



# The High Diagnostic Accuracy of Combined Test of Thyroid Transcription Factor 1 and Napsin A to Distinguish between Lung Adenocarcinoma and Squamous Cell Carcinoma: A Meta-Analysis

Li Li<sup>1</sup>, Xiaorong Li<sup>1</sup>, Jieyun Yin<sup>1</sup>, Xia Song<sup>1</sup>, Xiaochen Chen<sup>1</sup>, Jiane Feng<sup>2</sup>, Hongyu Gao<sup>3</sup>, Li Liu<sup>1\*</sup>, Sheng Wei<sup>1\*</sup>

**1** Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical College of Huazhong University of Science and Technology, Wuhan, Hubei, China, **2** Department of Cardiovascular Surgery, Union hospital, Tongji Medical College of Huazhong University of Science and Technology, Wuhan, Hubei, China, **3** Renal Division, Tongji Hospital, Tongji Medical College of Huazhong University of Science and Technology, Wuhan, Hubei, China

## Abstract

**Background:** Accurate classification of non-small cell lung cancer (NSCLC) using morphological features has several limitations. However, the use of thyroid transcription factor 1 (TTF-1) and Napsin A as markers for the identification of various subtypes of NSCLC has shown promise. This meta-analysis was designed to evaluate the diagnostic value of combined TTF-1 and Napsin A test to distinguish lung adenocarcinoma from squamous cell carcinoma.

**Methods:** The Medline, EMBASE and Web of Science databases were searched, along with the reference lists of relevant articles (up to May 4, 2014). Ten studies containing 1,446 subjects were identified. The sensitivity, specificity, diagnostic odds ratio (DOR) and area under the summary receiver operating characteristics curve (AUC) were calculated to estimate the combined diagnostic value of TTF-1 and Napsin A.

**Results:** The pooled sensitivity and specificity were 0.76 (95% CI: 0.69–0.83) and 1.00 (95% CI: 0.92–1.00), respectively. The positive and negative likelihood ratios were 877.60 (95% CI: 8.40–91533.40) and 0.24 (95% CI: 0.18–0.32). The DOR was 3719 (95% CI: 33–414884). The AUC was 0.92 (95%CI: 0.89–0.94). The patient's location was a source of heterogeneity for sensitivity. The patient's location, the study's sample size and the threshold used to determine positive staining were consistently found to be sources of heterogeneity for specificity in subgroup analyses and meta-regression.

**Conclusions:** The combined test of TTF-1 and Napsin A presents a promising alternative method, useful to distinguish between lung adenocarcinoma and squamous cell carcinoma.

**Citation:** Li L, Li X, Yin J, Song X, Chen X, et al. (2014) The High Diagnostic Accuracy of Combined Test of Thyroid Transcription Factor 1 and Napsin A to Distinguish between Lung Adenocarcinoma and Squamous Cell Carcinoma: A Meta-Analysis. PLoS ONE 9(7): e100837. doi:10.1371/journal.pone.0100837

**Editor:** John D. Minna, University of Texas Southwestern Medical Center at Dallas, United States of America

**Received:** February 12, 2014; **Accepted:** May 30, 2014; **Published:** July 8, 2014

**Copyright:** © 2014 Li et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by the National Science Foundation of China [http://www.nsf.gov.cn/\(NSF 81172754 and NSF 81172752\)](http://www.nsf.gov.cn/(NSF 81172754 and NSF 81172752)). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* Email: [gracefulliuly@gmail.com](mailto:gracefulliuly@gmail.com) (LL); [ws2008cn@gmail.com](mailto:ws2008cn@gmail.com) (SW)

## Introduction

Lung cancer is the leading cause of cancer mortality worldwide. Approximately eighty percent of lung cancers are determined to be non-small cell lung cancer (NSCLC), a heterogeneous group comprised of adenocarcinoma (AC) and squamous cell carcinoma (SQCC) [1]. Due to the prevalence of NSCLC, extensive researches in the field have enabled the development of various therapies targeting specific types of lung cancer [2]. Therefore, demand for the accurate classifications of different subtypes of NSCLC has risen to allow maximization of the subsequent therapeutic response and minimization of any adverse effects, especially for poorly differentiated non-small cell lung carcinomas, which were classified generally as non-small cell carcinoma not otherwise specified (NSC NOS) based on hematoxylin-eosin

(H&E) immunohistochemistry method [3–5]. However, poorly differentiated non-small cell lung carcinomas could be further divided into several subtypes, mostly into adenocarcinomas (AC) and squamous cell carcinoma (SQCC), which require different treatment strategies. For example, AC frequently harbour activating mutations of the epidermal growth factor receptor gene (EGFR) or EML4–ALK rearrangements, so epidermal growth factor receptor (EGFR) inhibitors and vascular endothelial growth factor (VEGF) inhibitors are more effective in treating AC than SQCC [6–8]. SQCC commonly express insulin-like growth factor 1 receptor (IGF-1R), which is a target for figitumumab. In addition, molecularly-targeted therapies such as Avastin should not be used to treat SQCC patients, as it has a 30% mortality rate, due to fatal hemoptysis [9,10]. Therefore, misdiagnosis or inaccurate diagnosis of poorly differentiated non-small cell lung

carcinoma by single hematoxylin-eosin (H&E) immunohistochemical method may lead to patients' insensitivity to subsequent therapies and maximization of adverse effects. And this will in turn lead to difficult management of lung cancer patients and huge consumption of medical and health resources. Alternatively, the use of new molecular biomarkers, whose identification only require a small amount of cells, has been proposed as a powerful tool to distinguish AC from SQCC [11–14]. The value of immunohistochemical markers has been well established in separating poorly differentiated AC from SQCC [12,15–17]. Two such markers recently showing promise are Thyroid transcription factor 1 (TTF-1) and Napsin A.

TTF-1, a highly-conserved homeodomain-containing transcriptional factor involved in the early development of lung. It is a well established immunomarker for lung adenocarcinoma, with sensitivity ranging from 60% to 100%, and specificity ranging from 97% to 100% [18–21]. Napsin A, a functional aspartic proteinase expressed in the cytoplasm of healthy lung parenchyma, has recently been reported as a promising immunomarker associated with lung adenocarcinomas [22,23]. And the potential of TTF-1 and Napsin A has been revealed in separating AC from SQCC even in poorly differentiated NSCLC [17]. Stoll et al. reported the sensitivity and specificity for Napsin A immunohistochemistry in diagnosing poorly differentiated pulmonary ADCs were 65% and 96%, respectively [16]. Besides the advantage in the diagnosis of lung adenocarcinoma, both TTF-1 and Napsin A have also been found in biopsies of lung squamous cell carcinoma at rates of 2% to 13% for TTF-1, and upwards of 26% for Napsin A [23–28]. This has led researchers to propose the combined test of TTF-1 and Napsin A in order to subclassify NSCLC [29,30]. However, the diagnostic accuracy of the combined test has varied among studies, with the specificities ranging from 0.88 to 1.00 [22,24,27,31], and the sensitivity between 0.47 and 1.00 [1,8,30,32]. Until now, the diagnostic value of the combined test has not been systematically evaluated. The present meta-analysis was designed to summarize the evidence behind the value of a combined TTF-1 and Napsin A test for subtype-classification of lung AC and SQCC in NSCLC.

## Materials and Methods

### Literature search

A thorough search of Medline, EMBASE and Web of Science (up to May 4, 2014) was conducted to identify eligible studies. The following search terms were used: “lung cancer” OR “lung neoplasms” OR “lung neoplasm” OR “lung carcinoma” OR “lung tumor” OR “pulmonary neoplasm” OR “pulmonary neoplasms” OR “pulmonary cancer” OR “carcinoma, non-small cell lung” OR “non-small cell lung cancer” OR “non-small cell lung carcinoma” OR NSCLC, “thyroid transcription factor-1” OR TTF-1, Napsin A, without language restrictions. The references of any relevant articles were also scanned for potentially missing studies.

### Selection criteria

Selection criteria were given as follows: (1) the combined detection of TTF-1 and Napsin A was used to distinguish between AC and SQCC; (2) the combined test of TTF-1 and Napsin A was serial test, which mean tissues with positive immunohistochemical staining for both TTF1 and Napsin A were determined as adenocarcinomas, otherwise, as squamous cell carcinoma; (3) morphological diagnoses, such as hematoxylin-eosin staining, were used as the reference diagnostic standard; (4) results were reported in numbers of true-positive, false-positive, true-negative, and false-

negative, or sufficiently detailed data were presented to derive these numbers. Studies were excluded if they were as follows: (1) a case report, review or conference proceeding; (2) containing no specific results regarding lung cancer; (3) with a sample size less than 50. For multiple or duplicate publications that covered the same dataset, only the most recent or complete study was included.

### Data extraction and quality assessment

The two investigators independently extracted data from all studies fulfilling the inclusion criteria. The following information was extracted from each study: first author, year of publication, patients' location, sample size, the number of AC and SQCC, specimen type and the threshold for staining positivity. Any disagreement was resolved by consensus. The quality of methodology for each study was assessed using the quality assessment of diagnostic accuracy studies (QUADAS) [33]. For each item, a score of 1 was applied if the answer was “yes”; otherwise, a score 0 was applied.

### Statistical analysis

For studies with true-positive, true-negative, false-positive and false-negative numbers available, we computed the following parameters and corresponding 95% confidence intervals for each study: sensitivity, specificity, and the diagnostic odds ratio combined by positive and negative likelihood ratios (LRs). Data were finally pooled in summarized receiver-operating characteristic curves (sROC), where the area under the sROC (AUC) measures test precision. The heterogeneity was explored with a Cochran's chi-square test and quantified by calculating the  $I^2$  statistic to reflect the degree of variability in results across studies. To assess any potential confounding factors, including the patient's location (Western vs. Asian), specimen types (resected vs. others), sample size (over 100 vs. less than 100), and threshold for staining positivity ( $\geq 1\%$  vs.  $\geq 10\%$ ), the subgroup analyses and meta-regression were performed taking the above factors into account. Additionally, sensitivity analysis was also performed to assess the influence of any individual study on the overall estimate.

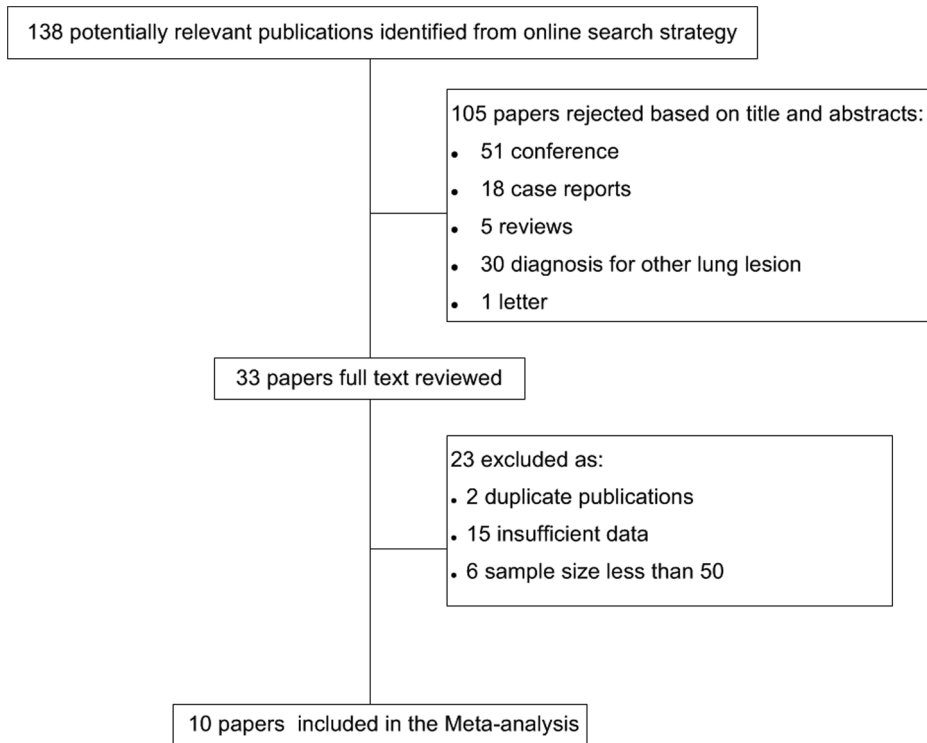
In order to test for publication bias, Deek's funnel plot method was applied. Statistical analyses were performed using Midas module in the Stata (Version 10.0), and all  $P$ -values calculated as two-sided. The association was considered significant if the  $P$ -value was less than 0.05.

## Results

### Characteristics and methodological quality of include studies

The search of the databases produced 222 studies, of which 84 were excluded for duplication, leaving 138 potentially relevant studies to be retrieved. One hundred and five studies were excluded following the first screening based on abstracts and titles. Thirty-three studies were then retrieved for full text review. After carefully reading the full text articles, 2 were excluded for duplication and 6 studies were excluded for containing a sample size less than 50 [1,9,10,12,29,34]. And 15 studies were excluded for containing insufficient data to calculate sensitivity and specificity. Finally, 10 studies were eligible for inclusion in this meta-analysis [8,11,22,24,27,31,32,35–37] (Figure 1). The characteristics of the included studies are detailed in Table 1.

The quality assessments on the included studies are presented in Figure 2. Out of 14 QUADAS items, item 1 (spectrum composition) and 2 (selection criteria) describe the variability of the studies, while items 8 (index test execution), 9 (reference standard execution) and 13 (uninterpretable test results) assess the



**Figure 1. Study flow chart for the process of selecting the eligible publications.**  
doi:10.1371/journal.pone.0100837.g001

quality of the reporting. The remainders are about the bias of the studies. The QUADAS scores of the studies ranged from 8 to 12 with a median score of 10. Items 1 and 2 were 60.0% and 30.0% fulfilled by studies respectively, indicating high variability among studies. The execution proportions of item 8, 9 and 13 were 60.0%, 20.0% and 40.0% respectively, suggesting poor reporting quality of studies. The remainders of the items reached a level of 100%, with exception of items 5 (verification of diagnosis), and 10 (blinding for index test results), which were only 60.0% and 50.0% fulfilled by the studies, respectively.

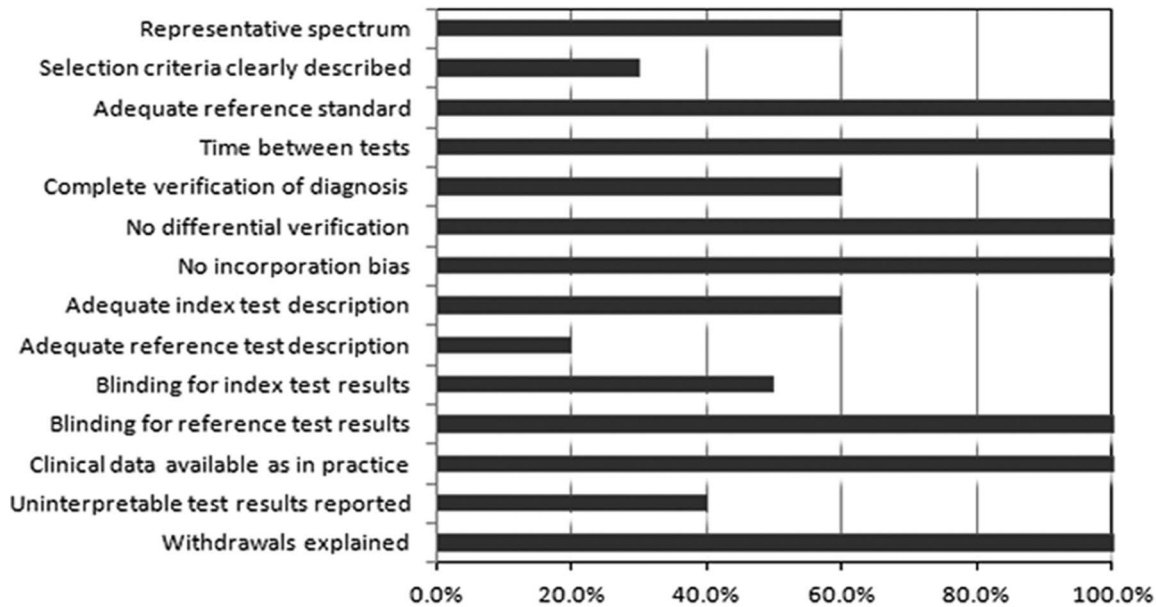
**Meta-analysis and heterogeneity**

Ten studies containing 1,446 subjects were summarized to assess the diagnostic accuracy of the combined test of TTF-1 and Napsin A. The pooled sensitivity and specificity were 0.76 (95% CI: 0.69–0.83), and 1.00 (95% CI: 0.92–1.00), respectively. Forest plots showed relative strength of the diagnostic accuracy of the combined test (Figure 3). The positive and negative LR<sub>s</sub> of the studies were 877.60 (95% CI: 8.40–91533.40) and 0.24 (95% CI: 0.18–0.32), respectively. The pooled diagnostic odds ratio was 3719 (95% CI: 33–414884). The area under the sROC was 0.92

**Table 1. Characteristics of Studies Included in the Meta-analysis.**

First Author	Year	Country	Sample size	AC/SQCC	Specimen type	Stain positive
Bishop	2010	America	102	50/52	resected specimens	≥1% <sup>a</sup>
Yang	2010	America	197	158/39	resected specimens	≥1%
Zhang	2010	China	297	212/85	resected specimens	≥10% <sup>b</sup>
Yanagita	2011	Japan	64	39/25	surgical, biopsy cases	≥1%
Fatima	2011	America	59	35/24	FNA	≥1%
Turner	2012	Japan	188	94/94	resected specimens	≥1%
Tacha	2012	America	210	115/95	resected specimens	≥10%
Noh	2012	Korea	74	36/38	resected specimens	≥1%
Collins	2013	America	69	35/34	FNA	≥10%
Brunnström	2013	Swedish	186	121/65	resected specimens	≥1%

NOTE:  
<sup>a</sup>, cases with more than 1% of tumor cells staining were classified as positive;  
<sup>b</sup>, cases with more than 10% of tumor cells staining were classified positive;  
 FNA, Fine-needle aspiration; AC, lung adenocarcinoma; SQCC, squamous cell carcinoma.  
 doi:10.1371/journal.pone.0100837.t001

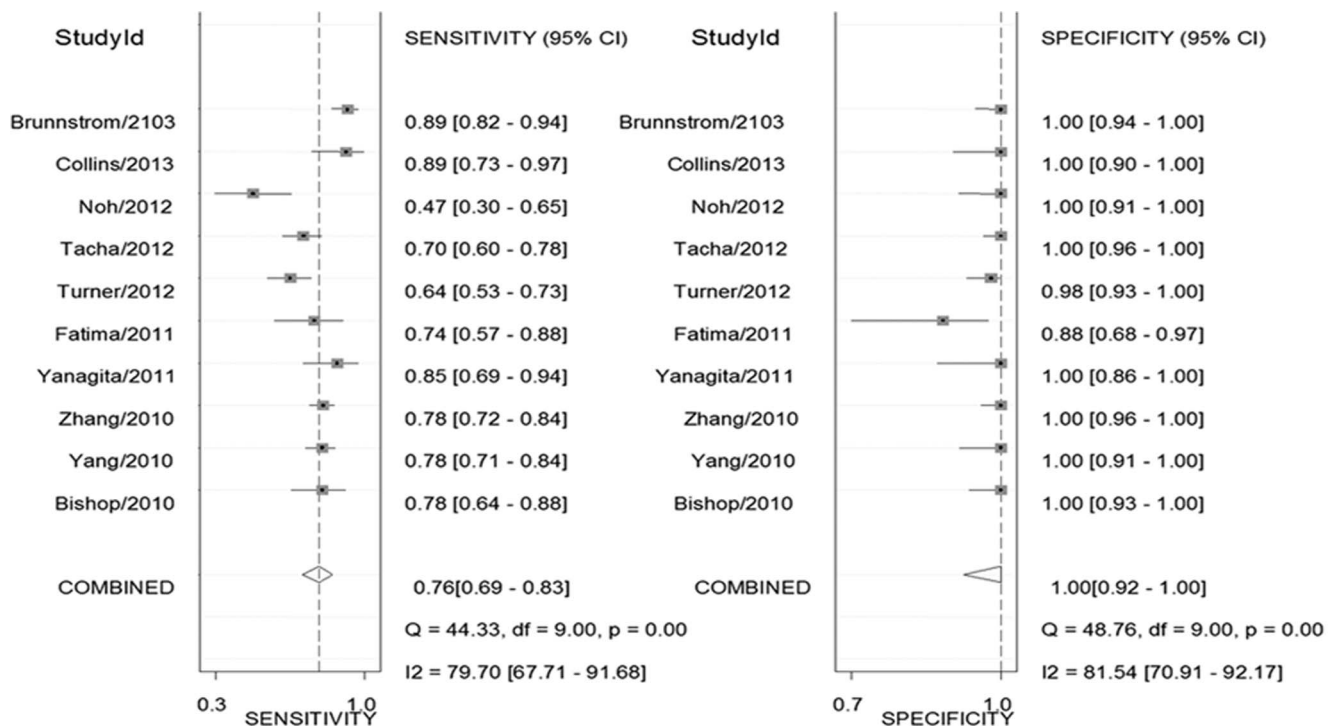


**Figure 2. The methodological quality assessments of studies included in the meta-analysis, the vertical coordinate presents 14 QUADAS items of the quality assessment of diagnostic accuracy studies (QUADAS); the horizontal axis presents percentages of 14 QUADAS items fulfilled by studies included in the meta-analysis.**  
doi:10.1371/journal.pone.0100837.g002

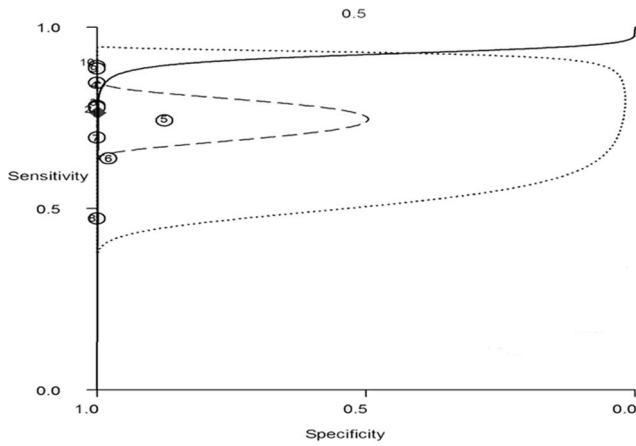
(95% CI: 0.89–0.94), indicating high precision of the combined test (Figure 4).

Significant heterogeneity was observed both in sensitivity ( $Q = 44.33$ ;  $P \leq 0.01$ ;  $I^2 = 79.70$ ) and specificity ( $Q = 48.76$ ;  $P \leq 0.01$ ;  $I^2 = 81.54$ ). Subsequently, subgroup analyses and meta-regression were conducted to identify the sources of heterogeneity

with respect to the patient’s location, sample size, specimen type, and threshold of staining positivity. The results of the subgroup analyses showed significant heterogeneity observed for sensitivity in all subgroups (Table 2). The heterogeneity of specificity in studies conducted in Asians was not statistically significant ( $I^2 = 18.84$ ;  $P = 0.30$ ). However, it was adverse in studies conduct-



**Figure 3. Pooled sensitivities and specificities of combined test of TTF-1 and Napsin A in distinction between AC and SQCC.**  
doi:10.1371/journal.pone.0100837.g003



**Figure 4. ROC curve analyses of combined test of TTF-1 and Napsin A in distinction between AC from SQCC.** ACU = 0.92[0.89–0.94], sensitivity = 0.76[0.69–0.83], specificity = 1.00[0.92–1.00]. Solid line presents sROC; circle presents observed data; diamond presents summary operating point; dash line presents 95% confidence contour; points dotted line presents 95% prediction contour.  
doi:10.1371/journal.pone.0100837.g004

ed in western countries ( $I^2 = 87.40$ ;  $P \leq 0.01$ ). When compared to holistic heterogeneity of specificity ( $I^2 = 81.54$ ;  $P \leq 0.01$ ), the heterogeneity was reduced and was not statistically significant for studies with large sample sizes ( $I^2 = 40.08$ ;  $P = 0.14$ ), similar in studies that only contained resected specimens as test materials ( $I^2 = 34.35$ ;  $P = 0.17$ ). And the sensitivity and specificity of the seven studies as just resected specimens used were not significant different from that of the ten studies as resected specimens and biopsies combined (0.74 (95%CI: 0.65–0.82) vs. 0.76 (95%CI: 0.69–0.83)), similar in specificity (1.00 (95%CI: 0.92–1.00) vs. 1.00 (95%CI: 0.92–1.00)). It indicated that the three studies [27,32,36] using small biopsy or cytology specimens as the reference diagnostic standard were not responsible for influencing heterogeneity in sensitivity and specificity. Through meta-regression, the source of heterogeneity for sensitivity was found to be the patient’s location ( $P \leq 0.01$ ). The source of heterogeneity for specificity was found to be a result of the patient’s location, sample size, and threshold of staining positivity (all with  $P$ -values of less than 0.001). Furthermore, sensitivity analysis showed that removing study of

Noh [24] resulted in a wider range in 95%CI of specificity but the specificity was stable (Table S1).

Finally, the funnel plots for the diagnostic value of combined test did not reveal any evidence of obvious asymmetry, with all  $P$ -values of the Egger’s test greater than 0.05 (Figure S1). Therefore, it was determined that publication bias did not have a statistically significant effect in this meta-analysis.

**Discussion**

The combined test of TTF-1 and Napsin A has been considered as a promising alternative tool to subclassify NSCLC in clinical practice. However, the sensitivity and specificity of this combined test has varied greatly in previous studies. The present meta-analysis confirmed that the diagnostic value of this combined test was acceptable in distinguishing AC from SQCC, with a relatively high AUC (0.92, 95%CI: 0.89–0.94) and specificity (1.00, 95%CI: 0.92–1.00), and moderate sensitivity (0.76, 95%CI: 0.69–0.83).

In recent years, the need for accurate subclassification of NSCLC has increased, as emerging evidence has suggested that the specific NSCLC subtypes will respond differently to targeted therapies. The current World Health Organization classification of lung cancer has been based almost entirely on the results of H&E staining. Although H&E evaluation could provide sufficient information to classify NSCLC subtypes in many cases, accurate diagnoses may be limited in case of where: only small biopsies or cytology specimens are available, there is poorly differentiated neoplasm, or there is a marked disruption of the histological architecture. What’s more, the concordance rates among pathologists often vary significantly. One study showed that concordance rates among pathologists in subclassifying NSCLC by H&E alone were only 81% [38]. As a result of these limitations, several molecular diagnoses have been proposed to subclassify NSCLC in recent years, such as *EGFR* mutations, *K-ras* mutations, *EML4-ALK* fusions and miRNA profiling [11,39–42]. *EGFR* mutations, *K-ras* mutations and *EML4-ALK* fusions have been shown to be primarily restricted to lung adenocarcinoma. Following identification, specific therapies could then be used, such as the *EGFR* inhibitors Erlotinib and Gefitinib that showed the greatest benefit in *EGFR* mutation-positive tumors, which were predominantly adenocarcinoma. Similarly, tumors presenting *EML4-ALK* fusion proteins or fusion gene could be treated with greater success by Crizotinib, an inhibitor of anaplastic lymphoma kinase (ALK). Although these molecular diagnostic markers were precise and contributed to targeted therapies, most of the studies used *EGFR*

**Table 2. Subgroup analysis.**

Subgroup	SEN (95% CI)	SPE (95% CI)	AUC	$I^2$ (SEN,95%CI)	$I^2$ (SPE,95%CI)
Patient’s location					
Western	0.80 (0.72–0.86)	1.00 (0.09–1.00)	0.90	70.57 (45.64–95.51)	87.40 (78.73–96.06)
Asian	0.70 (0.56–0.82)	0.99 (0.96–1.00)	— <sup>c</sup>	85.96 (73.40–98.52)	18.84 (0.00–100.00)
Sample size					
Large <sup>a</sup>	0.77 (0.70–0.83)	1.00 (0.82–1.00)	0.92	78.61 (61.77–95.46)	40.08 (0.00–95.51)
Small <sup>b</sup>	0.76 (0.58–0.88)	1.00 (0.53–1.00)	0.94	85.63 (72.69–98.56)	81.97 (64.83–99.11)
Specimen type					
Resected	0.74 (0.65–0.82)	1.00 (0.92–1.00)	0.96	84.18 (73.58–94.78)	34.35 (0.00–90.83)
Staining positive					
≥1% <sup>d</sup>	0.76 (0.65–0.84)	1.00 (0.90–1.00)	0.93	83.88 (73.02–94.74)	78.73 (63.35–94.11)

NOTE:

<sup>a</sup>, the sample size ≥ 100;

<sup>b</sup>, the sample size <100;

<sup>c</sup>, the AUC can’t be calculated to get a result;

<sup>d</sup>, cases with more than 1% of tumor cells staining were classified as positive;

SEN, sensitivity; SPE, specificity; AUC, the summary receiver operating characteristics curve;  $I^2$ ,  $I^2$  statistic; 95%CI, 95% confidence interval.

doi:10.1371/journal.pone.0100837.t002

and *K-ras* mutations to classify the subtypes of lung adenocarcinoma. An additional factor that must be considered is that *EMLA-ALK* fusions were detected in only about 5% of lung adenocarcinoma, and it was mutually exclusive to other *EGFR* mutations and *K-ras* mutations [43]. These problems were further compounded by the fact that, molecular testing for these markers was complex and not easy to perform. Similar problems were met in miRNA profiling. miRNA, as noncoding single-stranded RNAs regulating gene expression, were reported as a highly reliable method in distinguishing lung SQCC from AC [11,13]. However, miRNA profiling had high requirements for specimen preservation conditions as miRNA is easily degraded, which presents a challenge for clinical diagnoses. Compared to these markers, testing for TTF-1 and Napsin A not only had a high diagnostic value, but also was easy to perform. This would be a powerful alternative method to distinguish AC from SQCC in lung cancer diagnosis and treatment.

Heterogeneity was statistically significant for both sensitivity and specificity of the combined detection of TTF-1 and Napsin A. Subgroup analyses showed that the stratification of the confounding factors did not remove the heterogeneity for sensitivity. For specificity however, the heterogeneity was not statistically significant in studies that were conducted in Asians, had large sample sizes and studies in which resected specimens were used. Meta-regression suggested that the patient's location contributed the most to the heterogeneity of sensitivity, and the patient's location, sample size and threshold of staining positivity were the sources of heterogeneity for specificity. The heterogeneity among studies conducted in different locations may be attributed most to ethnic differences, which may lead to different expression levels of TTF-1 and Napsin A. The heterogeneity in specificity among studies with different sample sizes may be due to the fact that the entire spectrum of patients in studies with small sample sizes was not representative in most cases. In order to further explore influence of specimen types on the pool sensitivity and specificity of the combined test, we compared pool sensitivity and specificity of the rest seven studies using resected samples only with those of the ten studies as resected specimens and biopsies combined. The results indicated that the pool sensitivity and specificity of the rest seven studies were not significant different from that of the ten studies (0.74 (95%CI: 0.65–0.82) vs. 0.76 (95%CI: 0.69–0.83) and 1.00 (95%CI: 0.92–1.00) vs. 1.00 (95%CI: 0.92–1.00)), which was

consistently found not to be a source of heterogeneity for sensitivity and specificity in meta-regression analysis. Sensitivity analysis showed that the heterogeneity was not statistically significant until the Fatima (2011) study was excluded, supporting the notion that the Fatima (2011) study affected the heterogeneity of specificity the most. This was due to the study's poorly differentiated areas of AC and SQCC.

Although the study provides a comprehensive assessment on the diagnostic value of the combined test of TTF-1 and Napsin A, several precautions should be taken into account when interpreting the results. First, significant inter-study heterogeneity was found. Second, the estimates that we obtained were not adjusted to account for other variables such as tumor size, histologic differentiation, and clinical staging. However, we still observed trends supporting the high diagnostic value of the test in patients with well- and moderately-differentiated lung adenocarcinoma in their tissue sections. Third, population characteristics such as the smoking, sex and age should be put forward in the included studies, so that a more detailed subgroup analyses can be made to explain the heterogeneity.

Overall, our meta-analysis showed that the combined test of TTF-1 and Napsin A was a high diagnosis accuracy alternative diagnostic test in classifying AC and SQCC in NSCLC.

## Supporting Information

**Figure S1 Deeks' funnel Plot Asymmetry test of combined TTF-1 and Napsin A in distinction between AC from SQCC.** Circle presents study; solid line presents regression line. See Figure S1.tif file.

(TIF)

**Table S1** Sensitive analysis for studies included in the meta-analysis. See Table S1.doc file.

(DOC)

**Checklist S1 PRISMA 2009 Checklist.**

(DOC)

## Author Contributions

Conceived and designed the experiments: L. Li SW L. Liu. Performed the experiments: L. Li XLJY. Analyzed the data: L. Li SW. Wrote the paper: L. Li SW L. Liu. Review, and/or revision of the manuscript: SW L. Liu.

## References

- Brown AF, Sirohi D, Fukuoka J, Cagle PT, Policarpio-Nicolas M, et al. (2013) Tissue-preserving antibody cocktails to differentiate primary squamous cell carcinoma, adenocarcinoma, and small cell carcinoma of lung. *Arch Pathol Lab Med* 137:1274–1281.
- Savas P, Hughes B, Solomon B (2013) Targeted therapy in lung cancer: IPASS and beyond, keeping abreast of the explosion of targeted therapies for lung cancer. *J Thorac Dis* 5:S579–S592.
- Noh S, Shim H (2012) Optimal combination of immunohistochemical markers for subclassification of non-small cell lung carcinomas: A tissue microarray study of poorly differentiated areas. *Lung Cancer* 76:51–55.
- Righi L, Graziano P, Fornari A, Rossi G, Barbareschi M, et al. (2011) Immunohistochemical subtyping of nonsmall cell lung cancer not otherwise specified in fine-needle aspiration cytology: A retrospective study of 103 cases with surgical correlation. *Cancer* 117:3416–3423.
- Mukhopadhyay S, Katzenstein AL (2011) Subclassification of non-small cell lung carcinomas lacking morphologic differentiation on biopsy specimens: Utility of an immunohistochemical panel containing TTF-1, napsin A, p63, and CK5/6. *Am J Surg Pathol* 35:15–25.
- Cohen MH, Gootenberg J, Keegan P, Pazdur R (2007) FDA drug approval summary: bevacizumab (Avastin) plus Carboplatin and Paclitaxel as first-line treatment of advanced/metastatic recurrent nonsquamous non-small cell lung cancer. *Oncologist* 12:713–718.
- Fatima N, Cohen C, Siddiqui MT (2011) DOG1 utility in diagnosing gastrointestinal stromal tumors on fine-needle aspiration. *Cancer Cytopathol* 119:202–208.
- Herbst RS (2008) Bevacizumab/chemotherapy in non-small-cell lung cancer: looking for a few good men? *Clin Lung Cancer* 9:75–76.
- Sandler A (2007) Bevacizumab in non small cell lung cancer. *Clin Cancer Res* 13:s4613–s4616.
- Herbst RS, O'Neill VJ, Fehrenbacher L, Belan CP, Bonomi PD, et al. (2007) Phase II study of efficacy and safety of bevacizumab in combination with chemotherapy or erlotinib compared with chemotherapy alone for treatment of recurrent or refractory non small-cell lung cancer. *J Clin Oncol* 25:4743–4750.
- Bishop JA, Benjamin H, Cholakh H, Chajut A, Clark DP, et al. (2013) Accurate classification of non-small cell lung carcinoma using a novel microRNA-based approach. *Clin Cancer Res* 16:610–619.
- Roh MH, Schmidt L, Placido J, Farnen S, Fields KL, et al. (2012) The application and diagnostic utility of immunocytochemistry on direct smears in the diagnosis of pulmonary adenocarcinoma and squamous cell carcinoma. *Diagn Cytopathol* 40:949–955.
- Lebanony D, Benjamin H, Gilad S, Ezagouri M, Dov A, et al. (2013) Diagnostic assay based on hsa-miR-205 expression distinguishes squamous from non-squamous non-small-cell lung carcinoma. *J Clin Oncol* 27:2030–2037.
- Lee HJ, Kim YT, Kang CH, Zhao B, Tan Y, et al. (2013) Epidermal growth factor receptor mutation in lung adenocarcinomas: relationship with CT characteristics and histologic subtypes. *Radiology* 268:254–264.
- Loo PS, Thomas SC, Nicolson MC, Fyfe MN, Kerr KM (2010) Subtyping of undifferentiated non-small cell carcinomas in bronchial biopsy specimens. *J Thorac Oncol* 5:442–447.

16. Stoll LM, Johnson MW, Gabrielson E, Askin F, Clark DP, et al. (2010) The utility of napsin-A in the identification of primary and metastatic lung adenocarcinoma among cytologically poorly differentiated carcinomas. *Cancer Cytopathol* 118:441–449.
17. Tsuta K, Tanabe Y, Yoshida A, Takahashi F, Maeshima AM, et al. (2011) Utility of 10 immunohistochemical markers including novel markers (desmocollin-3, glypican 3, S100A2, S100A7, and Sox-2) for differential diagnosis of squamous cell carcinoma from adenocarcinoma of the Lung. *J Thorac Oncol* 6:1190–1199.
18. Johansson L (2004) Histopathologic classification of lung cancer: Relevance of cytokeratin and TTF-1 immunophenotyping. *Ann Diagn Pathol* 8:259–267.
19. Whithaus K, Fukuoka J, Prihoda TJ, Jagirdar J (2012) Evaluation of napsin A, cytokeratin 5/6, p63, and thyroid transcription factor 1 in adenocarcinoma versus squamous cell carcinoma of the lung. *Arch Pathol Lab Med* 136:155–162.
20. Schultz H, Marwitz S, Baron-Luhr B, Zissel G, Kugler C, et al. (2011) Generation and evaluation of a monoclonal antibody, designated MA<sub>DL</sub>, as a new specific marker for adenocarcinomas of the lung. *Br J Cancer* 105:673–681.
21. Gu K, Shah V, Ma C, Zhang L, Yang M (2007) Cytoplasmic immunoreactivity of thyroid transcription factor-1 (clone 8G7G3/1) in hepatocytes: true positivity or cross-reaction? *Am J Clin Pathol* 128:382–388.
22. Turner BM, Cagle PT, Sainz IM, Fukuoka J, Shen SS, et al. (2012) Napsin A, a new marker for lung adenocarcinoma, is complementary and more sensitive and specific than thyroid transcription factor 1 in the differential diagnosis of primary pulmonary carcinoma: evaluation of 1674 cases by tissue microarray. *Arch Pathol Lab Med* 136:163–171.
23. Ueno T, Linder S (2003) Elmerberger G. Aspartic proteinase napsin is a useful marker for diagnosis of primary lung adenocarcinoma. *Br J Cancer* 88:1229–1233.
24. Tacha D, Yu C, Bremer R, Qi W, Haas T (2012) A 6-antibody panel for the classification of lung adenocarcinoma versus squamous cell carcinoma. *Appl Immunohistochem Mol Morphol* 20:201–207.
25. Kim MJ, Shin HC, Shin KC, Ro JY (2013) Best immunohistochemical panel in distinguishing adenocarcinoma from squamous cell carcinoma of lung: tissue microarray assay in resected lung cancer specimens. *Ann Diagn Pathol* 17:85–90.
26. Stahlman MT, Gray ME, Whitsett JA (1996) Expression of thyroid transcription factor-1(TTF-1) in fetal and neonatal human lung. *J Histochem Cytochem* 44:673–678.
27. Fatima N, Cohen C, Lawson D, Siddiqui MT (2011) TTF-1 and Napsin A double stain: a useful marker for diagnosing lung adenocarcinoma on fine-needle aspiration cell blocks. *Cancer Cytopathol* 119:127–133.
28. Pereira TC, Share SM, Magalhaes AV, Silverman JF (2011) Can we tell the site of origin of metastatic squamous cell carcinoma? An immunohistochemical tissue microarray study of 194 cases. *Appl Immunohistochem Mol Morphol* 19:10–14.
29. Sethi S, Geng L, Shidham VB, Archuleta P, Bandyopadhyay S, et al. (2012) Dual color multiplex TTF-1 + Napsin A and p63 + CK5 immunostaining for subcategorizing of poorly differentiated pulmonary non-small carcinomas into adenocarcinoma and squamous cell carcinoma in fine needle aspiration specimens. *Cytojournal* 9:10.
30. Johnson H, Cohen C, Fatima N, Duncan D, Siddiqui MT (2012) Thyroid transcription factor 1 and napsin a double stain: utilizing different vendor antibodies for diagnosing lung adenocarcinoma. *Acta Cytol* 56:596–602.
31. Yang M, Nonaka D (2010) A study of immunohistochemical differential expression in pulmonary and mammary carcinomas. *Mod Pathol* 23:654–661.
32. Collins BT (2013) Endobronchial ultrasound fine-needle aspiration biopsy of pulmonary non-small cell carcinoma with subclassification by immunohistochemistry panel. *Cancer Cytopathol* 121:146–154.
33. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J (2003) The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 3:25.
34. Aikawa E, Kawahara A, Hattori S, Yamaguchi T, Abe H, et al. (2011) Comparison of the expression levels of napsin A, thyroid transcription factor-1, and p63 in nonsmall cell lung cancer using cytocentrifuged bronchial brushings. *Cancer Cytopathol* 119:335–345.
35. Brunnstrom H, Johansson L, Jirstrom K, Jonsson M, Jonsson P, et al. (2013) Immunohistochemistry in the differential diagnostics of primary lung cancer: an investigation within the southern Swedish lung cancer study. *Am J Clin Pathol* 140:37–46.
36. Yanagita E, Imagawa N, Ohbayashi C, Itoh T (2011) Rapid multiplex immunohistochemistry using the 4-antibody cocktail YANA-4 in differentiating primary adenocarcinoma from squamous cell carcinoma of the lung. *Appl Immunohistochem Mol Morphol* 19:509–513.
37. Zhang P, Han YP, Huang L, Li Q, Ma DL (2010) Value of napsin A and thyroid transcription factor-1 in the identification of primary lung adenocarcinoma. *Oncol Lett* 1:899–903.
38. Stang A, Pohlabein H, Muller KM, Jahn I, Giersiepen K, et al. (2006) Diagnostic agreement in the histopathological evaluation of lung cancer tissue in a population-based case-control study. *Lung Cancer* 52:29–36.
39. Li AR, Chitale D, Riely GJ, Pao W, Miller VA, et al. (2008) EGFR mutations in lung adenocarcinomas: clinical testing experience and relationship to EGFR gene copy number and immunohistochemical expression. *J Mol Diagn* 10:242–248.
40. Yoshida A, Tsuta K, Nakamura H, Kohno T, Takahashi F, et al. (2011) Comprehensive histologic analysis of ALK-rearranged lung carcinomas. *Am J Surg Pathol* 35:1226–1234.
41. Nicholson AG, Gonzalez D, Shah P, Pynegar MJ, Deshmukh M, et al. (2010) Refining the diagnosis and EGFR status of non-small cell lung carcinoma in biopsy and cytologic material, using a panel of mucin staining, TTF-1, cytokeratin 5/6, and P63, and EGFR mutation analysis. *J Thorac Oncol* 5:436–441.
42. Hamamoto J, Socjima K, Yoda S, Naoki K, Nakayama S, et al. (2013) Identification of microRNAs differentially expressed between lung squamous cell carcinoma and lung adenocarcinoma. *Mol Med Rep* 8:456–462.
43. Sakai Y, Nakai T, Ohbayashi C, Imagawa N, Yanagita E, et al. (2013) Immunohistochemical profiling of ALK fusion gene-positive adenocarcinomas of the lung. *Int J Surg Pathol* 21:476–482.