



Reproducibility of visual estimation of lung adenocarcinoma subtype proportions

Joanne Wright¹ · Andrew Churg¹ · Masanori Kitaichi² · Hui-Min Yang¹ · Dallas Hyde³ · Eunhee Yi⁴

Received: 11 March 2019 / Revised: 22 May 2019 / Accepted: 22 May 2019 / Published online: 23 June 2019
© The Author(s), under exclusive licence to United States & Canadian Academy of Pathology 2019

Abstract

In 2011, a new classification of lung adenocarcinoma was published. Since then there have been multiple studies regarding observer variability in predominant subtype determination, with levels of agreement generally in the weak to moderate range. In the 2011 and subsequent WHO 2015 classification, a recommendation was also made to visually assess and record the percentage of each subtype in 5% increments. The present study was initiated to determine the reproducibility of such gestalt assessments and to compare these data to a formal morphometric assessment. Five experienced pathologists reviewed multiple single images of 25 adenocarcinomas, taken at 2× and 10×, and estimated the percentage of lepidic, acinar, papillary, micropapillary, and solid components in 5% increments. After 2 months all the pathologists again reviewed the same images presented to them in a different order. We found that there was poor reproducibility within observers at 2× power using a 5% evaluation, but that this improved using 10% or 25% cutoffs. Use of 10× magnification allowed weak to moderate reproducibility at 5% increments, and this was again improved using 10% or 25% cutoffs. Correlation with morphometric assessment was poor except for the papillary and micropapillary subtypes. Differences among pathologists were generally low except for the acinar and, to a lesser degree, lepidic subtypes, which showed a wide spread of data. When estimating tumor subtype proportions, use of a 10× objective, and utilization of 10% or preferably 25% cutoffs provides a greater degree of consistency than a 5% cutoff.

Introduction

The International Association of the study of lung cancer 2011 classification of pulmonary adenocarcinoma [1] recommended that a diagnosis of pulmonary adenocarcinoma should not only include a listing of the predominant pattern, but that the percentage of each subtype should be visually be estimated in increments of 5%. The 2015 WHO monograph [2] also makes this recommendation and

reiterates the importance of recognition of minor subtypes. However, even reproducibility of determination of the predominant subtype has shown only weak to moderate agreement among pathologists [3–6]. Although it may be important to recognize the presence of minor subtypes, there have been little data to indicate that such a practice, and in particular use of a 5% increment, is reproducible within or among pathologists. The present study has been designed to investigate this issue.

Methods

Ethics approval was received from the University of British Columbia and the Vancouver General Hospital. Twenty-five invasive, nonmucinous adenocarcinomas of lung were chosen from the surgical pathology archives of Vancouver General Hospital. There were no minimally invasive adenocarcinomas in the study. From each case, two slides containing tumor were selected and photographs taken for evaluation. Five experienced pathologists were included in the study. In order to reduce training bias, none

✉ Joanne Wright
jlwright@mail.ubc.ca

¹ Departments of Pathology, St Paul's Hospital, Vancouver General Hospital, University of British Columbia, Vancouver, BC, Canada

² Department of Pathology, National Hospital Organization Minami Wakayama Medical Center, Tanabe City, Wakayama Prefecture, Japan

³ California National Primate Research Center, Davis, CA, USA

⁴ Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA

of the pathologists, who work in three separate countries, trained together, nor did any one pathologist train any of the others.

Definitions for each tumor pattern were first reviewed using the criteria set out in the summary article by Travis et al. [7]. In the first experiment, each observer examined two photographs of each case taken using a 2× objective with a field diameter of 10.5 mm. The fields to be photographed were selected using a number from the top left hand X- and Y-axes of the microscope stage obtained from a random number generator. The 50 photographs were arranged randomly. Each pathologist provided an overall visual percentage estimation, in 5% increments, of the proportions of the different variants of adenocarcinoma: lepidic, acinar, papillary, solid, and micropapillary. Percentage of the variant types was calculated using the combined data for each case.

In the second part of the study, each slide was placed on the microscope stage and co-ordinates around the tumor were obtained on the X- and Y-axes of the stage encompassing the area previously photographed at 2× power. Five sets of nonoverlapping co-ordinates were obtained for each case using stratified random sampling by taking the total number of fields on the slide divided by 5 to get the sampling period; a random number was used within the first period to select the starting point and then each additional field was incremented by the sampling period. These areas were then identified using a 10× objective magnification (field diameter 2.1 mm), photographed, and the randomly arranged 125 images supplied to each of the participating pathologists who again made a visual estimate of the percent of each histologic subtype. Percentage of the variant types was calculated using the combined data for each case.

After 2 months, the images were recoded and the process was repeated. The code was then broken, and data were arranged to determine the final percentages of each subtype for each case as indicated by each pathologist.

In a third part of the study, using a morphometric grid with 35 equally spaced points, one pathologist (J.L.W.) utilized the 10× photographs and categorized the subtypes at each point. These data were then combined for each case and analyzed as to percent of subtype in each case.

As a subsidiary question, the data were examined to determine agreement in the predominant pattern of the adenocarcinomas.

Statistical consultation was obtained prior to data analysis (see Discussion). Using the STATA statistical analysis system (StataCorp LLC USA), we determined the intrapathologist agreement for each pathologist on each of the two data sets by dividing these data into 5, 10, and 25% bins using the Kappa analysis function. Agreement between the morphometric data and the visual scoring was analyzed only for JLW. Agreement in predominant pattern was

assessed by Cohen's kappa analysis. An agreement of less than 40% was judged as minimal, 40–59% weak, 60–79% moderate, and 80–100% strong. The basis of these terms is derived from kappa usage.

To demonstrate the degree of inter-pathologist agreement, we utilized the second set of scores in each of the data sets and calculated all of the differences for each of the 25 cases. Specifically the data from each pathologist was subtracted from those of each of the other pathologists. These data were then graphed as histogram (see Discussion).

Results

Intrapathologist agreement

These data are shown in Tables 1 and 2 for 2× and 10× photographs. On low power evaluation (2× objective), a 5% cutoff provided only minimal to moderate agreement. This improved by approximately 10% using 10% bins, and an additional 10–20% using 25% bins. Agreement was best for papillary and micropapillary subtypes and worst for acinar subtype. We tested for an “outlier” pathologists, but no one pathologist had consistently different responses or less reproducibility than any other.

Analysis of the 10× photographs (Table 2) showed a generally weak to moderate degree of agreement using 5% cutoffs, which improved approximately 10% using 10% bins. Additional improvement in agreement was obtained using 25% bins for the lepidic, acinar, and solid subtypes.

Table 3 shows agreement between 2× and 10× analysis of photographs. Using 5% cutoff bins, there was only weak

Table 1 Intrapathologist agreement at 2× power using various percentage cutoffs

	5% Bins (%)	10% Bins (%)	25% Bins (%)
Lepidic	37	47	69
Acinar	18	26	46
Papillary	67	79	92
Micropapillary	66	74	85
Solid	48	53	70

Table 2 Intrapathologist agreement at 10× power using various percentage cutoffs

	5% Bins (%)	10% Bins (%)	25% Bins (%)
Lepidic	58	67	82
Acinar	22	29	61
Papillary	78	90	
Micropapillary	66	80	
Solid	48	56	65

Table 3 Intrapathologist agreement between 2× and 10× power using various percentage cutoffs

	5% Bins (%)	10% Bins (%)	25% Bins (%)
Lepidic	35	44	71
Acinar	15	24	45
Papillary	67	77	
Micropapillary	65	78	
Solid	48	56	70

Table 4 Agreement with morphometric analysis using various percentage cutoffs

	5% Bins (%)	10% Bins (%)	25% Bins (%)
Lepidic	40	44	72
Acinar	4	24	44
Papillary	88	88	
Micropapillary	72	72	
Solid	48	56	64

agreement for subtypes except for papillary and micropapillary which showed a moderate degree of agreement. This improved only slightly using 10% bins, but there was an approximately 30% increase in agreement using 25% cutoffs; papillary and micropapillary subtypes were all within a 25% range so additional improvement was not obtained for these subtypes.

Agreement with morphometric analysis

Table 4 shows the agreement between visual and morphometric scoring. Acinar subtype had an extremely weak degree of agreement at 5% cutoff which improved only to a moderate level even at 25% cutoff bins. Papillary and micropapillary subtypes had a reasonably strong degree of agreement even at 5% bins. Lepidic and solid subtypes obtained strong agreement only with 25% bin cutoffs.

Differences among pathologists

Figures 1–5 show histograms of the differences among pathologists for each subtype in both 2× and 10× photograph analysis. With a 2× objective, for all subtypes except acinar, differences among pathologists were generally modest, but there were also some very large differences. These differences were especially marked for the acinar variant and remained even with a 10× objective (see Discussion). Considerable differences were present for the lepidic variant even with 10× objective (see Discussion). However, papillary and micropapillary variants showed only small degrees of differences at 10×, and differences were diminished in the solid subtype. Agreement in predominant pattern was weak at both low (kappa 0.35) and high (kappa 0.25) objective powers. At low

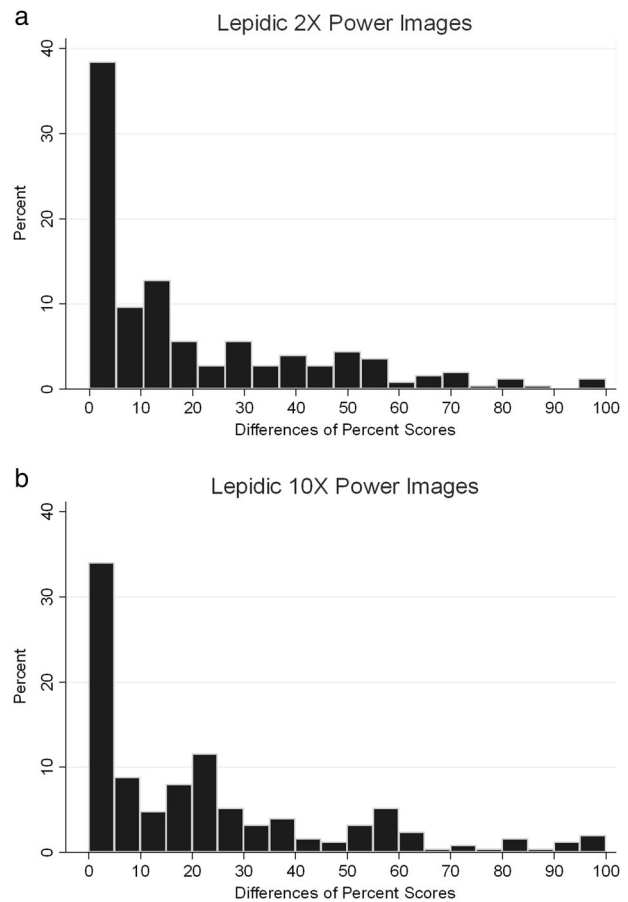


Fig. 1 Differences in percent scores of all pathologists for lepidic pattern at 2× (a) and 10× (b) objective powers

power, the discrepancies are most obvious between acinar and solid patterns. In five cases one or more pathologists had a discordant predominant pattern between these two subtypes. Similar findings regarding discrepancies between solid and acinar pattern were present at high power, but there were also variations in lepidic and acinar pattern with a single (but not always the same) pathologist differing in 6 six cases.

Discussion

Since the report of the 2011 IASLC pulmonary adenocarcinoma classification [1], a number of articles have been published regarding assessment of patterns [7], and the reproducibility of such assessments [3–5]. Reproducibility of determination of the predominant pattern on H&E stains ranges from a kappa of 0.25–0.77 using H&E only, findings recapitulated in the present study. A recent study [6] has suggested that this can be improved using a combination of immunohistochemistry and mucin stains. The studies by Thunnissen et al. [3] and Warth et al. [4, 5] noted that determination of a predominant pattern was easier for

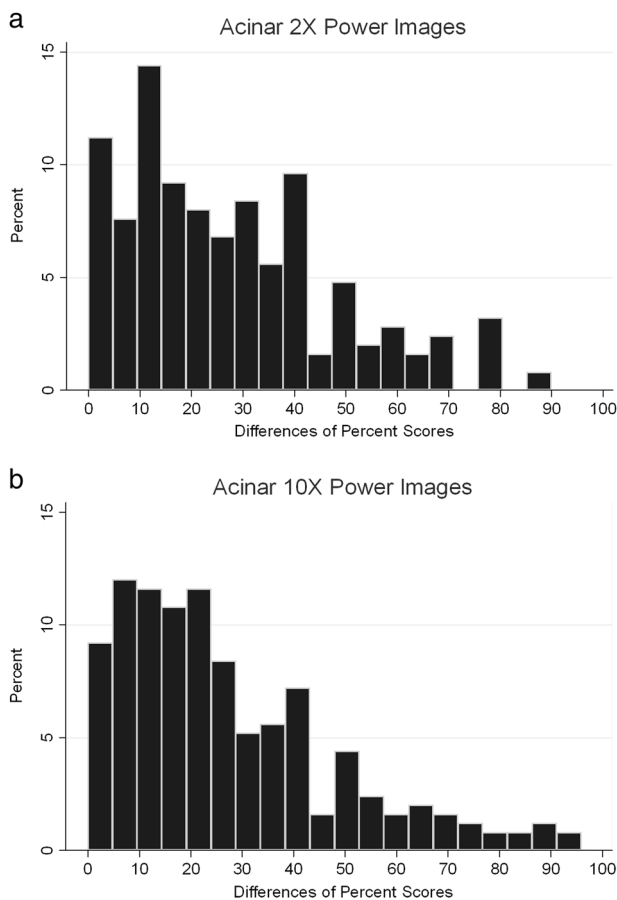


Fig. 2 Differences in percent scores of all pathologists for acinar pattern at 2× (a) and 10× (b) objective powers

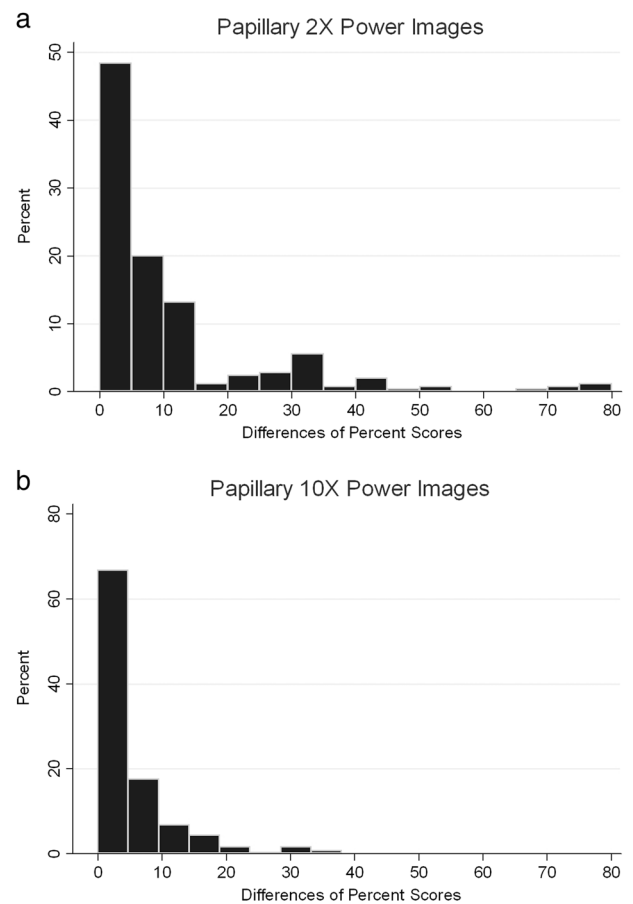


Fig. 3 Differences in percent scores of all pathologists for papillary pattern at 2× (a) and 10× (b) objective powers

lepidic and acinar patterns, and most difficult for papillary and micropapillary patterns, however, our study showed a better degree of agreement on the proportion of these latter two patterns indicating that they are less prone to an alternate interpretation (see below). However, since these patterns were the predominant pattern in only a minority of our samples, we cannot make this assertion with exactitude. Our study suggests that determination of predominant pattern is most difficult between acinar and solid patterns, a difference which may reflect the difficulties in assigning a major pattern to cribriform variant.

The determination of patterns appears to be important prognostically since the different histological subtypes conferred different patient outcomes (reviewed in [8]. For example, it has been recently shown [9] that micropapillary and solid variants had a greater recurrence rate and a poorer overall survival when they were the predominant patterns, and even a small amount of micropapillary pattern appears to have a negative effect upon overall survival [10].

In the original 2011 classification [1] a recommendation was introduced to provide a gestalt percentage of the types of patterns using 5% increments, although it was also noted

that this recommendation was weak and based on low-quality evidence. Our present study was performed to determine whether such a visual process was reproducible within and among pathologists at either a low (2×) or medium (10×) objective, and whether the results would correlate with a morphometric determination.

Our data show that intrapathologist agreement is weak using a 2× objective and better using a 10× objective analysis. There is a generally weak to moderate degree of agreement using the recommended 5% cutoffs; agreement was best for the papillary and micropapillary variants. However, agreement improved considerably using 10% bins and generally improved further using 25% bins. Surprisingly, we found that the least degree of agreement was present for the acinar and lepidic subtypes, suggesting that, on individual fields, it may be difficult to differentiate acinar subtype from lepidic; this issue probably arises from differences in whether one views areas with acinar glands and surrounding scars as lepidic tumor (i.e., Noguchi type B) or invasive acinar adenocarcinoma. This problem may, theoretically, be accentuated when assessing minimally invasive tumors; this is beyond the scope of the present

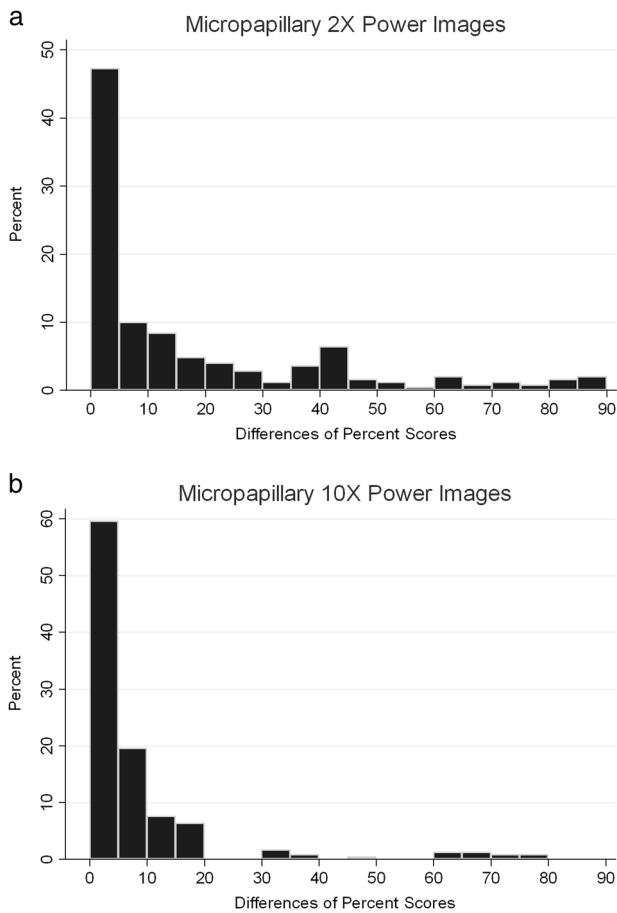


Fig. 4 Differences in percent scores of all pathologists for micropapillary pattern at 2× (a) and 10× (b) objective powers

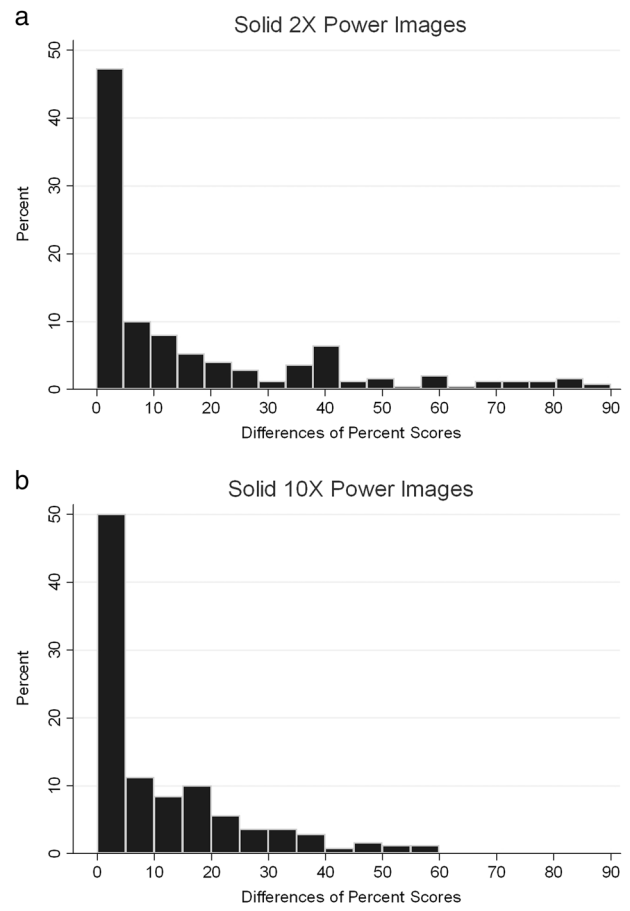


Fig. 5 Differences in percent scores of all pathologists for solid pattern at 2× (a) and 10× (b) objective powers

study, since we did not include minimally invasive tumors in the analysis. Acinar, and to a lesser extent lepidic variants also had the greatest degree of difference among the pathologists.

One could ask why we chose to utilize percent agreement tables and histograms to display our data rather than the commonly utilized kappa statistic. For intrapathologist agreement, it is important to recognize that at 5% cutoffs, there were 20 potential bins and at 10% there were 10 bins, and therefore percent agreement avoids problems related to data skew. The other advantage of the histograms is that they show degrees of difference among pathologist and are thus able to summarize an easy-to-interpret quantity (how (dis-)similar are the two observations), whereas the kappa value is an abstraction that does not provide information on how often observations are similar.

Our study results are not surprising when considering that reproducibility of dominant type only shows only a weak-to-moderate agreement. However, it does indicate that subtype fraction analysis is possible, but is probably best performed with 10–25% bin cutoffs, with the greatest accuracy using 25% bins. A study by Smits et al. [11] has

also found very low reliability when determining the percentages of tumor population for molecular analysis using 5% cutoffs in tissue samples. One might ask whether formal morphometric analysis would be more accurate, but morphometric analysis is not generally appropriate for other than research studies, and does not provide any additional information. Indeed the same difficulty in definitively determining acinar subtype appears to be present here using morphometric analysis.

In summary, this study has shown that determination of lung adenocarcinoma patterns according to the percentage of each subtype improves with larger cutoffs. Use of 5% cutoffs produces only weak to moderate reproducibility, and 10% or 25% subtype estimates appear to provide more accurate data.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger KR, Yatabe Y, et al. International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society International Multidisciplinary Classification of Lung Adenocarcinoma. *J Thorac Oncol*. 2011;6:244–85.
2. Travis WD, Noguchi M, Yatabe Y, Brambilla E, Nicholson AG, Aisner SC, et al. Adenocarcinoma. In: Travis WD, Brambilla E, Burke AP, Marx A, Nicholson AG, editors. WHO classification of tumors of the lung, pleura, thymus and heart, 4th edition, Vol. 7 Lyons, France: International Agency for Research on Cancer (IARC); 2015. p. 32.
3. Thunnissen E, Beasley MB, Borczuk AC, Brambilla E, Chirieac LR, Dacic S, et al. Reproducibility of histopathological subtypes and invasion in pulmonary adenocarcinoma. An international interobserver study. *Mod Pathol*. 2012;25:1574–83.
4. Warth A, Stenzinger A, von Brunnecke A-C, Goeppert B, Cortis J, Petersen I, et al. Interobserver variability in the application of the novel IASLC/ATS/ERS classification for pulmonary adenocarcinomas. *Eur Respir J*. 2012;40:1221–7.
5. Warth A, Cortis J, Fink L, Fisseler-Eckhoff A, Geddert H, Hager T, et al. Training increases concordance in classifying pulmonary adenocarcinomas according to the novel IASLC/ATS/ERS classification. *Virchows Arch*. 2012;461:185–93.
6. Funkhouser WK, Hayes DN, Moore DT, Funkhouser WK III, Fine JP, Jo HJ, et al. Interpathologist diagnostic agreement for non-small cell lung carcinomas using current and recent classifications. *Arch Pathol Lab Med*. 2018;142:1537–48.
7. Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger KR, Yatabe Y, et al. Diagnosis of lung adenocarcinoma in resected specimens. *Arch Pathol Lab Med*. 2013;137:685–705.
8. Borczuk AC. Prognostic considerations of the new World Health Organization classification of lung adenocarcinoma. *Eur Respir Rev*. 2016;142:364–71.
9. Hung J-J, Yeh Y-C, Jeng W-J, Wu K-J, Huang B-S, Yu-Chung Wu Y-C, et al. Predictive value of the International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society Classification of Lung Adenocarcinoma in Tumor Recurrence and Patient Survival. *J Clin Oncol*. 2014;32:2357–64.
10. Lee G, Lee HY, Jeong JY, Han J, Cha MJ, Lee KS, et al. Clinical impact of minimal micropapillary pattern in invasive lung adenocarcinoma: prognostic significance and survival outcomes. *Am J Surg Pathol*. 2015;39:660–6.
11. Smits AJJ, Jummer JA, de Bruijn PC, Bol M, van den Tweel JG, Seldenrijk KA, et al. The estimation of tumor cell percentage for molecular testing by pathologists is not accurate. *Mod Pathol*. 2014;27:168–74.