1574

Reproducibility of histopathological subtypes and invasion in pulmonary adenocarcinoma. An international interobserver study

Erik Thunnissen[1], Mary Beth Beasley[2], Alain C Borczuk[3], Elisabeth Brambilla[4], Lucian R Chirieac[5], Sanja Dacic[6], Douglas Flieder[7], Adi Gazdar[8], Kim Geisinger[9], Philip Hasleton[10], Yuichi Ishikawa[11], Keith M Kerr[12], Sylvie Lantejoul[4], Yoshiro Matsuno[13], Yuko Minami[13], Andre L Moreira[14], Noriko Motoi[11], Andrew G Nicholson[15], Masayuki Noguchi[13], Daisuke Nonaka[16], Giuseppe Pelosi[17], Iver Petersen[18], Natasha Rekhtman[14], Victor Roggli[19], William D Travis[14], Ming S Tsao[20], Ignacio Wistuba[21], Haodong Xu[22], Yasushi Yatabe[23], Maureen Zakowski[14], Birgit Witte[24] and Dirk Joop Kuik[24]

[1]*Department of Pathology, VU University Medical Centre, Amsterdam, The Netherlands;* [2]*Mount Sinai Hospital, New York, NY, USA;* [3]*Department of Pathology, Anatomic Pathology Columbia University Medical Center, New York, NY, USA;* [4]*Department of Pathology, CHU A Michallon, INSERM U 823-Institut A Bonniot-University J Fourier, Grenoble, France;* [5]*Department of Pathology, Brigham and Women's Hospital, Boston, MA, USA;* [6]*Department of Pathology, University of Pittsburgh, Pittsburgh, PA, USA;* [7]*Department of Pathology, Fox Chase Cancer Center, Philadelphia, PA, USA;* [8]*UT Southwestern Medical Center, Dallas, TX, USA;* [9]*Piedmont Pathology Associates, Hickory, NC, USA;* [10]*Department of Pathology, Hadassah Hospital, Jerusalem, Israel;* [11]*Division of Pathology, The Cancer Institute, Japanese Foundation for Cancer Research, Tokyo, Japan;* [12]*Department of Pathology, Aberdeen Royal Infirmary, Aberdeen, UK;* [13]*Department of Pathology, Faculty of Medicine, University of Tsukuba, Tsukuba, Japan;* [14]*Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA;* [15]*Department of Histopathology, Royal Brompton Hospital, London, UK;* [16]*Department of Histopathology, The Christie NHS Foundation, Manchester, UK;* [17]*Department of Pathology and Laboratory Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori and Università degli Studi di Milano, Milan, Italy;* [18]*Institute of Pathology, Jena University Hospital, Friedrich-Schiller-University, Jena, Germany;* [19]*DUMC 3712, Durham, NC, USA;* [20]*Department of Pathology, University Health Network-Princess Margaret Hospital and University of Toronto, Toronto, Canada;* [21]*Department of Pathology, University of Texas MD Anderson Cancer Center, Houston, TX, USA;* [22]*Department of Pathology and Laboratory Medicine, Aab Cardiovascular Research Institute, University of Rochester Medical Center, Rochester, NY, USA;* [23]*Department of Pathology and Molecular Diagnostics, Aichi Cancer Center, Nagoya, Japan and* [24]*Department of Epidemiology and Biostatistics, VU University Medical Centre, Amsterdam, The Netherlands*

Histological subtyping of pulmonary adenocarcinoma has recently been updated based on predominant pattern, but data on reproducibility are required for validation. This study first assesses reproducibility in subtyping adenocarcinomas and then assesses further the distinction between invasive and non-invasive (wholly lepidic) pattern of adenocarcinoma, among an international group of pulmonary pathologists. Two ring studies were performed using a micro-photographic image-based method, evaluating selected images of lung adenocarcinoma histologic patterns. In the first study, 26 pathologists reviewed representative images of typical and 'difficult' histologic patterns. A total number of scores for the typical patterns combined ($n = 94$) and

Correspondence: Dr E Thunnissen, MD, PhD, Department of Pathology, VU University Hospital, De Boelelaan 1117, 1081 HV Amsterdam, The Netherlands.
E-mail: e.thunnissen@vumc.nl

the difficult cases ($n = 21$) were 2444 and 546, respectively. The mean kappa score ($\pm$ s.d.) for the five typical patterns combined and for difficult cases were 0.77 ± 0.07 and 0.38 ± 0.14, respectively. Although 70% of the observers identified 12–65% of typical images as single pattern, highest for solid and least for micropapillary, recognizing the predominant pattern was achieved in 92–100%, of the images except for micropapillary pattern (62%). For the second study on invasion, identified as a key problem area from the first study, 28 pathologists submitted and reviewed 64 images representing typical as well as 'difficult' examples. The kappa for typical and difficult cases was 0.55 ± 0.06 and 0.08 ± 0.02, respectively, with consistent subdivision by the same pathologists into invasive and non-invasive categories, due to differing interpretation of terminology defining invasion. In pulmonary adenocarcinomas with classic morphology, which comprise the majority of cases, there is good reproducibility in identifying a predominant pattern and fair reproducibility distinguishing invasive from *in-situ* (wholly lepidic) patterns. However, more precise definitions and better education on interpretation of existing terminology are required to improve recognition of purely *in-situ* disease, this being an area of increasing importance.

The 2004 WHO classification of lung cancer contained four major patterns of adenocarcinoma: bronchioloalveolar, acinar, papillary and solid pattern with the most common pattern consisting of a mixture of these four subtypes.[1] In the recent IASLC/ATS/ERS lung adenocarcinoma classification several major changes are made.[2] First, the mixed subtype category is discontinued and tumors are subtyped according to the predominant pattern following a comprehensive semiquantitatively estimating the percentage of each of the adenocarcinoma histologic patterns. However, evidence for use of predominant patterns to improve reproducibility pattern diagnosis was at the time not available. Therefore, this was put forth as a weak recommendation with low quality of evidence (Pathology Recommendation 4).[1] Second, the term bronchioloalveolar carcinoma (BAC) is no longer used, as BAC was being interpreted in four different ways: (1) adenocarcinoma *in situ*, (2) minimally invasive adenocarcinoma, (3) overtly invasive adenocarcinoma with a lepidic pattern and (4) invasive mucinous adenocarcinoma (formerly mucinous BAC). In addition, micropapillary adenocarcinoma was added as a fifth major pattern due to its association with poor prognosis.[3,4] Diagnostic inconsistencies may originate from difficulties in interpretation due to subjective application of existing criteria.

In the past, the distinction between small-cell and non-small-cell lung cancers has been shown to have high accuracy and reproducibility.[5,6] Also, in resection specimen, accuracy in distinguishing squamous cell carcinoma from adenocarcinoma has been repeatedly demonstrated, even though cases that are difficult to classify by morphology alone may remain in poorly differentiated tumors.[5–10] However, there remains a lack of data on reproducibility in relation to identifying predominant patterns of adenocarcinoma.

The intention of this study was therefore to assess the reproducibility of histopathological subtyping for adenocarcinomas among pulmonary pathologists from three continents, with respect to both 'histologic patterns' and 'invasion'.

Materials and methods

To assess reproducibility of adenocarcinoma subtyping, two ring studies were performed. In the first study, 19 pathologists were asked to submit six cases, consisting of micro-photomicrographs, pasted into a PowerPoint slide. Five of the six cases represent one example each of the five typical histological patterns of adenocarcinoma: acinar, non-mucinous lepidic (formerly BAC), micropapillary, papillary and solid pattern, as perceived by the contributing pathologist.[1] The sixth case was regarded as a difficult case by the contributing pathologist. The PowerPoint slide contained two images: the left sided one very-low-magnification picture (objective 2 to $\times 4$) to represent the general architectural pattern, and the right side one showing a higher magnification (objective $\times 10$, or minimally four sizes of 'normal' alveolar spaces in the longest axis of the image) to highlight the diagnostic area to be evaluated. The assumption was made that assessments would be made only on the high-magnification images. Cases ($n = 115$) were randomized and blinded to participants who classified cases by their dominant pattern; if more than one pattern was recognized, then additional pattern(s) were provided additionally (Table 1).

After the first round, participants felt that difficult cases were centered on the concept of 'invasion' (distinguishing pure lepidic pattern from others), which led to the initiation of a second ring study. For the second study, 10 pathologists submitted photomicrographs in similar manner as before, but to show typical invasion ($n = 20$), no invasion ($n = 20$) and 'problem cases' ($n = 24$). All cases were randomized (JK) and for each case

Table 1 CRF contains the possible categories for diagnosing the case(s)

1	Pathologist study number	[.]
	Diagnosis	[.] one of the numbers 2–71
2	AC pure	
3	BA pure	
4	MP pure	
5	PA pure	
6	SO pure	
7	AC predominant with BA	
8	AC predominant with MP	
9	AC predominant with PA	
10	AC predominant with SO	
11	AC predominant with BA and MP	
12	AC predominant with MP and PA	
13	AC predominant with PA and SO	
14	AC predominant with SO and BA	
15	AC predominant with BA and MP and PA	
16	AC predominant with MP and PA and SO	
17	AC predominant with PA and SO and BA	
18	AC predominant with SO and BA and MP	
19	AC predominant with BA and MP and PA and SO	
20	BA predominant with MP	
21	BA predominant with PA	
22	BA predominant with SO	
23	BA predominant with AC	
24	BA predominant with MP and PA	
25	BA predominant with PA and SO	
26	BA predominant with SO and AC	
27	BA predominant with AC and MP	
28	BA predominant with MP and PA and SO	
29	BA predominant with PA and SO and AC	
30	BA predominant with SO and AC and MP	
31	BA predominant with AC and MP and PA	
32	BA predominant with MP and PA and SO and AC	
33	MP predominant with PA	
34	MP predominant with SO	
35	MP predominant with AC	
36	MP predominant with BA	
37	MP predominant with PA and SO	
38	MP predominant with SO and AC	
39	MP predominant with AC and BA	
40	MP predominant with BA and PA	
41	MP predominant with PA and SO and AC	
42	MP predominant with SO and AC and BA	
43	MP predominant with AC and BA and PA	
44	MP predominant with BA and PA and SO	
45	MP predominant with PA and SO and AC and BA	
46	PA predominant with SO	
47	PA predominant with AC	
48	PA predominant with BA	
49	PA predominant with MP	
50	PA predominant with SO and AC	
51	PA predominant with AC and BA	
52	PA predominant with BA and MP	
53	PA predominant with MP and SO	
54	PA predominant with SO and AC and BA	
55	PA predominant with AC and BA and MP	
56	PA predominant with BA and MP and SO	
57	PA predominant with MP and SO and AC	
58	PA predominant with SO and AC and BA and MP	
59	SO predominant with AC	
60	SO predominant with BA	
61	SO predominant with MP	
62	SO predominant with PA	
63	SO predominant with AC and BA	
64	SO predominant with BA and MP	
65	SO predominant with MP and PA	
66	SO predominant with PA and AC	
67	SO predominant with AC and BA and MP	
68	SO predominant with BA and MP and PA	
69	SO predominant with MP and PA and AC	
70	SO predominant with PA and AC and BA	
71	SO predominant with AC and BA and MP and PA	

The amount of certainty is defined as follows
72 = certain of diagnosis
73 = probable, consult colleague
74 = uncertain, consult colleague

AMOUNT of certainty		[.] one of the numbers 72–74

AC, acinar; BA, bronchioloalveolar; CRF, case report form; MP, micropapillary; PA, papillary; SO, solid.
As it is not excluded that an image may contain in the eye of a 'blinded' reader more than one pattern, all possible perturbations for diagnosis are numbered individually. If two patterns are equally present, please make choice of one of the predominant options.
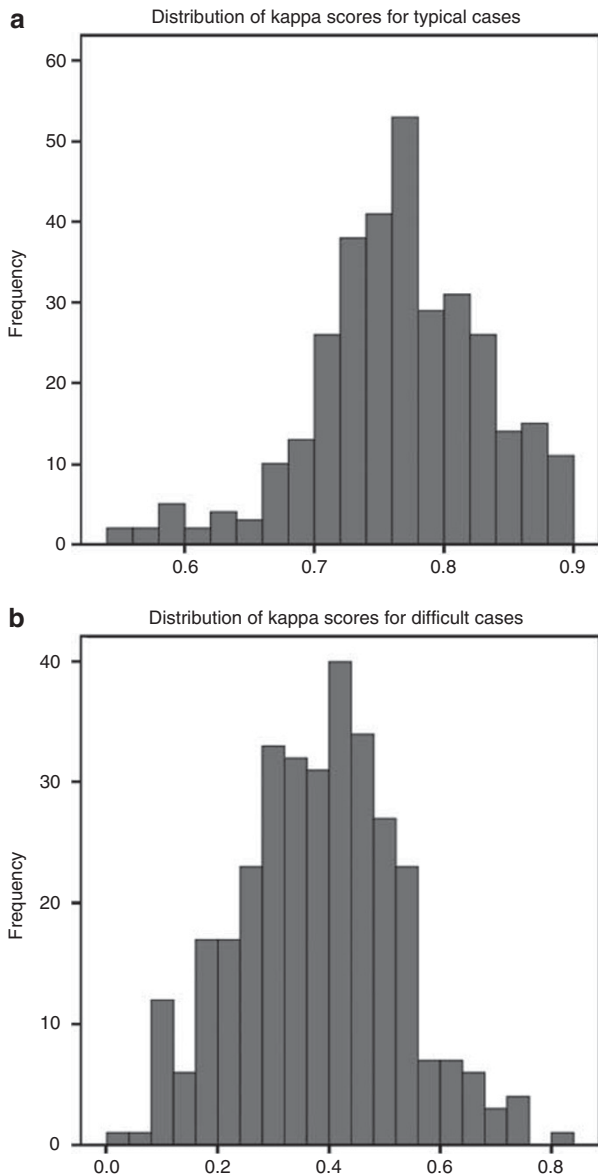
Figure 1 Distribution of kappa scores between all pairs of pathologists for typical cases (**a**) and difficult (**b**) cases.

Figure 2 Box plot distribution of the dominant pattern score (1 = perfect agreement, 0 = no agreement) is shown for the 'typical' patterns according to the submitting pathologist for each of the five histologic subtypes: acinar, lepidic, microp(apillary), papillary and solid. Note that box represents interquartile range (IQR), line in the box is median, wiskers 1.5 × IQR and occasional outliers (o,*) are numbered cases.

Table 2 Cross table of 5 patterns of adenocarcinoma subclassification denoting overlap areas of the diagnostic scores (in percentages) with more than one pattern

Pattern	Acinar (%)	Lepidic (%)	Micropapillary (%)	Papillary (%)	Solid (%)
Lepidic	9				
Micropapillary	6	6			
Papillary	14	10	22		
Solid	15	0	3	2	
Total/pattern	44	25	37	48	20

Table 3 The number of pathologists among 26, who scored at least 70% of the cases correctly as a single pattern or predominant pattern

Submitted pattern	Single pattern (%)	Predominant pattern (%)
Acinar (n = 20)	17/26 (65)	25/26 (96)
Lepidic (BAC; n = 19)	11/26 (42)	24/26 (92)
Micropapillary (n = 16)	3/26 (12)	16/26 (62)
Papillary (n = 19)	5/26 (19)	25/26 (96)
Solid (n = 20)	17/26 (65)	26/26 (100)

the reviewer were asked to provide a 'score' on invasion using five categories: invasion, definite or probable, no invasion, definite or probable, and undetermined.

Statistical Analysis

Kappa score was calculated for the typical cases (separately for five typical patterns and typical invasion) and difficult cases (separately for patterns and invasion) by comparing the scores of submitting pathologist with the 'blind' reading of 26 pathologists. For difficult cases, kappa score was calculated for combinations of all pathologists. For the pattern subtyping a dominant and a subscore were calcu-
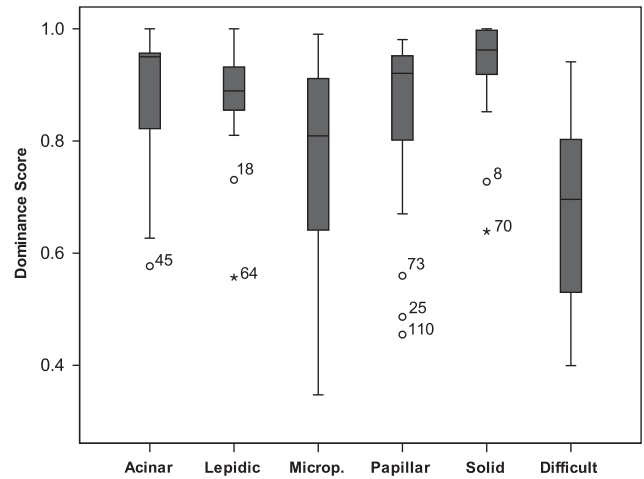
lated for each case and pathologist. The subscore for each subpattern (= non-dominant histologic pattern) was 32, 24, 20 or 16 in case of 1, 2, 3 or 4 subpatterns. The dominant pattern received 512 points minus the number of subpatterns times the subscore. The total score for each pattern was obtained by summing the scores over all pathologists. The overall dominant score for each case was then defined as the highest total score divided by
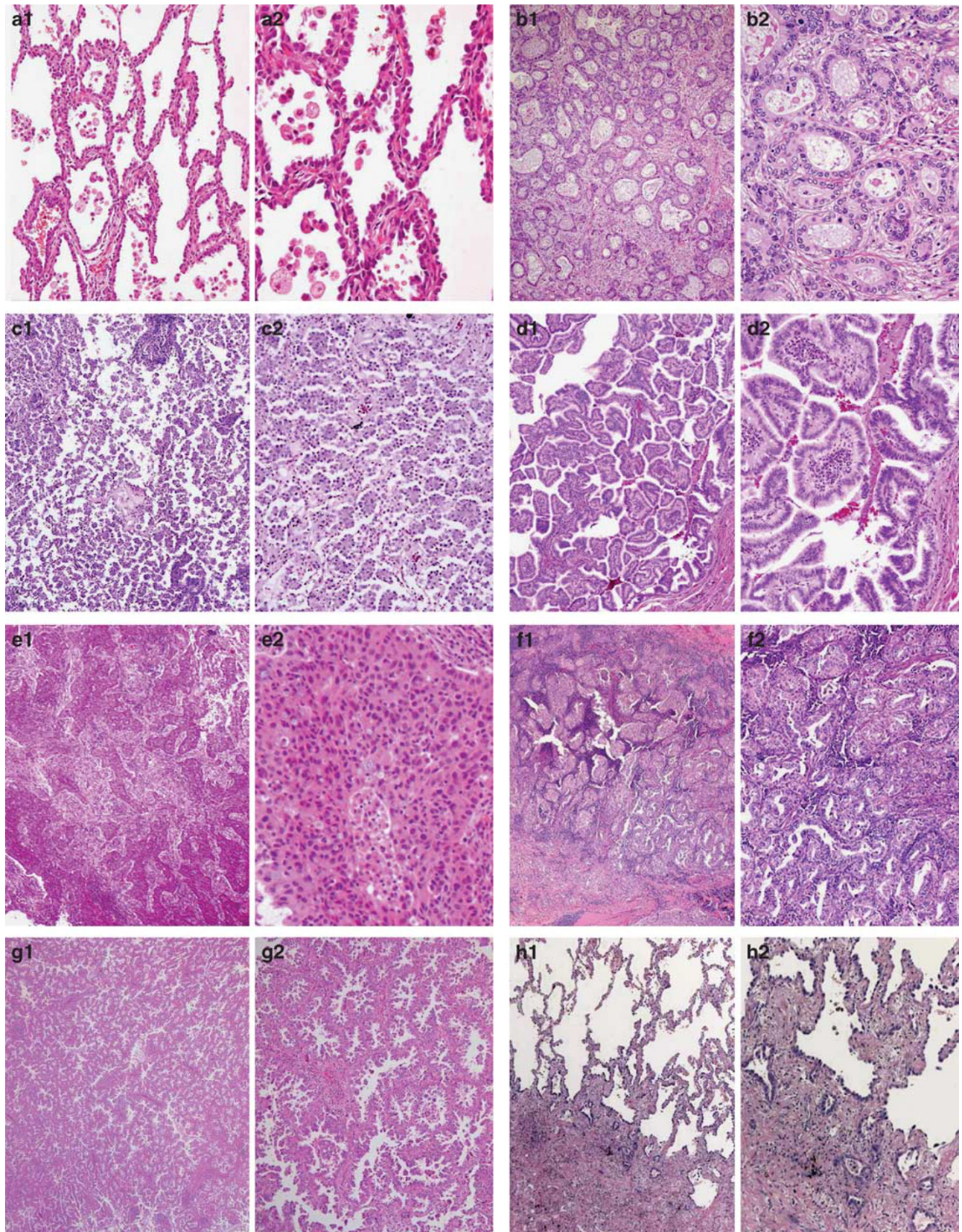
Reproducibility pulmonary adenocarcinoma
1578
E Thunnissen *et al*

Figure 3 Examples of unanimous lepidic (**a**), acinar (**b**), micropapillary (**c**), papillary (**d**) and solid (**e**) pattern are shown, as well as examples of more than one pattern with percent of pathologists recording that pattern (with judgement on second image, **f–j**). Patterns scored by >10% of the pathologists are mentioned. **f**: overlap solid (96%), acinar (93%), micropapillary (15%), papillary (15%); **g**: overlap micropapillary (96%), papillary (11%); **h**: overlap AIS (92%), acinar (78%); **i**: overlap AIS (78%), papillary (54%); **j**: overlap AIS (66%), micropapillary (42%), papillary (30%) and acinar (27%).
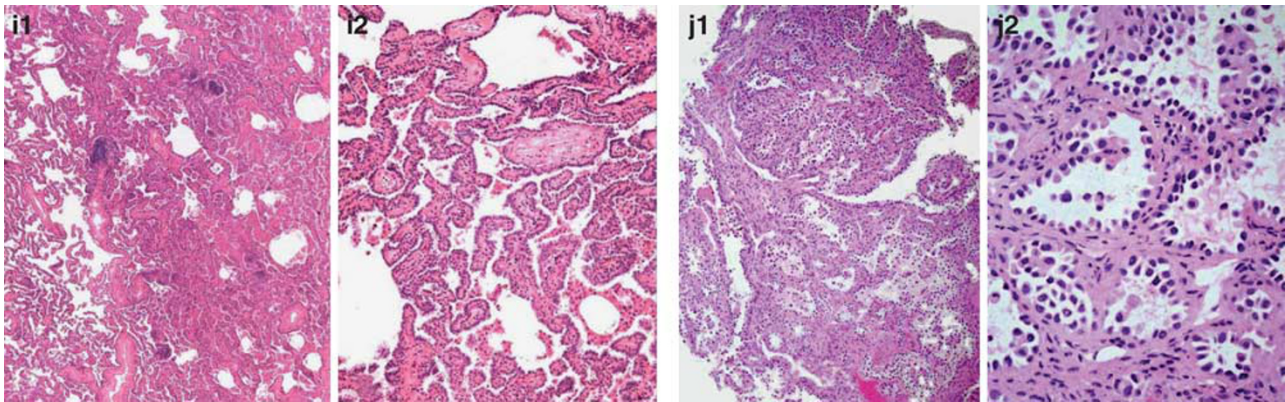
Figure 3 Continued.

the sum of all total scores. For each 'typical' pattern, the number of cases where 70 of the pathologists scored a submitted pattern as single or a predominant pattern were also calculated.

In the second study on 'invasion', kappa was analyzed for five and three categories. For each case an invasion score was calculated: definite invasion = one point; probable invasion = two points; undetermined = three points; probable no-invasion = four points; definite non-invasive = five points.

Results

Pattern Reproducibility

In the first ring study, the reproducibility of pattern classification was performed on 115 cases by 26 pathologists. The 115 cases were as follows: acinar ($n = 20$), lepidic (formerly BAC, $n = 19$), micropapillary ($n = 16$), papillary ($n = 19$), solid ($n = 20$) and difficult cases ($n = 21$). In total, the number of scores for the typical patterns combined and for the difficult cases were 2444 and 546, respectively.

The kappa score (mean and s.d.) calculated between all pathologists for the five typical patterns combined was 0.77 ± 0.07. For the difficult cases, mean and s.d. for Kappa was 0.38 ± 0.14. The distribution of kappa scores for all cases is shown in Figure 1.

Dominant scores were calculated as described before. A dominant score close to 1 is perfect agreement, whereas a score close to 0 corresponds to major disagreement. The distribution of the scores for each pattern is shown in Figure 2. Overall, two patterns have reasonable to good concordance: lepidic and solid. The range of dominant scores for acinar, micropapillary and papillary carcinoma varied to a larger extend. The difficult case category had the lowest average score of agreement.

For the typical cases more than one pattern was recorded in 848 of the 2444 (35%) scores, indicating heterogeneity of adenocarcinoma patterns. In 1048 of the 1205 (87%) diagnostic scores with more

than one pattern, two patterns were scored. Three or more scores were present in 13% of the cases. Call overlap existed between all patterns, except between solid and lepidic pattern. Patterns of overlap in adenocarcinoma sub-classification are shown in Table 2. Perhaps not surprisingly, the highest overlap was noted between papillary and micropapillary patterns.

The concordance rates among 26 pathologists to recognize 70% of the submitted typical patterns as a single pattern ranged from 12–65%, lowest for micropapillary and highest for solid types (Table 3). However, when the submitted pattern was recognized as the predominant pattern (combining single plus multiple patterns with the submitted pattern being the predominant one), the concordance rates between submitted and dominant patterns reached 62–100% (Table 3); four of the patterns scoring $\geq 92\%$ except micropapillary. In general, the pattern of overlap was similar for both submitted 'typical' and 'difficult' cases. Examples of overlap in pattern diagnoses are shown in Figure 3.

Invasion Reproducibility

In the second ring study with an emphasis on reproducibility of invasion, 28 pathologists scored the 64 cases. For each case, an invasion score was calculated. A score of 1 with standard deviation of 0 indicated that all 28 scores were definite invasion and vice versa a score of 5 for a case showed perfect agreement for non-invasion. In Table 4, the distribution of invasion scores is shown. Complete agreement was present in only 6 out of 64 cases, combining the probable and definite categories. In 37 cases, at least five readings differed from the majority score (presence or absence of invasion).

In 15 cases, the scores were equivocal for invasion and non-invasion: at least 9 scores were for invasion and at least 9 against invasion. It turned out that the same pathologists were systematically scoring invasion and another group of

Table 4 The invasion score (mean ± standard deviation, s.d.) is shown for each case, ranked from high (4.93 non-invasive score) to low (1.0 invasive score)

Case #	Invasion score		Scores per category					Total
	Mean	s.d.	1 Inv	2 Pr Inv	3 Undet.	4 Pr-NonI	5 NonInv	
71	4.93	0.26	0	0	0	2	26	28
73	4.93	0.26	0	0	0	2	26	28
11	4.86	0.76	1	0	0	0	27	28
74	4.86	0.76	1	0	0	0	27	28
39	4.64	1.06	2	0	0	2	24	28
3	4.57	1.17	2	1	0	1	24	28
57	4.43	1.35	3	1	0	1	23	28
31	4.21	1.40	3	2	0	4	19	28
34	4.21	1.40	3	2	0	4	19	28
46	4.18	1.54	4	2	0	1	21	28
24	4.00	1.54	4	2	2	2	18	28
7	3.25	1.69	8	3	0	8	9	28
17	3.00	1.70	10	2	1	8	7	28
28	3.00	1.56	8	4	1	10	5	28
66	3.04	1.79	11	1	1	6	9	28
52	3.18	1.81	9	3	2	2	12	28
51	2.89	1.77	11	3	0	6	8	28
64	3.00	1.72	9	4	2	4	9	28
4	2.89	1.77	10	5	0	4	9	28
36	2.82	1.76	11	4	0	5	8	28
53	2.61	1.73	13	3	0	6	6	28
14	2.54	1.69	13	3	2	4	6	28
35	2.61	1.73	12	5	0	4	7	28
55	2.64	1.83	13	4	0	2	9	28
9	2.46	1.71	14	3	1	4	6	28
19	2.46	1.45	10	7	2	6	3	28
27	2.14	1.60	17	2	1	4	4	28
59	2.25	1.67	15	5	0	2	6	28
61	2.11	1.45	15	5	0	6	2	28
65	1.89	1.40	18	3	1	4	2	28
15	1.86	1.35	18	3	2	3	2	28
38	1.96	1.40	16	5	2	2	3	28
56	1.93	1.51	18	4	0	2	4	28
40	1.86	1.46	18	5	0	1	4	28
10	1.68	1.28	20	3	1	2	2	28
25	1.54	1.20	23	0	1	3	1	28
49	1.75	1.11	17	4	5	1	1	28
63	1.64	1.28	21	2	1	2	2	28
70	1.57	1.14	21	2	2	2	1	28
12	1.50	1.11	22	2	1	2	1	28
32	1.61	0.92	17	7	2	2	0	28
41	1.50	0.96	20	5	0	3	0	28
45	1.46	0.92	21	3	2	2	0	28
48	1.50	1.04	21	3	2	1	1	28
69	1.50	1.26	23	2	0	0	3	28
75	1.50	0.88	20	3	4	1	0	28
8	1.43	0.84	21	3	3	1	0	28
26	1.46	0.84	20	4	3	1	0	28
67	1.50	1.00	20	5	1	1	1	28
5	1.32	0.94	24	2	0	1	1	28
50	1.39	1.07	23	3	0	0	2	28
16	1.32	0.72	22	4	1	1	0	28
18	1.21	0.69	25	1	1	1	0	28
13	1.25	0.80	24	3	0	0	1	28
29	1.36	0.83	21	6	0	0	1	28
37	1.11	0.57	27	0	0	1	0	28
42	1.14	0.59	26	1	0	1	0	28
47	1.14	0.76	27	0	0	0	1	28
30	1.11	0.42	26	1	1	0	0	28
72	1.11	0.42	26	1	1	0	0	28
21	1.04	0.19	27	1	0	0	0	28
44	1.00	0.00	28	0	0	0	0	28
54	1.00	0.00	28	0	0	0	0	28
68	1.11	0.31	25	3	0	0	0	28

1, definite invasion; 2, probable invasion; 3, undetermined; 4, probable no-invasion; 5, definite no-invasion.

pathologists consistently diagnosed no invasion on the same cases.

Examples of definite non-invasion were similar to AIS/lepidic pattern. Examples of tumor cases with equivocal and definite invasion scores are shown in Figure 4. Morphologic features attributable to discrepant interpretation of the pattern and invasion judgments appear to include: (i) the characteristics of background stroma (fibroblastic reaction *versus* dense fibrosis), (ii) occurrence of inflammation, (iii) tumor architecture, (iv) presence of micropapillary component, and (v) detection of mucinous component.

Kappa statistics for easy cases of invasion (mean value ± s.d.) was 0.55 ± 0.06 when relying on five as well as on three categories (ie, putting together probable invasion and invasion, and probable no-invasion and non-invasion). For the difficult invasion cases, the kappa value was 0.08 ± 0.02 when relying on five and 0.15 ± 0.05 on three categories. Splitting the pathologists in two categories based on the 15 cases with equivocal scoring into a group A, which favored invasion ($n = 14$), and group B, which favored no invasion, the kappa scores for groups A and B for the easy cases (3 categories) were 0.61 ± 0.06 and 0.59 ± 0.07. In contrast, the kappa scores for the difficult cases for group A and group B were 0.16 ± 0.09 and 0.27 ± 0.15, respectively. The improvement in kappa scores for pathologists in groups A and B supports a difference in diagnostic interpretation. Comparing the composition of the two groups, there was some segregation of pathologists in relation to different countries (Chi-2, $P = 0.02$).

Discussion

In these image-based ring studies, substantial reproducibility was found for typical patterns of pulmonary adenocarcinoma subtypes. When multiple patterns are present, and for the assessment of invasion in pulmonary adenocarcinoma, the reproducibility level was good (0.77) for cases showing classic architectural patterns and fair for classic invasion (0.55), but low to poor for problematic pattern and invasion cases.

The kappa score for adenocarcinoma subtyping (0.77–0.38) was higher than in a previously reported study using an older classification (0.18).[11] In our study, it was evident that solid and lepidic patterns without collapse were more reliably recognized than others, such as micropapillary *versus* papillary and acinar *versus* papillary *versus* lepidic, in particular in relation to what constituted invasion (wholly lepidic *versus* other). Thus, a second study was undertaken to examine this area more closely, with the kappa value being very similar overall (0.40) to that for submitted problematic cases in the first phase. Although issues in distinguishing micropapillary *versus* papillary patterns are self-evident,
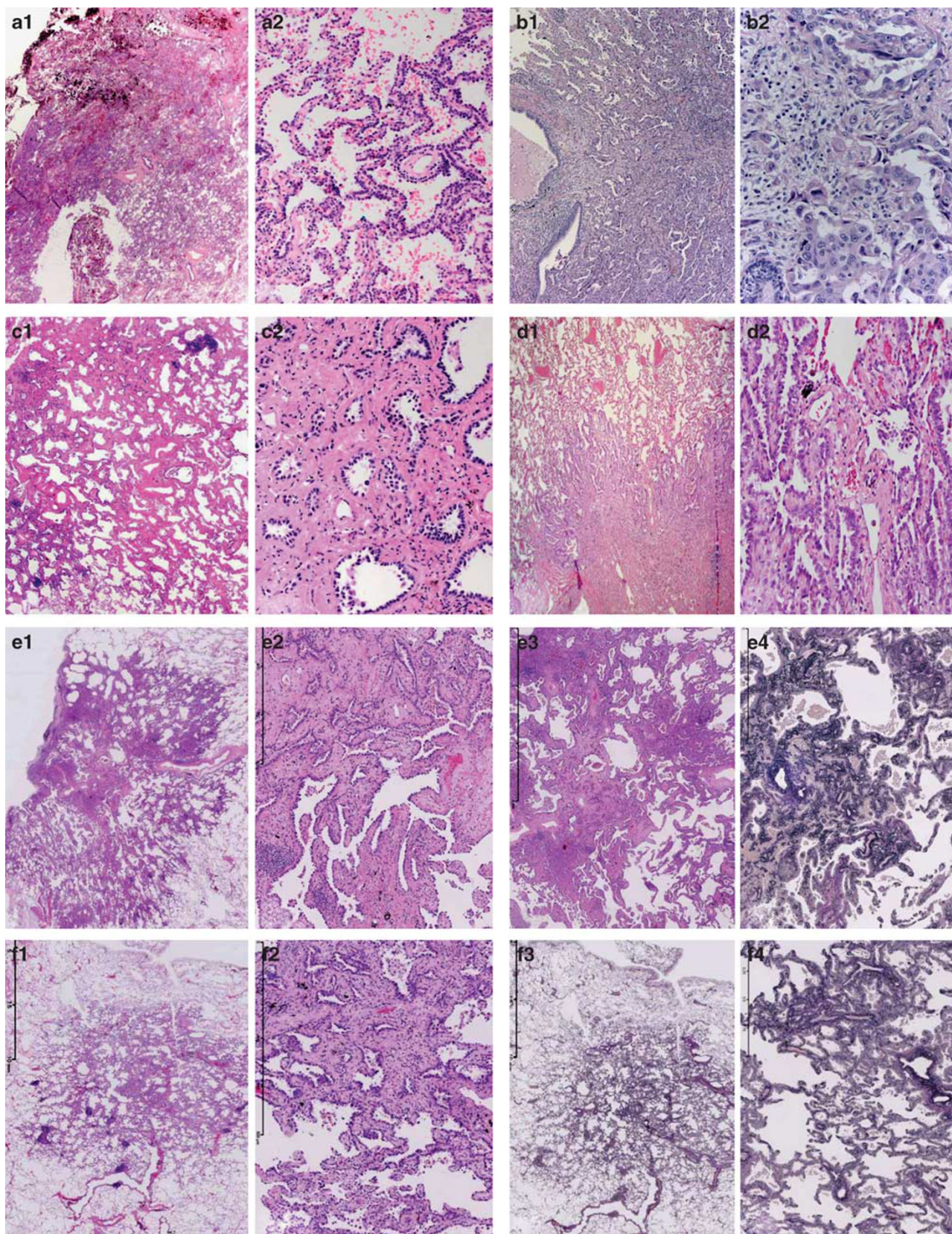
Figure 4 Examples of unanimous absence of invasion (**a**) and definite invasion (**b**) are shown, as well examples of cases (judgement on second image) with split opinion (**c–h**) having at least nine pathologists for invasion ('invasion yes ≥9') and a different group of at least 9 for non-invasion ('NO≥9'). In two cases images of another slide (same case) was available as well (**e3**, **e4** elastic stain and **f3**, **f4** elastic stains).
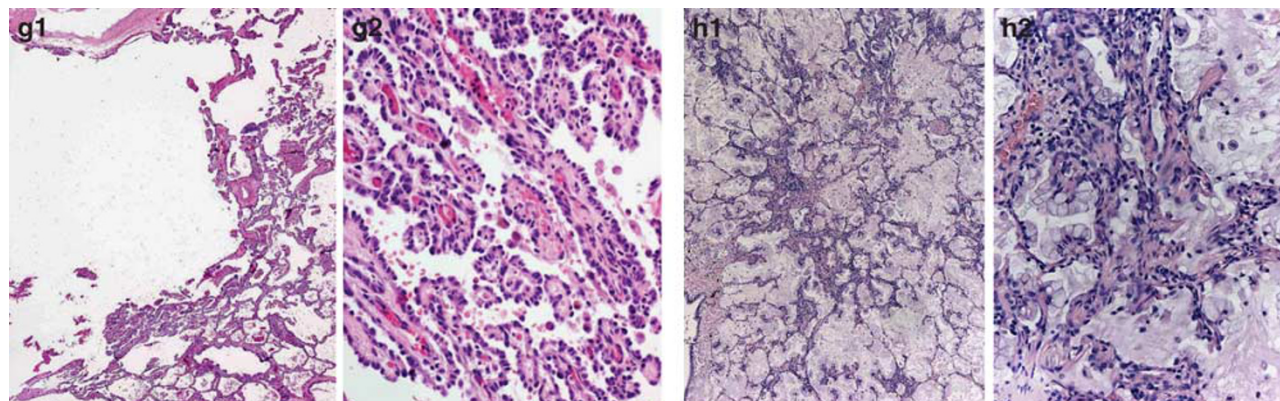
Figure 4 Continued.

problems in distinguishing acinar/papillary patterns from lepidic pattern are not so obvious. However, pulmonary adenocarcinoma poses particular challenges for pathologists through the superimposition of its neoplastic growth on the underlying lung architecture. As they grow into aerated, alveolar tissue, cross-cutting of growth along alveolar walls (lepidic) will mimic papillary structures and desmoplastic reaction will produce acinar structures which in reality are collapsed areas lacking invasion (Figure 4d, g and i). This can be further complicated by pre-existing lung architectural changes such as emphysema or interstitial fibrosis, and inconsistent use of formalin inflation technique to fix the tumor specimens.

Thus, the prime confounder is using a two-dimensional histological section to diagnose a lesion with a complex three-dimensional architecture an issue that not adequately addressed within current definitions, given kappa values in this study. This problem was borne out in both parts of this study in which there were clearly two 'groups', one being very literal (group A) in their application of diagnostic criteria and one being more interpretive (group B), whereas others are more interpretive in their approach and translating 'two-dimensional data' into a 'three-dimensional categorization'. In addition, a difference in the two groups may be related to country of practice. We believe a constructive approach is to improve definitions and increase education on the usage of this terminology.[12,13] Further studies are ongoing in this respect.

Post-study discussion also identified variation in interpretation of various morphological features. First, some pathologists interpreted a stromal component as tumor-related stroma with fibroblasts (also called desmoplastic stroma), whereas others considered the same feature as benign scarring/fibroelastosis (Figure 4c and i). Second, the presence of elastin was variably weighed as representing native alveolar wall by some pathologists but not by others (Figure 4e and f). Third, inflammation in alveolar walls implied invasive disease to some.

Fourth, although there was good agreement between pathologists in cases with a prominent micropapillary component, there was variation in interpretation between what some interpreted as focal micropapillary component and tangential cutting of both lepidic and true papillary structures. Finally, some pathologists interpreted a mucinous lepidic component as being invasive, based upon the reasonable assumption that elsewhere in the tumor an invasive component with scarring is highly likely, whereas others interpreted the image in itself as non-invasive (Figure 4h). It is therefore notable that much of the interobserver variation stems from interpretation based on operator experience and opinion, and improved definitions and better education on their usage are required to reduce interobserver variability.

The main limitation of the study was using digitized photographic image to present relevant images of pulmonary adenocarcinoma. The advantages of this approach were that there was precision of diagnosis on specific areas and the study was undertaken in timely manner. The main disadvantage was that the whole section was not examined and therefore the procedure was not representative of daily diagnostic practice. Therefore, to add some context, a low power image was added to the case. Although this did not compensate fully for using a microscope-based approach on whole sections, we believe that this did allow appropriate examination of a pathologist's ability to identify histological patterns in a robust manner. Another limitation is that the review of highly selected images does not represent the frequency of problems encountered in routine practice. For example, in this study great emphasis was placed on distinguishing invasive *versus* non-invasive patterns, when the frequency of AIS, MIA and lepidic predominant patterns occur in 10–20% of all early-stage resected lung adenocarcinomas there is no prognostic difference between the AIS and MIA categories both of which should have 100% 5-year disease-free survival, if completely resected.[3,4]

In conclusion, given that most cases of pulmonary adenocarcinoma show mixed morphology in relation to the five major histological patterns, this study provides strong evidence that a predominant pattern can be reproducibly identified with high concordance among pathologists in resection specimens, thus supportive of the adoption of 'predominant pattern' for subtyping invasive adenocarcinoma in the updated classification, as more data are published highlighting the clinical relevance of this approach. Recognition of the adenocarcinoma *in-situ* pattern is more problematic though kappa values are fair, but this area could be improved by having more precise definitions and subsequent better education on interpretation of existing terminology, and/or additional markers of invasion.

Disclosure/conflict of interest

The authors declare no conflict of interest.

References

1 Travis WD, Brambilla E, Muller-Hermelink KM. Pathology and Genetics: Tumours of the Lung, Pleura, Thymus and Heart. IARC: Lyon, 2004.

2 Travis WD, Brambilla E, Noguchi M, *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;6:244–285.

3 Russell PA, Wainer Z, Wright GM, *et al.* Does lung adenocarcinoma subtype predict patient survival?: A clinicopathologic study based on the new International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society international multidisciplinary lung adenocarcinoma classification. J Thorac Oncol 2011;6:1496–1504.

4 Yoshizawa A, Motoi N, Riely GJ, *et al.* Impact of proposed IASLC/ATS/ERS classification of lung adenocarcinoma: prognostic subgroups and implications for further revision of staging based on analysis of 514 stage I cases. Mod Pathol 2011;24:653–664.

5 Campobasso O, Andrion A, Ribotta M, *et al.* The value of the 1981 WHO histological classification in interobserver reproducibility and changing pattern of lung cancer. Int J Cancer 1993;53:205–208.

6 Stang A, Pohlabeln H, Muller KM, *et al.* Diagnostic agreement in the histopathological evaluation of lung cancer tissue in a population-based case-control study. Lung Cancer 2006;52:29–36.

7 Dale I, Lexow P, Skjorten F, *et al.* Reproducibility of tumour typing of lung carcinomas performed according to WHO's recommendation. APMIS 1989;97:351–356.

8 Yesner R. Large cell carcinoma of the lung. Semin Diagn Pathol 1985;2:255–269.

9 Roggli VL, Vollmer RT, Greenberg SD, *et al.* Lung cancer heterogeneity: a blinded and randomized study of 100 consecutive cases. Hum Pathol 1985;16:569–579.

10 Feinstein AR, Wells CK. Lung cancer staging. A critical evaluation. Clin Chest Med 1982;3:291–305.

11 Sorensen JB, Hirsch FR, Gazdar A, *et al.* Interobserver variability in histopathologic subtyping and grading of pulmonary adenocarcinoma. Cancer 1993;71:2971–2976.

12 Noguchi M, Minami Y, Iijima T, *et al.* Reproducibility of the diagnosis of small adenocarcinoma of the lung and usefulness of an educational program for the diagnostic criteria. Pathol Int 2005;55:8–13.

13 Thunnissen FB, Kerr KM, Brambilla E, *et al.* EU-USA pathology panel for uniform diagnosis in randomised controlled trials for HRCT screening in lung cancer. Eur Respir J 2006;28:1186–1189.