

# Immunohistochemical markers to diagnose primary squamous cell carcinoma of the lung: a meta-analysis of diagnostic test accuracy

Hao Chen\* , Seigo Katakura\* , Nobuyuki Horita , Ho Namkoong, Ikuma Kato, Yu Hara, Nobuaki Kobayashi , Satoshi Fujii and Takeshi Kaneko

Ther Adv Med Oncol

2022, Vol. 14: 1–11

DOI: 10.1177/  
17588359211065152

© The Author(s), 2022.  
Article reuse guidelines:  
sagepub.com/journals-  
permissions

## Abstract

**Background:** Inconsistent diagnostic test accuracies of immunohistological staining for squamous cell carcinoma (SQC) of the lung have been frequently reported. There have been few meta-analyses of the diagnostic accuracies of the immunohistochemical markers.

**Methods:** A systematic review and meta-analysis were performed following standard guidelines for systematic reviews of diagnostic test accuracy. Immunohistochemical markers (p40, p63, CK5/6, and DSC3) were evaluated as index tests for SQC. The diagnostic odds ratio (DOR) was obtained by the DerSimonian–Laird variate model. Summary estimates of sensitivity and specificity were calculated using a bivariate model. The protocol registration ID is UMIN000041664.

**Results:** The meta-analysis included 85 of the 1353 first-screened articles. The total number of patients was 17,893, which consisted 6151 SQC cases and 11,742 non-squamous non-small-cell lung cancer cases. The DOR was better for p40 (377, 95% confidence interval (CI) = 213–644,  $I^2 = 0\%$ ) than for CK5/6 (120, 95% CI = 78–184,  $I^2 = 2.5\%$ ), p63 (70, 95% CI = 55–88,  $I^2 = 9.1\%$ ), and DSC3 (94, 95% CI = 35–250,  $I^2 = 3.7\%$ ). Summary estimates of sensitivity and specificity were followings: p40 sensitivity 0.92 (95% CI = 0.89–0.95), specificity 0.94 (95% CI = 0.93–0.96); p63 sensitivity 0.92 (95% CI = 0.90–0.94), specificity 0.83 (95% CI = 0.80–0.86); CK5/6 sensitivity 0.90 (95% CI = 0.87–0.93), specificity 0.91 (95% CI = 0.89–0.93); DSC3 sensitivity 0.81 (95% CI = 0.73–0.88), and specificity 0.95 (95% CI = 0.85–0.98).

**Conclusion:** P40 had the best DOR to diagnose SQC in non-small-cell lung carcinoma. Despite its lower sensitivity, DSC3 had the best specificity among the four markers and might be useful to rule-in the diagnosis of SQC.

**Keywords:** accuracy, immunohistochemistry, sensitivity, specificity, squamous cell carcinoma

Received: 12 May 2021; revised manuscript accepted: 12 November 2021.

## Introduction

Lung cancer is the leading cause of cancer-related death. In 2017, there were 2.2 million incident cases of lung cancer and 1.9 million deaths.<sup>1</sup> Lung cancer is divided histologically into two main subtypes: small-cell lung carcinoma (SCLC) and non-small-cell lung carcinoma (NSCLC), accounting for 15% and 85% of all cases, respectively. NSCLC is further classified into three

main types: squamous cell carcinoma (SQC), adenocarcinoma (ADC), and large-cell carcinoma. SQC accounts for 25–30% of all lung cancer cases.<sup>2</sup> ADC and large-cell carcinoma are usually called non-squamous non-small-cell lung cancer (NSQ-NSCLC).

The selection of medical management is based on the histological subtype.<sup>3</sup> Although the majority

Correspondence to:  
**Nobuyuki Horita**  
Department of  
Pulmonology, Yokohama  
City University Graduate  
School of Medicine, 3-9  
Fukuura, Kanazawa-ku,  
Yokohama 236-0004,  
Japan  
[horitano@yokohama-cu.ac.jp](mailto:horitano@yokohama-cu.ac.jp)

**Hao Chen**  
Department of  
Pulmonology, Yokohama  
City University Graduate  
School of Medicine,  
Yokohama, Japan

Department of Oncology,  
Graduate School  
of Medicine, Teikyo  
University, Tokyo, Japan

**Seigo Katakura**  
**Yu Hara**  
**Nobuaki Kobayashi**  
**Takeshi Kaneko**  
Department of  
Pulmonology, Yokohama  
City University Graduate  
School of Medicine,  
Yokohama, Japan

**Ho Namkoong**  
Department of Infectious  
Diseases, Keio University  
School of Medicine, Tokyo,  
Japan

**Ikuma Kato**  
**Satoshi Fujii**  
Department of Molecular  
Pathology, Yokohama City  
University Graduate School  
of Medicine, Yokohama,  
Japan

\*Hao Chen and Seigo  
Katakura contributed  
equally to the work

of anti-cancer agents had similar efficacy for NSQ-NSCLC and SQC of the lung, drugs such as pemetrexed and bevacizumab are only effective for patients with NSQ-NSCLC.<sup>4,5</sup> Based on molecular advances and the clinical demand for accurate subclassification of lung cancer, the World Health Organization (WHO) updated the Classification of Tumors of the Lung, Pleura, Thymus, and Heart in 2015, which emphasized the expanded use of immunohistochemical techniques even for the diagnosis of SQC and NSQ-NSCLC and explicitly included some immunohistochemical markers.<sup>6</sup> The incidence of large-cell cancer of the lung has been decreasing since 2015 because these immunohistochemical markers can discern the difference between poorly differentiated SQC and ADC.<sup>7</sup>

Numerous immunohistochemical and immunocytochemical markers have been explored to distinguish between pulmonary SQC and NSQ-NSCLC. p40, p63, cytokeratin 5/6 (CK5/6), and desmocollin-3 (DSC3) have been frequently used in the diagnosis of SQC.<sup>8</sup> Sensitivity and specificity are key metrics to understand the diagnostic test accuracy of immunohistochemical staining techniques. To the best of our knowledge, no systematic review has evaluated the diagnostic test accuracy of SQC immunohistochemical markers. The current systematic review and meta-analysis aimed to summarize data from the previous studies of diagnostic test accuracy of immunohistochemical markers used for the diagnosis of SQC.

## Methods

### Study overview

The protocol of this systematic review and meta-analysis of diagnostic test accuracy was prepared following standard guidelines for systematic reviews of diagnostic test accuracy and registered on the website of the University Hospital Medical Information Network Clinical Trials Registration (UMIN000041664).<sup>9,10</sup> Approval of the Institutional Review Board was not required because of the nature of this study. A checklist of PRISMA was shown in Supplementary Table 1.

### Study search

Four major online databases, PubMed, Web of Science, Cochrane, and Embase, were searched (January 31, 2020). The following search strategy

was used for PubMed: ((p40 OR deltaNp63 OR ΔNP63) OR (p63 OR DBR16.1) OR (ck5/6 OR Cytokeratin 5/6) OR (desmocollin 3 OR desmocollin-3 OR DSC3 OR DSC-3) OR (TTF1 OR TTF-1 OR Thyroid transcription factor-1 OR Thyroid transcription factor 1) OR (NapsinA OR Napsin A OR TA02 OR aspartic protease) OR (CK7 OR cytokeratin7 OR cytokeratin 7)) AND (sensitivity and specificity) AND (NSCLC OR lung OR pulmonary OR bronchial OR pleural OR respiratory OR bronchoscopy) AND (NSCLC OR adenocarcinoma OR squamous OR squamous-cell OR non-small OR non small). The detailed information of the research stratagem was shown in Supplementary Table 2.

Two authors (SK and NH) independently screened the titles and abstracts and carefully evaluated full text to select eligible articles; in cases of discrepancy, they reached a consensus through discussion. Review articles and included original articles were hand-searched (HC and NH) for additional research papers that met the inclusion criteria.

### Study selection

Full articles, brief reports, and conference abstracts published in any language that provided data for sensitivity and specificity of immunohistochemical markers to diagnose lung SQC were included. An article that provided data of either sensitivity or specificity was excluded since bivariate analysis is not applicable for such data.<sup>9</sup> A case-control study design that consisted of patients with ADC and SQC was accepted, though a case-control design may be considered to have a risk of bias according to Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2).<sup>11</sup>

The target population was patients with NSCLC. Commonly used pathological criteria were accepted along with WHO 2015 criteria. A study that collectively evaluated both NSCLC and SCLC was excluded since such a study did not fit the clinical question. Studies focusing on non-pulmonary cancers and metastatic lung cancers of non-pulmonary origin were also excluded. Similarly, studies that compared NSCLC subtypes and mesothelioma were not accepted. Studies including patients with only ADC or SQC diagnosis were considered two-gate studies, and studies including all NSCLC patients were considered one-gate studies.

Specimens outside the lung such as lymph nodes and pleural effusion were accepted as well. Immunocytochemical staining using lung cytology or pleural effusion cell blocks was also accepted along with immunohistochemical staining. Small samples from cell blocks, lymph nodes, and pleural effusion was classified as biopsy specimen.

Target immunohistochemical markers included p40, p63, CK5/6, and DSC3 for SQC. Immunohistochemical techniques using any commercially available antibodies and non-commercial antibodies were accepted. The reference test had to be a pathological diagnosis by pathologists.

#### *Risk of bias*

QUADAS2 was applied to assess the risk of bias in each study.<sup>11</sup>

#### *Outcomes*

Sensitivity, specificity, area under the curve (AUC), and the diagnostic odds ratio (DOR) were evaluated. If two or more cutoffs were applied in an original article, all of the weakly, moderately, and strongly positive were collectively considered positive. To diagnose SQC, both SQC and adenosquamous carcinoma were counted since adenosquamous carcinoma has a squamous cell component, whereas large-cell carcinoma and NSCLC not otherwise specified were not counted as SQC.

#### *Data extraction*

Two review authors, SK and NH, independently extracted data, including the name of the first author, publication year, publication country, types of immunohistochemical markers, numbers of patients with positive results, numbers of patients evaluated, and QUADAS-2-related information.

#### *Statistics*

A bivariate model was used to obtain pooled sensitivity and specificity and to draw a summary receiver operating characteristic curve (SROC).<sup>12</sup> The DOR was obtained by the DerSimonian–Laird random model. The DOR was calculated by the ‘madauni’ command (‘netmeta’ package of R project, Gerta Rücker, Denmark). Sensitivity, specificity, and AUC were pooled by the ‘reitsma’

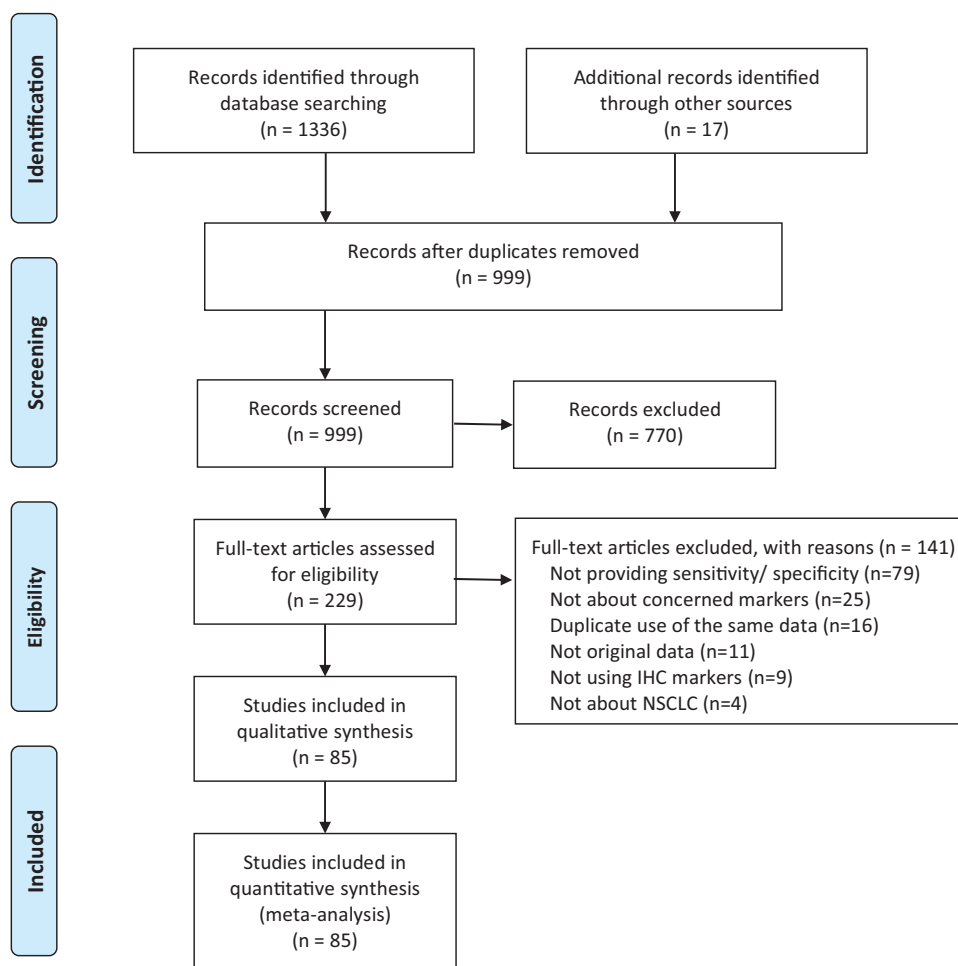
command (‘netmeta’ package of R project, Gerta Rücker, Denmark). AUCs were interpreted as follows:  $\geq 0.97$ , excellent; 0.93–0.96, very good; 0.75–0.92, very good; and 0.5–0.74, fair.<sup>13</sup> The threshold for significance was set at 0.05. Heterogeneity evaluated using  $I^2$  statistics was interpreted as follows:  $I^2 = 0\%$ , no heterogeneity;  $I^2 > 0\%$  but  $< 25\%$ , minimal heterogeneity;  $I^2 \geq 25\%$  but  $< 50\%$ , mild heterogeneity;  $I^2 \geq 50\%$  but  $< 75\%$ , moderate heterogeneity; and  $I^2 \geq 75\%$ , strong heterogeneity.<sup>14</sup>

## Results

### *Study search and study characteristics*

A total of 1346 articles, including 1336 articles through database search and 17 articles by hand search, were identified; 999, 229, and 85 articles were left after removing duplication, screening, and full-article reading, respectively (Figure 1). Finally, 85 reports, comprising 75 full-length articles and 10 conference abstracts, were included (Table 1). All were written in the English language except for one article written in the Chinese language. Prospective study designs were adopted in four articles, and the other 81 were retrospective studies. Of the 85 reports, 28 were from the United States, nine were from China, six were from Germany and Japan, five were from Turkey and the United Kingdom. Of the 17,893 patients who were enrolled in this study, 6151 had SQC based on the pathological diagnosis, and 11,742 had NSQ-NSCLC. Surgical specimens were assessed in 34 studies, and 31 studies evaluated biopsied specimens, whereas 10 studies collected both surgical and biopsy samples. Ten reports did not specify specimen type. Fifty-one studies were two-gate case–control studies, enrolling SQC and ADC, respectively, and the other 34 studies were one-gate studies that enrolled NSCLC specimens. The WHO classification of lung cancer pathology was used in 67 articles, and the other 18 studies did not mention classification criteria. The cutoff values for immunohistochemical markers were 1% in 29 studies, 5% in 6 studies, 10% in 15 studies, and 35 studies that did not report cutoff values.

Clones of used immunohistochemical markers were shown in Supplementary Table 3. Although different clones were used in studies, more than half of the studies used the same clone ploy antibody, 4A4, D5-16B4, and DSC3-U114 for p40, p63, CK5/6, and DSC3, respectively. The risk of



**Figure 1.** Flow diagram of this study.  
IHC, immunohistochemical; NSCLC, non-small-cell lung cancer.

bias assessment is shown in Figure 2. There were 45 studies with high patient selection bias, and 26 studies showed an unclear risk of selection bias. A total of 12 studies and 24 studies with high and unclear risk of bias compared to the reference. No study showed bias in patient selection applicability concerns, index test, index test applicability concerns, reference standard applicability concerns, and flow and timing.

#### Diagnostic accuracy of p40

Thirty-four studies with 6788 samples yielded a DOR of 377 (95% confidence interval (CI) = 213–644;  $P = 0\%$ ) and an AUC of 0.976. This AUC suggested that p40 had ‘excellent’ diagnostic test accuracy for SQC (Figure 3(a), Table 2).<sup>13</sup> The summary estimates of sensitivity and specificity were 0.92 (95% CI = 0.89–0.95) and 0.94 (95% CI = 0.92–0.96), respectively. The one-gate

subgroup analysis including 14 studies found similar DOR, AUC, sensitivity, and specificity of 477 (95% CI = 154–1479;  $P^2 = 0\%$ ), 0.976, 0.92 (95% CI = 0.88–0.95), and 0.94 (95% CI = 0.92–0.97), respectively (Table 2). The paired forest plots of sensitivity and specificity for each study of p40 are shown in Supplementary Figure 1. Fagan’s nomogram for p40 wash is shown in Supplementary Figure 5. For p40, likelihood positive (LR+) is 15.3, likelihood negative (LR–) is 0.85. In this example, the pretest probability is 90%. Posttest probability is 99.3% for the positive test and is 46% for the negative test.

#### Diagnostic accuracy of p63

Data of 11,898 samples from 66 reports suggested a DOR of 70 (95% CI = 55–88;  $P^2 = 9.1\%$ ) and an AUC of 0.942, which means that p63 had ‘very good’ diagnostic test accuracy for SQC

**Table 1.** Characteristics of included studies.

Author	Year	Country	Types	Nature	No. of NSCLC	No. of SQC	Standard test	Pathological type	Specimen	Cutoff value
Affandi	2018	Malaysia	FA	Retro	70	35	WHO 2015	Ad, Sq	NS.	5%
Aikawa	2011	Japan	FA	Retro	154	77	NS	Ad, Sq	B	1%
Alexander	2017	USA	FA	Retro	190	31	WHO 2015	Ad, Sq	B	1%
Allison	2015	UK	FA	Retro	246	30	NS	Ad, Sq	B	NS
Ao	2014	USA	FA	Retro	200	77	WHO 2004	Ad, Sq	S	1%
Argon	2015	Turkey	FA	Retro	120	89	WHO 2004	NSCLC	NS.	1%
Attanoos	2003	UK	FA	Retro	53	4	WHO 1999	NSCLC	S	NS
Bernardi	2018	Sao Paulo	FA	Retro	340	124	WHO 2015	Ad, Sq	B	10%
Bir	2016	Turkey	FA	Retro	100	50	WHO 2015	Ad, Sq	S	NS
Bishop	2012	USA	FA	Retro	470	81	NS	Ad, Sq	S	1%
Brown	2013	USA	FA	Retro	200	89	WHO 2004	Ad, Sq	S	1%
Cadioli	2014	USA	FA	Retro	172	46	WHO 2015	NSCLC	S	NS
Chen	2013	China	CA	Retro	61	22	WHO 2004	Ad, Sq	B	1%
Collins	2013	USA	FA	Retro	100	29	NS	NSCLC	B	5%
Comin	2014	Italy	FA	Retro	247	38	WHO 2004	NSCLC	S	1%
Delgado	2017	USA	CA	Retro	76	55	WHO 2015	Ad, Sq	B and S	NS
DePeralta-Venturina	2010	USA	CA	Retro	47	47	WHO 2004	Ad, Sq	B	1%
Downey	2008	Ireland	FA	Retro	45	21	NS	Ad, Sq	S	NS
Dvorak	2016	USA	FA	Retro	538	189	WHO 2004	Ad, Sq	B	1%
Ezzat	2016	Egypt	FA	Retro	60	32	WHO 2004	Ad, Sq	B and S	10%
Fatima	2012	USA	FA	Retro	58	44	WHO 2004	Ad, Sq	B	NS
Galindo	2020	Spain	FA	Retro	85	26	WHO 2015	Ad, Sq	S	5%
Guo	2019	China	FA	Retro	58	24	WHO 2015	Ad, Sq	S	NS
Gurda	2015	USA	FA	Retro	246	34	WHO 2004	Ad, Sq	B	5%
Hammer	2015	Turkey	FA	Retro	165	26	WHO 2004	NSCLC	S	1%
Ikeda	2015	Japan	FA	Retro	70	44	WHO 2004	NSCLC	S	1%
Kargi	2007	Turkey	FA	Retro	77	39	WHO 2004	Ad, Sq	B	NS
Kaufmann	2001	Germany	FA	Retro	248	15	WHO 1999	Ad, Sq	S	10%
Kawai	2015	Japan	FA	Retro	215	96	NS	Ad, Sq	S	5%
Khoor	2015	USA	CA	Retro	214	101	NS	Ad, Sq	NS	NS
Kim	2013	South Korea	FA	Retro	129	48	WHO 2004	Ad, Sq	S	10%
Kimbrell	2012	USA	FA	Retro	140	12	WHO 2004	NSCLC	S	1%

*(Continued)*

Table 1. (Continued)

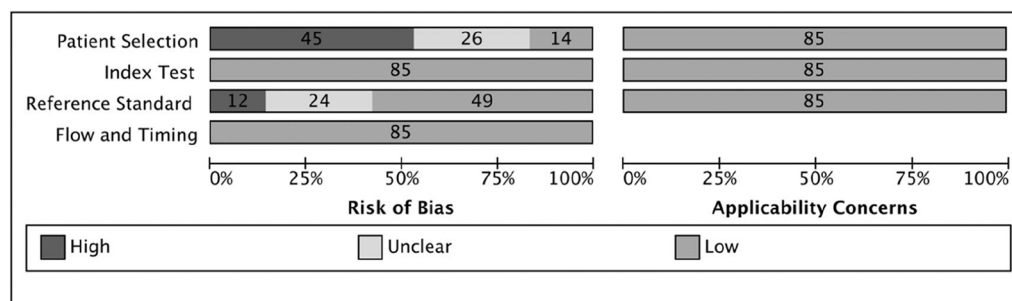
Author	Year	Country	Types	Nature	No. of NSCLC	No. of SQC	Standard test	Pathological type	Specimen	Cutoff value
Koh	2014	South Korea	FA	Retro	186	59	WHO 2004	NSCLC	B and S	10%
Kriegsmann	2019	Germany	FA	Retro	1244	569	WHO 2015	Ad, Sq	B and S	NS
Kriegsmann	2016	Germany	FA	Retro	208	98	WHO 2015	Ad, Sq	S	1%
Lilo	2016	UK	FA	Retro	144	53	WHO 2015	Ad, Sq	S	NS
Liu	2017	China	FA	Retro	23	6	WHO 2015	Ad, Sq	B	NS
Loo	2010	UK	FA	Retro	82	25	WHO 2004	NSCLC	B	1%
Marson	2004	France	FA	Retro	202	33	WHO 1999	NSCLC	NS	10%
Mukhopadhyay	2011	USA	FA	Retro	39	98	WHO 2004	NSCLC	B	10%
Nishino	2016	USA	FA	Retro	241	50	NS	Ad, Sq	B and S	1%
Noh	2012	Korea	FA	Retro	82	38	NS	NSCLC	S	10%
Nonaka	2012	UK	CA	Retro	460	30	WHO 2004	Ad, Sq	S	1%
Ocque	2011	USA	FA	Retro	448	38	WHO 2004	Ad, Sq	B	NS
Pelosi	2013	Italy	FA	Pros	119	116	WHO 2004	NSCLC	S	10%
Prabhakaran	2019	Austria	FA	Retro	200	115	WHO 2015	Ad, Sq	S	NS
Rekhtman	2011	USA	FA	Pros	315	115	WHO 2004	Ad, Sq	S	1%
Righi	2011	Italy	FA	Retro	103	25	WHO 2004	NSCLC	B	NS
Roberts	2020	USA	FA	Retro	264	99	NS	NSCLC	B and S	NS
Roh	2012	USA	FA	Retro	25	32	WHO 2004	NSCLC	B	10%
Savci-Heijink	2009	USA	FA	Retro	414	25	WHO 2004	Ad, Sq	NS	NS
Schultz	2011	Germany	FA	Retro	362	156	WHO 2004	NSCLC	B and S	NS
Sekar	2017	India	FA	Pros	60	15	WHO 2015	NSCLC	B	NS
Sethi	2012	USA	FA	Retro	35	20	NS	NSCLC	B	NS
Shah	2019	Germany	FA	Retro	100	23	WHO 2015	Ad, Sq	B	1%
Sharma	2016	USA	FA	Retro	109	37	WHO 2015	NSCLC	B	1%
Shim	2011	South Korea	CA	Retro	82	38	WHO 2004	NSCLC	S	NS
Siddiqui	2013	USA	CA	Retro	60	30	NS	Ad, Sq	B	NS
Sinna	2013	Egypt	FA	Retro	40	31	WHO 2004	NSCLC	S	NS
Sisakht	2020	Iran	FA	Retro	83	37	WHO 2015	Ad, Sq	S	1%
Sterlacci	2012	Austria	FA	Retro	371	129	WHO 2004	NSCLC	S	NS
Stojisic	2013	Serbia	FA	Retro	50	13	WHO 2004	NSCLC	B	NS
Szade	2019	Poland	FA	Retro	123	61	WHO 2015	NSCLC	B and S	NS
Tacha	2010	USA	CA	Retro	97	56	WHO 2004	Ad, Sq	B	NS

(Continued)

**Table 1.** (Continued)

Author	Year	Country	Types	Nature	No. of NSCLC	No. of SQC	Standard test	Pathological type	Specimen	Cutoff value
Tacha	2012	USA	FA	Retro	210	95	NS	Ad, Sq	NS	10%
Tacha	2014	USA	FA	Retro	527	107	NS	NSCLC	NS	1%
Tatsumori	2014	Japan	FA	Retro	580	158	WHO 2004	NSCLC	S	10%
Terry	2010	Canada	FA	Retro	425	225	WHO 2004	Ad, Sq	S	1%
Thunnissen	2012	Netherlands	FA	Pros	110	25	WHO 2004	NSCLC	B	NS
Tsuta	2011	Japan	FA	Retro	309	150	WHO 2004	Ad, Sq	S	10%
Uke	2010	India	FA	Retro	100	21	WHO 2004	Ad, Sq	NS	1%
vanZyl	2019	South Africa	FA	Retro	271	53	WHO 2015	Ad, Sq	B	NS
Vidarsdottir	2019	Sweden	FA	Retro	669	202	WHO 2004	NSCLC	B and S	1%
Vogt	2013	USA	CA	Retro	60	30	NS	Ad, Sq	NS	NS
Walia	2017	India	FA	Retro	263	58	WHO 2015	NSCLC	B	1%
Wang	2020	China	FA	Retro	314	50	WHO 2015	Ad, Sq	B	1%
Warth	2012	Germany	FA	Retro	1145	503	WHO 2004	Ad, Sq	S	1%
Whithaus	2012	USA	FA	Retro	291	66	NS	Ad, Sq	NS	1%
Xu	2014	China	FA	Retro	210	99	WHO 2004	Ad, Sq	B	NS
Yaman	2015	Turkey	FA	Retro	80	24	WHO 2004	NSCLC	S	5%
Yanagita	2011	Japan	FA	Retro	64	25	NS	Ad, Sq	B	NS
Zhan	2015	China	FA	Retro	50	50	WHO 2004	Ad, Sq	S	NS
Zhang	2009	China	FA	Retro	404	174	WHO 2004	NSCLC	B and S	10%
Zhang	2013	China	CA	Retro	198	75	NS	Ad, Sq	S	NS
Zhao	2014	China	FA	Retro	48	16	WHO 2004	NSCLC	B	10%

Ad, adenocarcinoma; B, biopsy; CA, conference abstract; FA, full article; NS, not specified; NSCLC, non-small-cell lung cancer; Pros, prospective study; Retro, retrospective study; S, surgery; SQC, squamous cell carcinoma; WHO, World Health Organization.

**Figure 2.** Selection bias of studies.

**Table 2.** Summary of diagnostic accuracy of markers.

	Studies	DOR (95% CI)	$I^2$ (%)	AUC	Sensitivity (95% CI)	Specificity (95% CI)
p40						
One gate	14	477 (154–1479)	0	0.976	0.92 (0.88–0.95)	0.94 (0.92–0.97)
Two gates	20	285 (154–525)	16	0.975	0.92 (0.87–0.95)	0.94 (0.91–0.96)
Overall	34	377 (213–664)	0	0.976	0.92 (0.89–0.95)	0.94 (0.92–0.96)
p63						
One gate	25	92 (57–148)	0	0.950	0.91 (0.87–0.94)	0.88 (0.83–0.91)
Two gates	41	61 (47–79)	10.5	0.938	0.93 (0.91–0.95)	0.80 (0.76–0.83)
Overall	66	70 (55–88)	9.1	0.942	0.92 (0.90–0.94)	0.83 (0.80–0.86)
CK 5/6						
One gate	21	131 (62–282)	13.8	0.957	0.89 (0.82–0.93)	0.92 (0.89–0.94)
Two gates	28	116 (69–195)	0	0.956	0.91 (0.87–0.94)	0.90 (0.85–0.93)
Overall	49	120 (78,184)	2.5	0.957	0.90 (0.87–0.92)	0.91 (0.89–0.93)
DSC3						
One gate	2	198 (77–506)	0	0.899	0.76 (0.63–0.85)	0.98 (0.90,0.99)
Two gates	8	90 (26–307)	4.2	0.913	0.83 (0.73,0.90)	0.94 (0.79–0.99)
Overall	10	94 (35–250)	3.7	0.909	0.81 (0.73–0.88)	0.95 (0.85–0.98)

AUC, area under the curve; CI, confidence interval; DOR, diagnostic odds ratio.

(Figure 3(b), Table 2). The summary estimates of sensitivity and specificity were 0.92 (95% CI = 0.90–0.94) and 0.83 (95% CI = 0.80–0.86), respectively. One-gate subgroup analyses focusing on studies including all NSCLC were performed, and DOR, AUC, sensitivity, and specificity were 92 (95% CI = 57–148;  $I^2$  = 0%), 0.950, 0.89 (95% CI = 0.87–0.94), and 0.88 (95% CI = 0.83–0.91), respectively. Sensitivity and specificity for studies of p63 are shown in paired forest plots in Supplementary Figure 2.

#### Diagnostic accuracy of CK5/6

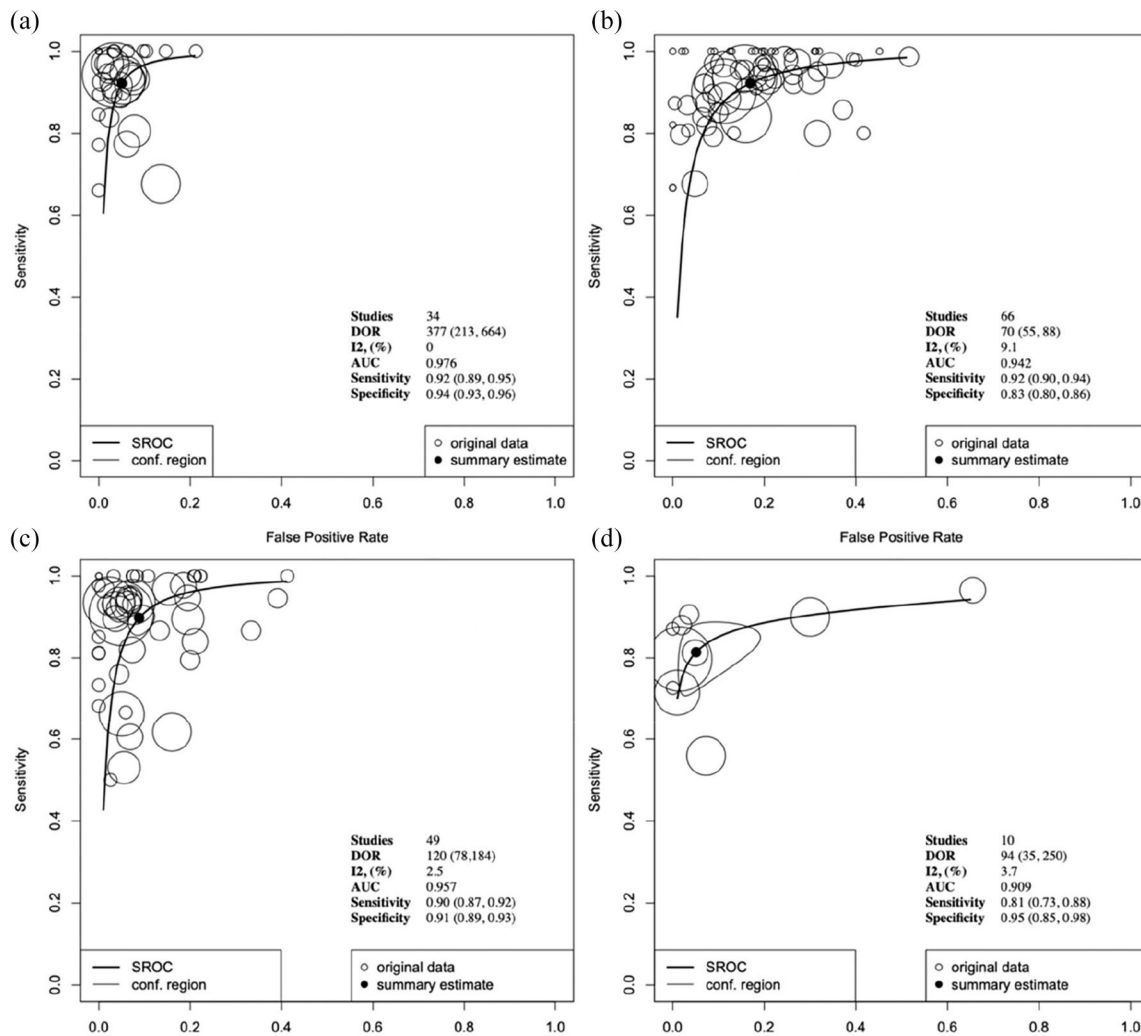
Forty-nine studies with 8962 specimens yielded a DOR of 120 (95% CI = 78–184;  $I^2$  = 2.5%) and an AUC of 0.957. This AUC value suggests that CK5/6 had ‘very good’ diagnostic test accuracy for SQC (Figure 3, Table 2). Using the data from 49 cohorts, the summary estimates of sensitivity and specificity were 0.90 (95% CI = 0.87–0.92) and 0.91 (95% CI = 0.88–0.93), respectively. A one-gate subgroup analysis including 21 cohorts

of NSCLC yielded DOR, AUC, sensitivity, and specificity of 131 (95% CI = 62–282.4;  $I^2$  = 13.8%), 0.957, 0.89 (95% CI = 0.82–0.93), and 0.92 (95% CI = 0.89–0.94), respectively. The paired forest plots of sensitivity and specificity for each study of CK5/6 are shown in Supplementary Figure 3.

#### Diagnostic accuracy of DSC3

The diagnostic test accuracy of DSC3 was examined in 2664 samples of ADC and SQC in 10 cohorts. The DOR was 93.9 (95% CI = 35.3–249.7;  $I^2$  = 3.7%), and AUC was 0.909. The sensitivity and specificity were 0.81 (95% CI = 0.73–0.88) and 0.95 (95% CI = 0.85–0.98), respectively (Figure 3(d), Table 2). DSC3 showed a ‘good’ diagnostic accuracy, though in relatively limited studies compared with other markers. There were only two cohorts including NSCLC that suggested DOR, AUC, sensitivity, and specificity in a one-gate subgroup analysis of 198 (95% CI = 77.4–506.4;  $I^2$  = 0%), 0.899, 0.76





**Figure 3.** Diagnosis accuracy of IHC markers: (a) p40, (b) p63, (c) CK5/6, and (d) DSC3. AUC, area under the curve; CI, confidence interval; DOR, diagnostic odds ratio; IHC, immunohistochemical.

(95% CI = 0.63–0.85), and 0.98 (95% CI = 0.90–0.99), respectively. The paired forest plots of sensitivity and specificity for each study of CK5/6 are shown in Supplementary Figure 4.

## Discussion

The diagnostic test accuracies of the immunohistochemical tumor markers p40, p63, CK5/6, and DSC3 in SQC were systematically reviewed. Based on our analysis, p40 showed the best DOR and AUC among these four markers, and the systematic review and meta-analysis provided evidence supporting the use of p40 as the first choice in the algorithm of diagnosis of predicting SQC, as in current guidelines.<sup>6,15</sup> Given the AUCs of p63 and CK5/6, which were at least 0.93, suggesting ‘very good’ diagnostic test accuracy,<sup>16</sup>

p63, and CK5/6 were all capable in the diagnosis of SQC as a choice, as suggested by some guidelines.<sup>17,18</sup> DSC3 did not have ‘very good’ diagnostic accuracy; however, DSC3 had the highest specificity and may be useful for ruling-in SQC when p40 and some markers for ADC are all positive. This finding supported the recommendation of using p40 in the diagnosis of predicting SQC from Lung Cancer/American Thoracic Society/European Respiratory Society (IASLC/ATS/ERS),<sup>17</sup> 2015 WHO classification of lung tumors,<sup>6</sup> and The European Society for Medical Oncology (ESMO).<sup>18</sup>

Although the detailed diagnosis accuracies of immunohistochemical tumor markers were a little different in one-gate and two-gate analysis. The expression of p40, p63, CK5/6, or DSC3 might

be seen in, for example, LCNEC or other non-ADC NSCLC. The sequence of diagnostic accuracy of each tumor marker kept the same with the result in the overall analysis. Data of studies used in this meta-analysis compared diagnosis accuracy between SQC with ADC or NSQ-NSCLC. The test accuracy of the above immunohistochemical tumor markers to identify metastases to the lungs or salivary gland-type carcinomas was still unclear. The results are seen as just the markers' ability to separate SQC from ADC.

The combination of TTF1 and p40 was recommended to identify SQC or ADC among NSCLC specimens. TTF1 single-positive suggests ADC of the lung, and p40 single-positive diagnoses SQC. When TTF1 and p40 are double-positive, the specimen should be further stained by highly specific markers such as Napsin A and DSC3, a protein found in desmosomes.<sup>19</sup> On the contrary, when TTF1 and p40 are double-negative, another sensitive marker for ADC, such as CK7. Although CK7 cannot be regarded as an ADC marker, for example, a significant proportion of SQC are positive for CK7, while the addition of CK7 or broad keratin in TTF1/p40-negative NSCC without clear morphology is recommended.<sup>20</sup> Additional sensitive markers for SQC are also required; p63 and CK5/6 are candidates additional immunohistochemical stains. It is true that p63 is more sensitive than CK5/6 for the diagnosis of SQC. Nonetheless, since p40 is the N-terminally truncated isoform of p63,<sup>21</sup> IHC results of p40 and p63 correlate with each other. CK5/6, intermediate-sized basic keratins with a molecular mass of 58 kDa,<sup>22</sup> had a different immunostaining target from p40. Although p63 was slightly more sensitive than CK5/6, CK5/6 might be a better additional marker when TTF1 and p40 are double-negative.

The largest number of studies of SQC IHC markers was conducted for p63, followed by CK5/6, p40, and DSC3. CK5/6 and p63 were the previous standards to diagnose SQC, whereas p40 and DSC3 have been investigated since around 2011. Although studies of p40 and DSC3 were relatively fewer, both of them had abundant samples. Across all analyses, observed heterogeneities were almost absent ( $I^2 < 25\%$ ).

There were several limitations in this study. First, the included studies shown by QUADAS-2 were the two-gate study design. A high risk of patient selection was observed. However, results from sensitivity subgroup analysis focusing on one-gate

studies were compatible with those from two-gate studies. Second, we searched data from 2001, and diagnosis standard was different with different periods. A total of 36 studies showed a high or unclear risk of reference standard. Third, although more than half of the studies used the same clones of immunohistochemical markers, different clones, and protocols might potentially exit a selection bias in this study.

### Conclusion

P40 was the only marker with 'excellent' AUC to diagnose SQC among NSCLC. Both CK5/6 and p63 showed 'very good' AUC; however, CK5/6 may have slightly better diagnostic test accuracy. Despite the lower sensitivity, DSC3 had the best specificity among the four markers, and it might be useful to rule-in the diagnosis of SQC.

### Author contribution

**Hao Chen:** Conceptualization; Data curation; Writing-original draft.

**Seigo Katakura:** Conceptualization; Data curation.

**Nobuyuki Horita:** Conceptualization; Project administration.

**Ho Namkoong:** Investigation.

**Ikuma Kato:** Investigation.

**Yu Hara:** Investigation.

**Nobuaki Kobayashi:** Investigation.

**Satoshi Fujii:** Investigation.

**Takeshi Kaneko:** Investigation.

### Conflict of interest statement

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

### ORCID iDs

Hao Chen  <https://orcid.org/0000-0003-3074-2235>

Seigo Katakura  <https://orcid.org/0000-0003-4851-3377>

Nobuyuki Horita  <https://orcid.org/0000-0002-8200-0340>

Nobuaki Kobayashi  <https://orcid.org/0000-0002-7064-320X>

### Data availability statement

The raw data are available by email on reasonable request to the corresponding author. Email is [horitano@yokohama-cu.ac.jp](mailto:horitano@yokohama-cu.ac.jp).

### Supplemental material

Supplemental material for this article is available online.

### References

1. Fitzmaurice C, Abate D, Abbasi N, *et al.* Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 29 cancer groups, 1990 to 2017: a systematic analysis for the global burden of disease study. *JAMA Oncol* 2019; 5: 1749–1768.
2. Denisenko TV, Budkevich IN and Zhivotovskiy B. Cell death-based treatment of lung adenocarcinoma. *Cell Death Dis* 2018; 9: 117.
3. Reck M and Rabe KF. Precision diagnosis and treatment for advanced non-small-cell lung cancer. *N Engl J Med* 2017; 377: 849–861.
4. Davidson MR, Gazdar AF and Clarke BE. The pivotal role of pathology in the management of lung cancer. *J Thorac Dis* 2013; 5(Suppl. 5): S463–S478.
5. Scagliotti GV, Parikh P, Von Pawel J, *et al.* Phase III study comparing cisplatin plus gemcitabine with cisplatin plus pemetrexed in chemotherapy-naïve patients with advanced-stage non-small-cell lung cancer. *J Clin Oncol* 2008; 26: 3543–3551.
6. Travis WD, Brambilla E, Nicholson AG, *et al.* The 2015 World Health Organization classification of lung tumors: impact of genetic, clinical and radiologic advances since the 2004 classification. *J Thorac Oncol* 2015; 10: 1243–1260.
7. Stoll LM, Johnson MW, Gabrielson E, *et al.* The utility of napsin-A in the identification of primary and metastatic lung adenocarcinoma among cytologically poorly differentiated carcinomas. *Cancer Cytopathol* 2010; 118: 441–449.
8. Righi L, Graziano P, Fornari A, *et al.* Immunohistochemical subtyping of non-small cell lung cancer not otherwise specified in fine-needle aspiration cytology: a retrospective study of 103 cases with surgical correlation. *Cancer* 2011; 117: 3416–3423.
9. Leeflang MM, Deeks JJ, Takwoingi Y, *et al.* Cochrane diagnostic test accuracy reviews. *Syst Rev* 2013; 2: 82.
10. UMIN000041664, [https://upload.umin.ac.jp/cgi-open-bin/ctr/ctr\\_view.cgi?recptno=R000047560](https://upload.umin.ac.jp/cgi-open-bin/ctr/ctr_view.cgi?recptno=R000047560)
11. Whiting PF, Rutjes AW, Westwood ME, *et al.* QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; 155: 529–536.
12. Higgins JPT and Green S. Cochrane handbook for systematic reviews of interventions, <https://training.cochrane.org/handbook>
13. Jones CM and Athanasiou T. Summary receiver operating characteristic curve analysis techniques in the evaluation of diagnostic tests. *Ann Thorac Surg* 2005; 79: 16–20.
14. Higgins JP, Thompson SG, Deeks JJ, *et al.* Measuring inconsistency in meta-analyses. *BMJ* 2003; 327: 557–560.
15. Sung S, Heymann JJ, Crapanzano JP, *et al.* Lung cancer cytology and small biopsy specimens: diagnosis, predictive biomarker testing, acquisition, triage, and management. *J Am Soc Cytopathol* 2020; 9: 332–345.
16. Hosmer DW and Lemeshow S. *Applied logistic regression*. 2nd ed. New York: John Wiley & Sons, 2000, pp. 160–164.
17. 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.
18. Kerr KM, Bubendorf L, Edelman MJ, *et al.* Second ESMO consensus conference on lung cancer: pathology and molecular biomarkers for non-small-cell lung cancer. *Ann Oncol* 2014; 25: 1681–1690.
19. Monica V, Ceppi P, Righi L, *et al.* Desmocollin-3: a new marker of squamous differentiation in undifferentiated large-cell carcinoma of the lung. *Mod Pathol* 2009; 22: 709–717.
20. Yatabe Y, Dacic S, Borczuk AC, *et al.* Best practices recommendations for diagnostic immunohistochemistry in lung cancer. *J Thorac Oncol* 2019; 14: 377–407.
21. Bishop JA, Teruya-Feldstein J, Westra WH, *et al.* P40 ( $\Delta$ Np63) is superior to p63 for the diagnosis of pulmonary squamous cell carcinoma. *Mod Pathol* 2012; 25: 405–415.
22. Chu PG and Weiss LM. Expression of cytokeratin 5/6 in epithelial neoplasms: an immunohistochemical study of 509 cases. *Mod Pathol* 2002; 15: 6–10.