to ensure accuracy. A small subset of the cases was excluded from the study (n = 29) because of the inability to develop a reliable object classifier secondary to low cellularity, staining properties, or other artifacts.

### *Tumor Density and Tumor cellularity*

The absolute number of tumor cells was recorded and normalized by the surface area of the tumors on the slides to generate a measure of tumor density. To generate a measure of overall tumor cellularity, we multiplied the tumor density by the surface area of the total tumor using the greatest linear dimension as the diameter. The greatest dimension was abstracted from the pathology report available in the TCGA data set.

### *Correlation of Tumor Density With Tumor grades and Histologic Patterns*

Given the known variability in the cell density of NSCLCs, we selected representative 1260 × 1260-μm panels in which the tumors could be completely manually contoured, and the histologic pattern and grade could be reliably determined. Representative panels with an ADC growth pattern or SCC morphologic grade were chosen. In these areas, we manually contoured all tumor cells using the brush tool in QuPath, with all contours reviewed and refined by a thoracic pathologist (M.J.C. or I.S.). Within the contours, the cell detection algorithm described above was utilized to enumerate the total number of tumor cells in the annotated area.

### *Patient Demographics and Outcome Measurements*

Available patient data were abstracted from the TCGA database and pathology reports. The pathology reports were reviewed and restaged based on the eighth edition of the TNM staging system by a thoracic pathologist (M.J.C.). Patient demographics and tumor characteristics were outlined using descriptive statistics. Overall, the progression-free survival was determined and assessed using Kaplan-Meier curves and log-rank tests. Multivariable analyses were performed using Cox proportional hazards regression. Clinicopathologic factors, including sex, age, tumor size, stage,

**Table 2**
Multivariable analysis of progression-free survival

| Characteristic | HR | 95% CI | *P* value |
|---|---|---|---|
| Cellularity | | | |
| Low | — | — | |
| High | 2.13 | 1.24-3.66 | **.006** |
| Sex | | | |
| Male | — | — | |
| Female | 0.98 | 0.67-1.43 | .91 |
| Age | 1.01 | 0.99-1.02 | .54 |
| Tumor size (cm) | 0.92 | 0.79-1.07 | .26 |
| N stage | | | |
| 0 | — | — | |
| 1 | 1.54 | 1.00-2.39 | **.050** |
| >2 | 0.77 | 0.43-1.38 | .38 |
| T stage | | | |
| <pT1b | — | — | |
| pT1c | 0.90 | 0.43-1.89 | .78 |
| pT2a | 1.01 | 0.52-1.97 | .98 |
| pT2b | 1.67 | 0.72-3.86 | .23 |
| pT3 | 1.62 | 0.63-4.19 | .32 |
| pT4 | 3.40 | 0.85-13.7 | .085 |
| Diagnosis | | | |
| Adenocarcinoma | — | — | |
| Squamous cell carcinoma | 1.28 | 0.80-2.04 | .31 |
| Grade | | | |
| Well or moderate | — | — | |
| Poor | 1.36 | 0.92-2.00 | .13 |

Bolded values indicate $P < .05$.

**Table 3**
Stepwise cellularity analysis of overall and progression-free survival

| Characteristic | Overall survival | | | Progression-free survival | | |
|---|---|---|---|---|---|---|
| | HR | 95% CI | *P* value | HR | 95% CI | *P* value |
| Cellularity bins | | | | | | |
| Q1 − 1.1e08 | — | — | | — | — | |
| Q2 − 2.4e10 | 1.65 | 0.85-3.07 | .11 | 1.75 | 0.99-3.10 | .055 |
| Q3 − 5.4e10 | 1.12 | 0.55-2.27 | .76 | 0.96 | 0.48-1.92 | .90 |
| Q4 − 1.3e12 | 2.26 | 1.00-5.11 | **.049** | 2.17 | 1.00-4.72 | .051 |
| Sex | | | | | | |
| Male | — | — | | — | — | |
| Female | 1.17 | 0.76-1.80 | .49 | 1.02 | 0.69-1.50 | .93 |
| Age | 1.00 | 0.99-1.03 | .63 | 1.01 | 0.99-1.03 | .46 |
| Tumor size (cm) | 0.83 | 0.64-1.07 | .15 | 0.92 | 0.79-1.07 | .31 |
| N stage | | | | | | |
| 0 | — | — | | — | — | |
| 1 | 1.93 | 1.24-3.00 | **.004** | 1.60 | 1.04-2.47 | **.033** |
| >2 | 0.74 | 0.38-1.46 | .38 | 0.73 | 0.41-1.32 | .30 |
| T stage | | | | | | |
| <pT1b | — | — | | — | — | |
| pT1c | 0.64 | 0.26-1.54 | .32 | 0.77 | 0.36-1.66 | .51 |
| pT2a | 1.02 | 0.47-2.24 | .95 | 0.97 | 0.47-1.98 | .93 |
| pT2b | 1.63 | 0.56-4.76 | .37 | 1.72 | 0.68-4.39 | .25 |
| pT3 | 2.22 | 0.69-7.14 | .18 | 1.73 | 0.66-4.55 | .26 |
| pT4 | 3.78 | 0.65-22.0 | .14 | 3.70 | 0.91-15.0 | .067 |
| Diagnosis | | | | | | |
| Adenocarcinoma | — | — | | — | — | |
| Squamous cell carcinoma | 0.80 | 0.50-1.28 | .35 | 1.27 | 0.79-2.03 | .32 |
| Grade | | | | | | |
| Well or moderate | — | — | | — | — | |
| Poor | 1.14 | 0.74-1.75 | .56 | 1.46 | 0.96-2.14 | .077 |

Bolded values indicate $P < .05$.

**Table 4**
Characteristics of patients with non–small cell lung carcinoma

| | ADC (n = 213) | SCC (n = 90) |
|---|---|---|
| Diagnostic age (y) | | |
| Mean | 64.4 | 67.1 |
| SD | 10.3 | 9.1 |
| Sex, n (%) | | |
| Male | 123 (57.7) | 23 (25.6) |
| Female | 90 (42.3) | 67 (74.4) |
| Tumor size (cm) | | |
| Mean | 3.8 | 4.6 |
| SD | 2.4 | 2.1 |
| N stage, n (%) | | |
| 0 | 147 (69.0) | 56 (62.2) |
| 1 | 41 (19.2) | 25 (27.8) |
| >2 | 25 (11.7) | 9 (10.0) |
| T stage, n (%) | | |
| T1 | 76 (35.7) | 13 (14.4) |
| T2 | 91 (42.7) | 47 (52.2) |
| T3 | 31 (14.6) | 21 (23.3) |
| T4 | 15 (7.0) | 9 (10.0) |

ADC, adenocarcinoma; SCC, squamous cell carcinoma.

grade, and NSCLC subtype, were adjusted for in the multivariable analyses (Tables 1-3).

## Results

### *Patient Demographics and Determining Tumor Cellularity*

The cases of lung ADC (n = 213) and SCC (n = 90) were assessed. The patient demographics and tumor characteristics were evaluated (Table 4). In this cohort, the majority of the tumors were staged as ≥T2, and a subset of patients (31.0% of those with ADC and 37.8% of those with SCC) had lymph node involvement (pN1 or pN2).

Using Qupath,[11] we developed a semiautomated system to detect tumor cells on the slides (Fig. 1). By comparing automated and manual cell counts, the accuracy of the object classifier was estimated to be at least 85% to 95% for all the cases. The variability in tumor density among the cases was multifactorial and related to the presence of fibrosis, necrosis, and differences in tumor growth patterns.

### *Association Between Tumor Grade and Cell Density*

To better quantify the cellularity of each of the various tumor growth patterns, we manually contoured tumor areas in the representative 1260 × 1260-μm panels. In these areas, a single predominant histologic type could be defined in order to study a more homogenous population of tumor cells. There was a significant difference in the density of the different patterns of lung ADC, with high-grade patterns (micropapillary, solid, and complex glandular) having a higher density than other patterns (Fig. 2). Similar findings were also observed in squamous lung cancers, with poorly differentiated tumors having an increased density compared with more differentiated tumors (Fig. 3).

The complex glandular group noted a range of density (Fig. 2). Cases with the complex glandular pattern can be further divided into those with cribriform and those with fused glands.[12] Both the patterns had a range of cellularity, including the fused glands group, with some cases having lower cellularity, characterized by small, fused nests of tumor cells in a desmoplastic stroma (Supplementary Fig. S1A). Other cases had higher cellularity, with irregular back-to-back glands without intervening stroma (Supplementary Fig. S1B).

### *Association of Cellularity and Patient Outcomes*

Given the association between tumor density and histologic grade, we sought to evaluate whether an estimate of total tumor cellularity could be utilized to generate a novel prognostic marker that combines grading and staging information. We plotted the tumor cellularity versus the survival of the patients and identified an apparent inflection point in survival at $5.5 \times 10^{10}$ cells (Supplementary Fig. S2); we used this as a cutoff to define cases of low ($<5.5 \times 10^{10}$ cells) and high ($>5.5 \times 10^{10}$ cells) cellularity. We identified a significant difference in both overall survival (Fig. 4A) and progression-free survival (Fig. 4D). In NSCLC, there was a difference in both progression-free survival (hazard ratio [HR], 2.69; 95% CI, 1.67-4.32; $P < .0001$) and overall survival (HR, 2.22; 95% CI, 1.35-3.63; $P < .0001$). A difference in progression-free survival was also identified for both SCC (HR, 2.61; 95% CI, 1.03-6.60; $P = .0077$) and ADC (HR, 2.64; 95% CI, 1.53-4.56; $P < .0001$). A difference was also observed in overall survival for the SCC (HR, 2.42; 95% CI, 1.05-5.57; $P = .0061$) and ADC cases (HR, 1.99; 95% CI, 1.10-3.62; $P = .0052$). After adjusting for clinicopathologic factors, cellularity was found to be an independent predictor of patient outcomes in terms of both overall survival (HR, 2.06; 95% CI, 1.16-3.66; $P = .014$) and progression-free survival (HR, 2.22; 95% CI, 1.30-3.80; $P = .004$) in patients with NSCLC. We further explored the use of cutoff cellularity using an additional stepwise analysis for overall survival and progression-free survival using 4 quartiles of cellularity (Table 3). We found a significant difference in the highest quartile of cellularity for overall survival, which trended to significance for progression-free survival.

## Discussion

One of the main purposes of pathology reporting is to provide prognostic and predictive information that can be utilized to guide increasingly complex patient management decisions. The stage and grade of tumors are both important aspects of pathology reporting, and stage influences the management of patients with lung cancer.[13] In lung cancer, the T stage is largely defined by the size of the tumor, which may be a surrogate measure of the absolute number of tumor cells present in the tumor. There is a range of tumor densities that can be influenced by features such as fibrosis, necrosis, and the growth pattern of the tumor cells. Utilizing our semiautomated approach for tumor detection, we were able to capture this variability to provide a more specific measure of tumor cellularity that can account for these differences in tumor composition.

The grading of lung ADC recently underwent a series of revisions. The newest system set out in the fifth edition of the WHO classification of tumors of thoracic organs outlines a system that defines tumors composed of >20% high-grade patterns as poorly differentiated.[3] Accurate measurement of the contributions of various patterns in an overall tumor can be a very challenging and imprecise exercise. Further, there is only moderate agreement among pathologists on various distributions of histologic patterns.[6-9] In this study, we observed a greater tumor cell density in the high-grade patterns of lung ADC and poorly differentiated SCC

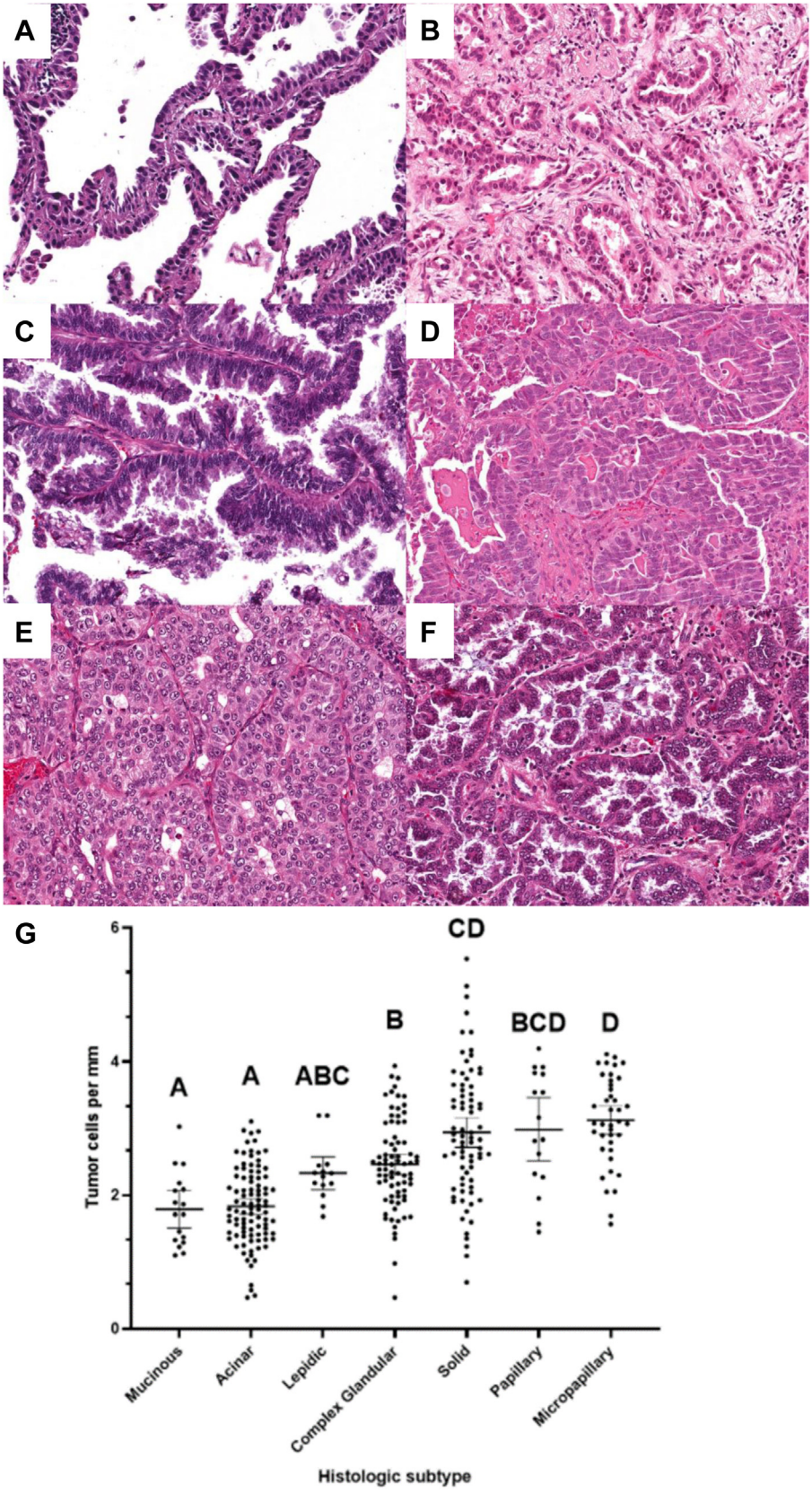


**Figure 2.**
Tumor density in lung adenocarcinoma. Representative areas of tumor demonstrating various morphologic patterns of adenocarcinoma, including (A) lepidic, (B) acinar, (C) papillary, (D) complex glandular, (E) solid, and (F) micropapillary patterns. (G) Distinct cell densities were observed in various histologic patterns of lung adenocarcinoma.

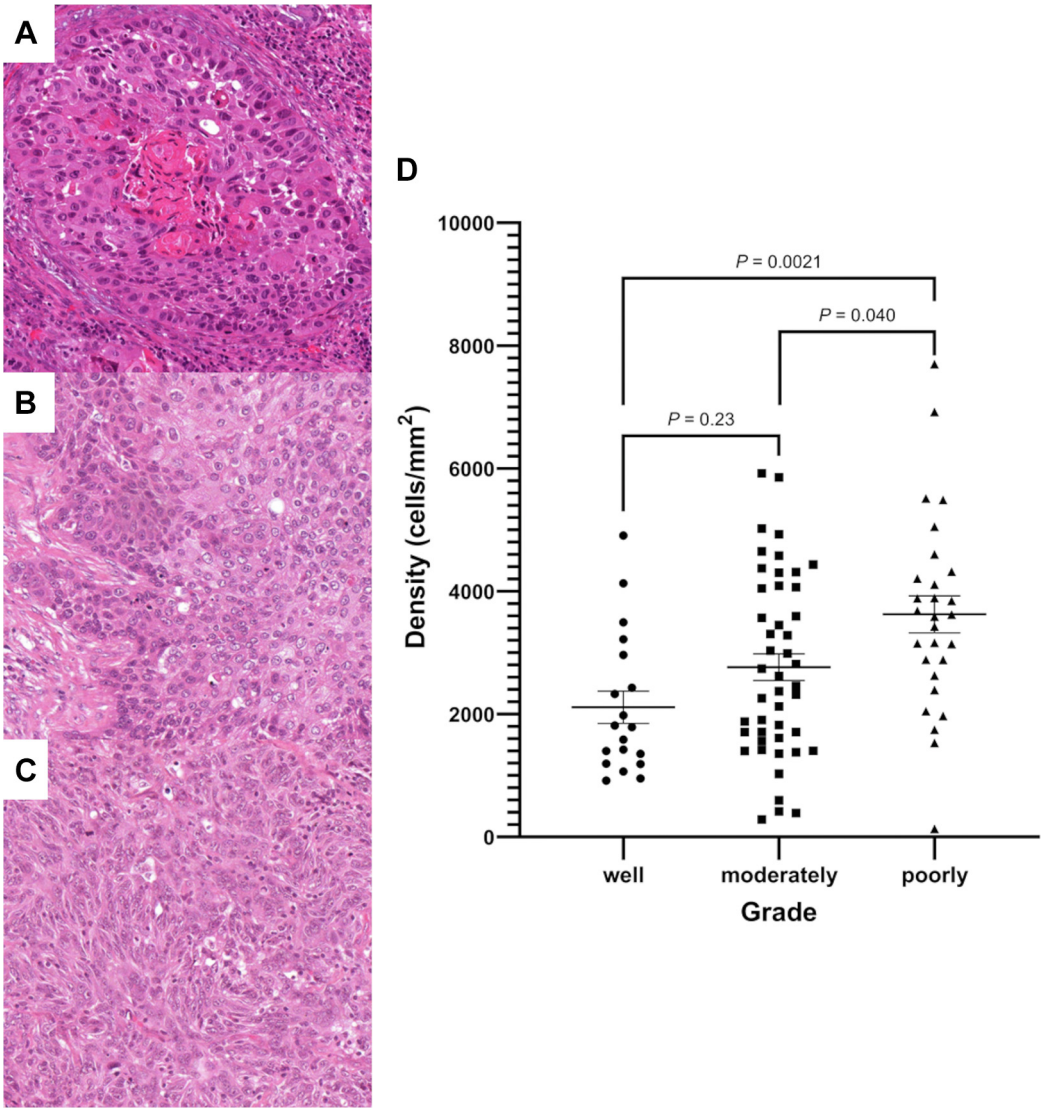


**Figure 3.**
Tumor density in lung squamous cell carcinoma. Representative areas of tumor demonstrating (A) well-differentiated, (B) moderately differentiated, and (C) poorly differentiated squamous cell carcinoma. (D) Distinct cell densities were observed in various histologic patterns of lung squamous cell carcinoma.

compared with that in low-grade patterns and well-differentiated tumors, respectively. Our system of tumor cellularity can provide a composite measure that captures both tumor stage and grade. Further, because our methodology is not based on subjective interpretation of histologic growth patterns but rather uses simple objective quantification of tumor cells, its incorporation into reporting may reduce interobserver variability. Surgical collapse has also been shown to interfere with the assessment of architectural patterns and the degree of invasion.[14,15] Quantification of the absolute number of tumor cells normalized to a specific surface area should conceptually minimize the effect of biases due to surgical collapse.

We observed a range of cellular density in the complex glandular group, with variation in both the fused gland and cribriform subgroups (Fig. 2). The current study was underpowered to quantify the difference between these subpatterns of complex glandular formations. However, it is an important question to be addressed in larger studies to determine whether all patterns captured in the fused gland definition of complex glandular are equally associated with cellularity and outcomes in patients with lung ADC.

We found that the overall cellularity was able to provide prognostic information about NSCLC, with improved overall and progression-free survival in patients with low cellularity. The difference in outcomes was most significant in patients with SCC, with a median overall survival of 25 months in those with tumors with high cellularity compared with 64 months in those with tumors with low cellularity. A significant difference was also observed in the ADC data set, with a median overall survival of 34 months in those with tumors with high cellularity compared with 49 months in those with tumors with low cellularity. SCCs often have large areas of necrosis and fibrosis associated with the tumors, which are typically measured as part of the overall size of the tumors. By utilizing the measurement of tumor cellularity, we may be better able to control for this variability and provide a more accurate means of predicting outcomes.

Our preliminary inflection point served as an initial means to develop a rational cutoff between low- and high-cellularity

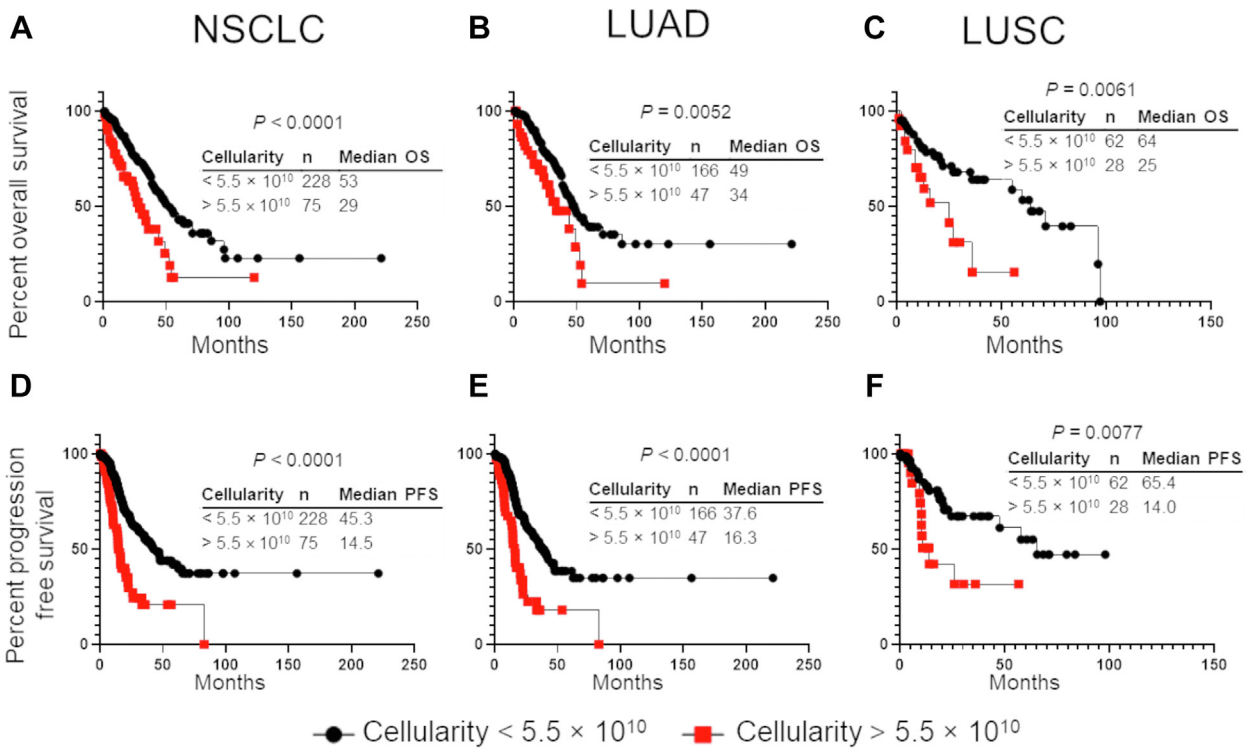


**Figure 4.**
Outcomes of cases based on calculated cellularity. Kaplan-Meier curves for (A-C) overall survival and (D-F) progression-free survival comparing low ($<5.5 \times 10^{10}$) versus high ($>5.5 \times 10^{10}$) cellularity. Data shown are for (A, D) all non–small cell lung carcinomas as well as subdivided into (B, E) lung adenocarcinoma and (C, F) squamous cell carcinoma. Median overall survival or progression-free survival, sample sizes, and *P* values from the log-rank analysis are shown in their respective figures. LUAD, lung adenocarcinoma; NSCLC, non–small cell lung carcinoma; OS, overall survival; PFS, progression-free survival.

tumors. We performed an additional stepwise analysis using 4 quartiles of cellularity (Table 3). We found that our initial inflection point from the scatterplot fell within the third and fourth quartiles of cellularity and was significantly different for overall survival ($P = .049$) and trended to significance for progression-free survival (Table 3). Larger, external validation studies are required to further refine the cutoff point between groups of low- and high-cellularity tumors and integrate other elements of T staging, such as pleural invasion, into a potentially revised TNM classification system.

We were able to efficiently apply our semiautomated approach to cases using standard computers and free, open-source tools (QuPath). Using this system, we needed to retrain the object classifier among the cases. This is due, in part, to the variation in hematoxylin and eosin staining between the different centers utilized in the TCGA data sets. However, many new algorithms are being developed by both academic and commercial laboratories that have demonstrated the ability to reliably detect NSCLCs across a range of staining with minimal training.[16,17] These platforms could be utilized to generate a completely automated workflow.

There are limitations to the current study, with many related to the use of the TCGA slides. Although this data set provides a resource to test novel hypotheses, there are limited data on the treatment of patients. Further, the cases span a range of timelines, and many patients would have been treated with regimes that would be outside the current standard of care. Notably, most of the cases predate the routine use of immunotherapy, which has revolutionized the treatment of this disease.[18] Further, only 1 histology slide was available for the cases, and it is conceivable that this may not have been representative of the overall cellularity of the tumors. The majority of the tumors were <5 cm (Table 4); however, it is possible that in larger tumors, the measure based on a single slide was not reflective of the overall cellularity. Ongoing work that is beyond the scope of this study seeks to validate and refine our findings of tumor cellularity as a novel prognostic feature in more recent data sets treated with defined therapies with all slides available for assessment.

As we increasingly move toward digital workflows in pathology, this study highlights the potential of digital image analysis tools to produce valuable data that cannot be generated using traditional analog microscopes. Using these data, we were able to highlight tumor cellularity as a novel prognostic factor for NSCLCs. Our pipeline uses all open-source, freely available software, and we have made our methodology for this approach universally available. We are also currently working to apply this technique to smaller images of tumors that can be taken using a standard camera or smartphone mounted to a microscope. We anticipate that this approach will allow for the utilization of this methodology in laboratories with varying funding and infrastructure levels across the globe. We hope that novel approaches based on image analysis of tumors can be integrated into future TNM staging systems to better provide more accurate prognostication tools to guide clinical decision making.

## *Author Contributions*

S.L., I.S., and M.J.C. conceptualized the study and performed methodology validation, writing, review, and revision of the paper. S.L., S.K., B.B., D.T., J.M., J.L., M.P., N.W., T.M., and V.L. performed

contouring of tumor annotations. I.S. and M.J.C. reviewed contours. J.P.S. and E.S. performed analysis and interpretation of data and statistical analysis. All authors read and approved the final paper.

### *Data Availability*

The data sets used and/or analyzed for this study are available from the corresponding author upon reasonable request.

### *Funding*

Funding support is gratefully acknowledged from the Megan J. Davey Opportunity Fund used to support a workstation to facilitate the computational analysis in this study and travel costs to present an earlier version of this content at the United States & Canadian Academy of Pathology 111th Annual Meeting (2022).

### *Declaration of Competing Interest*

The authors declare that they have no competing interests that are relevant to this work.

### *Ethics Approval and Consent to Participate*

Not applicable. All slides were from the publicly available The Cancer Genome Atlas database.

## Supplementary Material

The online version contains supplementary material available at https://doi.org/10.1016/j.modpat.2022.100055

## References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin*. 2019;69(1):7–34.
2. Amin MB, American Joint Committee on Cancer, American Cancer Society, eds. *AJCC Cancer Staging Manual*. 8th ed. American Joint Committee on Cancer, Springer; 2017:1024.
3. Nicholson AG, Tsao MS, Beasley MB, et al. The 2021 WHO classification of lung tumors: impact of advances since 2015. *J Thorac Oncol*. 2022;17(3):362–387.
4. Moreira AL, Ocampo PS, Xia Y, 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*. 2020;15(10):1599–1610.
5. Solis LM, Behrens C, Raso MG, et al. Histologic patterns and molecular characteristics of lung adenocarcinoma associated with clinical outcome. *Cancer*. 2012;118(11):2889–2899.
6. Wright J, Churg A, Kitaichi M, Yang HM, Hyde D, Yi E. Reproducibility of visual estimation of lung adenocarcinoma subtype proportions. *Mod Pathol*. 2019;32(11):1587–1592.
7. Thunnissen E, Beasley MB, Borczuk AC, et al. Reproducibility of histopathological subtypes and invasion in pulmonary adenocarcinoma. An international interobserver study. *Mod Pathol*. 2012;25(12):1574–1583.
8. Mlika M, Helal I, Braham E, Ayadi A, Mrabet A, Mezni F. The 2015 classification of lung adenocarcinomas: reproducibility in a Tunisian department specialised in thoracic pathology. *Ann Pathol*. 2017;37(6):467–471.
9. Warth A, Stenzinger A, von Brünneck AC, et al. Interobserver variability in the application of the novel IASLC/ATS/ERS classification for pulmonary adenocarcinomas. *Eur Respir J*. 2012;40(5):1221–1227.
10. Goldstraw P, Chansky K, Crowley J, et al. The IASLC lung cancer staging project: proposals for revision of the TNM stage groupings in the forthcoming (eighth) edition of the TNM classification for lung cancer. *J Thorac Oncol*. 2016;11(1):39–51.
11. Bankhead P, Loughrey MB, Fernández JA, et al. QuPath: open source software for digital pathology image analysis. *Sci Rep*. 2017;7(1):1–7.
12. Moreira AL, Joubert P, Downey RJ, Rekhtman N. Cribriform and fused glands are patterns of high-grade pulmonary adenocarcinoma. *Hum Pathol*. 2014;45(2):213–220.
13. Hwang JK, Page BJ, Flynn D, et al. Validation of the eighth edition TNM lung cancer staging system. *J Thorac Oncol Off Publ Int Assoc Study Lung Cancer*. 2020;15(4):649–654.
14. Blaauwgeers H, Radonic T, Lissenberg-Witte B, et al. P06. 02 incorporating surgical collapse in the pathological assessment of resected adenocarcinoma in situ of the lung. A proof of principle study. *J Thorac Oncol*. 2021;16(10):S985.
15. Thunnissen E, Motoi N, Minami Y, et al. Elastin in pulmonary pathology: relevance in tumours with a lepidic or papillary appearance. A comprehensive understanding from a morphological viewpoint. *Histopathology*. 2022;80(3):457–467.
16. Sikpa D, Fouquet JP, Lebel R, Diamandis P, Richer M, Lepage M. Automated detection and quantification of breast cancer brain metastases in an animal model using democratized machine learning tools. *Sci Rep*. 2019;9(1):1–8.
17. Wang S, Wang T, Yang L, et al. ConvPath: a software tool for lung adenocarcinoma digital pathological image analysis aided by a convolutional neural network. *EBioMedicine*. 2019;50:103–110.
18. Shields MD, Marin-Acevedo JA, Pellini B. Immunotherapy for advanced non-small cell lung cancer: a decade of progress. *Am Soc Clin Oncol Educ Book*. 2021;41:e105–e127.