RESEARCH ARTICLE

Aberrant Methylation of Genes in Sputum Samples as Diagnostic Biomarkers for Non-small Cell Lung Cancer: a Meta-analysis

Xu Wang[1], Li Ling[2], Hong Su[1]*, Jian Cheng[1], Liu Jin[1]

Abstract

Background: We aimed to comprehensively review the evidence for using sputum DNA to detect non-small cell lung cancer (NSCLC). Materials and Methods: We searched PubMed, Science Direct, Web of Science, Chinese Biological Medicine (CBM), Chinese National Knowledge Infrastructure (CNKI), Wanfang, Vip Databases and Google Scholar from 2003 to 2013. The meta-analysis was carried out using a random-effect model with sensitivity, specificity, diagnostic odd ratios (DOR), summary receiver operating characteristic curves (ROC curves), area under the curve (AUC), and 95% confidence intervals (CI) as effect measurements. Results: There were twenty-two studies meeting the inclusion criteria for the meta-analysis. Combined sensitivity and specificity were 0.62 (95%CI: 0.59-0.65) and 0.73 (95%CI: 0.70-0.75), respectively. The DOR was 10.3 (95%CI: 5.88-18.1) and the AUC was 0.78. Conclusions: The overall accuracy of the test was currently not strong enough for the detection of NSCLC for clinical application. Dscovery and evaluation of additional biomarkers with improved sensitivity and specificity from studies rated high quality deserve further attention.

Keywords: NSCLC - methylation - sputum - diagnostic - meta-analysis.

Asian Pac J Cancer Prev, **15 (11)**, 4467-4474

Introduction

Lung cancer killed over one million people worldwide every year, and as the leading cause of cancer death in men and second leading cause in women, the significance of this worldwide public health burden was evident (Molina et al., 2008). In United States, lung cancer is the number one cancer killer for both men and women, leading to over 160,000 deaths each year (Jemal et al., 2007). Lung cancer is divided into two sub-types clinically, small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). SCLC is the more aggressive sub-type, and accounts for 10-15% of all cases. The remaining 85-90% of cases is classified as NSCLC. Early detection of NSCLC, which is the more common and less aggressive sub-type, has the highest potential for saving lives (Paul et al., 2008).

At the present time, various screening options are available for the early detection of NSCLC, including chest X-ray (Gavelli et al., 2000), sputum cytology (Bach et al., 2003), low dose spiral computed tomography (LDSCT) (Carter et al., 2007), auto-fluorescence bronchoscopy (AFB) (Feller et al., 2005) and so on (Ziaian et al., 2014). However, none of these methods is truly optimal, either on account of improper sensitivity or specificity of the tests, or the methods are costly and invasive (Melvyn et al., 2000; Paul et al., 2008). Since these options have not been proven effective as early detection methods, sensitive and specific diagnostic methods remains to be found (Parkin et al., 2001).

To fill this void, research focus has been moved to molecular approaches (Suzuki et al., 2008; Ramshankar et al., 2013). The goal was to identify molecular biomarkers (generally DNA) that can be utilized for early detection of these lesions at the pre-invasive stage. Initial DNA methylation studies for NSCLC focused on single gene that was chosen because of its potential function in cancer (Jarmalaite et al., 2003; Xie et al., 2006). However, single DNA methylation marker cannot be expected to detect all cases of a particular cancer. The way to solve this problem is to discuss the DNA methylation status of multiple loci (a panel). Further studies employ panels with more than one loci for DNA methylation profiling to detect a designated cancer (Tsou et al., 2007; Feng et al., 2008).

Many studies have shown that sputum could be a promising "remote medium" for early detection of NSCLC (Miozzo et al., 1996; Palmisano et al., 2000; Belinsky et al., 2005). The advantages of sputum as "remote media" included its non-invasive procurement, and the fact that it contains cells from the lungs and lower respiratory tract (Olaussen et al., 2005; Paul et al., 2008). Shin et al. (2012)

[1]Department of Epidemiology and Biostatistics, School of Public Health, [2]Department of Social Medicine and Health Service Administration, School of Health Management, Anhui Medical University, 81 Meishan Road, Hefei 230032, Anhui Province, China
**For correspondence: hongsu20135151@163.com.*

described a combination of 2 loci, MAGE A3 and p16 that served as a good panel for early detection of lung cancer in sputum specimen. Similarly, Li et al. (2007) reported a combination of FHIT and HYAL2 in sputum samples with 76% sensitivity and 85% specificity. Hwang et al. (2011) recommended HOXA9 gene methylation in sputum, the sensitivity was 90.5% and the specificity was 97.5%, and Belinsky (2006) showed that the combined effect of methylation of at least one of the four most significant genes in sputum increased the positive predictive value to 86%. In contrast, Cirincione et al. (2006) reported that 3 loci, RARβ2, p16 and RASSF1A genes in sputum had a limited diagnostic value in early detection of lung cancer.

During the past years, an increasing number of researches had been published utilizing aberrant methylation of sputum DNA as diagnostic biomarkers for NSCLC (Georgios et al., 2012; Skin et al., 2012; Hubers et al., 2013). However, the results of these studies were variable and inconsistent. To the best of our knowledge, there was still no comprehensive evaluation on the diagnostic accuracy of methylation markers in sputum samples for NSCLC. Hence, we performed a comprehensive review on the diagnostic value of sputum DNA testing for patients with NSCLC.

Materials and Methods

Search strategy and selection criteria

The PubMed, Science Direct, Web of Science, Chinese Biological Medicine (CBM), Chinese National Knowledge Infrastructure (CNKI), Wanfang, Vip Databases and Google Scholar were systematically searched by two authors (Wang and Ling). Key words including "lung cancer or lung carcinoma or non-small cell lung cancer or NSCLC", "sputum or flema", "diagnostic", "sensitivity and specificity" and "methylation or hypermethylation or hypomethylation or demethylation" were used to identify appropriate studies published in English and Chinese from Jan 1, 2003 to 2013. In addition, the reference lists of all identified studies were manually searched to identify any additional studies. Duplicated results, irrelevant articles were removed from this study. To be eligible for inclusion, studies had to utilize one or more methylated biomarkers in sputum samples for detecting NSCLC.

Two reviewers (Wang and Ling) independently evaluated the full text of each manuscript. Studies were chosen to investigated the association between sputum DNA with patients' diagnosis by histopathology (endoscopy) and provided data on the numbers of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN). Articles were excluded if date were insufficient to calculate the numbers of TP, FP, TN and FN; subjects were enrolled without a diagnosis; non-case-control studies, this study was based on tissue or animals, the study's purpose was to evaluate technical or mechanical aspects of the NSCLC detection assay, review article or letter, single case report, or conference summary or memorandum.

Data extraction procedure and quality assessment

Two reviewers (Wang and Ling) independently extracted the following data from each article: name of first author, year of publication, source populations, number of patients and controls included, TP, FP, FN, TN, detection techniques used, target gene (s), the score of the quality assessment of studies of diagnostic accuracy (QUADAS) and studies with or without blinded. Any disagreements between the reviewers were resolved by discussion.

The standards for reporting of diagnostic accuracy (STARD) initiative and QUADAS guidelines were utilized to assess the methodological quality of each study (Bossuyt et al., 2003; Whiting et al., 2004). There are 25 items in the STARD initiative checklist, and a score of 1 was given when the item was yielded (Bossuyt et al., 2003). While 14 items were included in the QUADAS tool, whereby a score of 1 was given when a specific item was fulfilled, 0 if this item was unclear and -1 for the item not achieved (Whiting et al., 2004). All of these studies were evaluated independently and discussed by the reviewers until a consensus was reached.

Statistical methods

We used standard methods recommended for meta-analysis of diagnostic test evaluations (Deville et al., 2004). Analyses were performed by using two statistical software programs (Meta-Disc 1.4 for Windows and Stata, version 10.0). Pooled estimates on sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR) and DOR were employed to examine the diagnostic accuracy of sputum DNA testing. Spearman's correlation coefficient was assessed to determine the threshold effect (Lijmer et al., 1999). The heterogeneity amongst studies was assessed on the basis of the χ^2 test using the Cochran Q statistic. The I^2 statistic, which measures the extent of inconsistency between studies, was also assessed (Deville et al., 2002).

For further explore heterogeneity, subgroup analyses were conducted. Separate analyses investigated the effect of source populations (Asians and non-Asians), the classification of control groups (cancer-free smokers, patients with pulmonary benign disease and both of all), studies with or without blindness (blind method and not mentioned), study quality (high quality: QUQDAS ≥8, medium quality: QUADAS =7 and low quality: QUADAS ≤6), assay method (qualitative and quantitative) and biomarkers status (single genes and multiple genes), given their potential impact on test performance.

Publication bias was detected by the Deeks' funnel plot asymmetry test. A P value of less than 0.1 for the slope coefficient was considered as significant asymmetry, which indicated potential publication bias (Pai et al., 2003).

Results

Baseline study characteristics

The systematic literature search yielded a total of 22 studies including a total of 1208 patients and 1200 controls for final analysis (Destro et al., 2004; Zhang et al., 2004; Konno et al., 2004; Wang et al., 2004; Belinsky et al., 2005; Olaussen et al., 2005; Wang et al., 2006; Cirincione et al., 2006; Georgiou et al., 2007; Hsu et al.,

DOI:http://dx.doi.org/10.7314/APJCP.2014.15.11.4467

Table 1. Baseline Characteristics of All Included Studies

Study/year	Country	Patients/controls (n)	Control groups	Biomarkers	Assay method	No.of TP	No.of FP	No.of FN	No.of TN	Blinded	STARD	QUADAS
Destro/2004	Italy	24/100	cancer-free smokers	p16	MSP	16	4	8	96	Yes	21	12
Konno/2004	Japan	78/94	cancer-free smokers	p16/APC/RARβ	MSP	44	20	34	74	Unknown	20	11
Wang/2004	China	34/21	Pulmonary benign disease	p16	PCR	11	0	23	21	Yes	16	7
Belinsky /2005	USA	53/118	cancer-free smokers	p16/RASSF1A/MGMT /DAPK/H-cadherin /PAX5α/PAX5β	MSP	45	77	8	41	Yes	20	9
Olaussen /2005	Republic of Korea	22/56	cancer-free smokers	HOX A9/p16 /MAGE A1/MAGE B2	MSP	21	12	1	44	Unknown	18	8
Cirincinone /2006	Italy	18/112	cancer-free smokers	RARβ2/p16/RASSF1A	MSP	9	69	9	43	Yes	15	7
Wang/2006	Taiwan	79/22	cancer-free smokers	FHIT/p16/RARβ	MSP	64	6	15	16	Unknown	12	4
Georgiou /2007	Greece	80/40	cancer-free smokers	p16	MSP	55	9	25	31	Unknown	16	9
Hsu/2007	Taiwan	82/37	cancer-free smokers	FHIT/p16/RARβ	fluorescent nested-MSP	67	9	15	28	Unknown	16	7
Shivapurkar /2007	USA	13/23	cancer-free smokers + pulmonary benign disease	3-OST-2/RASSF1A /P16/APC	quantitative realtime PCR	8	0	5	23	Yes	22	13
Liu/2008	China	58/107	cancer-free smokers	p16	MSP	41	55	17	52	Unknown	18	10
Van der Dirft /2008	The Netherlands	28/68	COPD	RASSF1A	real-time β-globin PCR	13	1	15	67	Unknown	19	11
Hwang/2011	Republic of Korea	76/109	cancer-free smokers + pulmonary benign disease	HOXA9	MSP	54	49	22	60	Unknown	20	9
Hubers/2012	The Netherlands	53/47	COPD	RASSF1A/APC/CYGB	qMSP	34	4	19	43	Yes	18	10
Zhang/2004	China	44/20	cancer-free smokers + pulmonary benign disease	p16	MSP	27	3	17	17	Yes	15	7
Guo/2008	China	100/50	cancer-free smokers	p16	MSP	61	0	39	50	Unknown	14	6
Cao/2008	China	34/21	pulmonary benign disease	p16	PCR	11	0	23	21	Yes	15	7
Hu/2009	China	42/25	pulmonary benign disease	p16	MSP	20	0	22	25	Yes	16	9
Peng/2010	China	82/25	pulmonary benign disease	RASSF1A/p16/DAPK	MSP	60	0	22	25	Unknown	18	6
Kang/2011	China	47/24	pulmonary benign disease	FHIT/p16/MGMT /RASSF1A/APC	MSP	38	0	9	24	Unknown	20	8
Zhang/2012	China	41/15	cancer-free smokers + pulmonary benign disease	p16	MSP	21	2	20	13	Unknown	17	6
Sun/2012	China	120/66	cancer-free smokers	p16/RASSF1A	MSP	32	8	88	58	Unknown	19	7

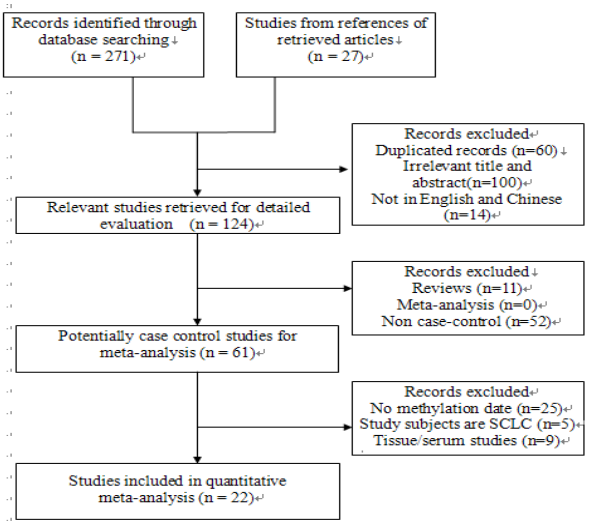

Figure 1. Flow Diagram of Studies of Studies through the Review Process

2007; Shivapurkar et al., 2007; Liu et al., 2008; Van der Drift et al., 2008; Guo et al., 2008; Cao et al., 2008; Hu et al., 2009; Peng et al., 2010; Kang et al., 2011; Hwang et al., 2011; Zhang et al., 2012; Hubers et al., 2013; Sun et al., 2013) (Figure 1). The baseline characteristics of the included studies were presented in Table 1. The studies originated from 8 countries or regions (including Italy, Japan, USA, Republic of Korea, Greece, China, The Netherlands and Taiwan) and they were published from 2003 to 2013. The sample size of these studies ranged from 13 to 120 individuals. Eleven studies (Zhang et al., 2004; Destro et al., 2004; Wang et al., 2004; Georgiou et al., 2007; Liu et al., 2008; Vander et al., 2008; Guo et al., 2008; Cao et al., 2008; Hu et al., 2009; Hwang et al., 2011; Zhang et al., 2012) evaluated methylation of a single gene as a diagnostic biomarker, while other studies evaluated methylation of multiple genes. Four studies (Zhang et al., 2004; Shivapurkar et al., 2007; Hwang et al., 2011; Zhang et al., 2012) evaluated controls from both cancer-free individual and pulmonary benign disease; seven studies (Wang et al., 2004; Van der Drift et al., 2008; Cao et al., 2008; Hu et al., 2009; Peng et al., 2010; Kang et al., 2011; Hubers et al., 2013) focused on patients with pulmonary benign disease only. Whereas, the rest of studies investigated cancer-free smokers only. Methylation-specific PCR (MSP) was used for biomarker detection in 16 studies (Zhang et al., 2004; Destro et al., 2004; Konno et al., 2004; Belinsky et al., 2005; Olaussen et al., 2005; Cirincione et al., 2006; Wang et al., 2006; Georgiou et al., 2007; Liu et al., 2008; Guo et al., 2008; Hu et al., 2009; Peng et al., 2010; Kang et al., 2011; Hwang et al., 2011; Zhang et al., 2012; Sun et al., 2013), quantitative methylation-specific PCR (qMSP, real time PCR) was used in three studies (Shivapurkar et al., 2007; Van der Drift et al., 2008; Hubers et al., 2013); the PCR-based methylation analysis in two studies (Wang et al., 2004; Cao et al., 2008); the fluorescent nested-MSP in one study (Hsu et al., 2008);

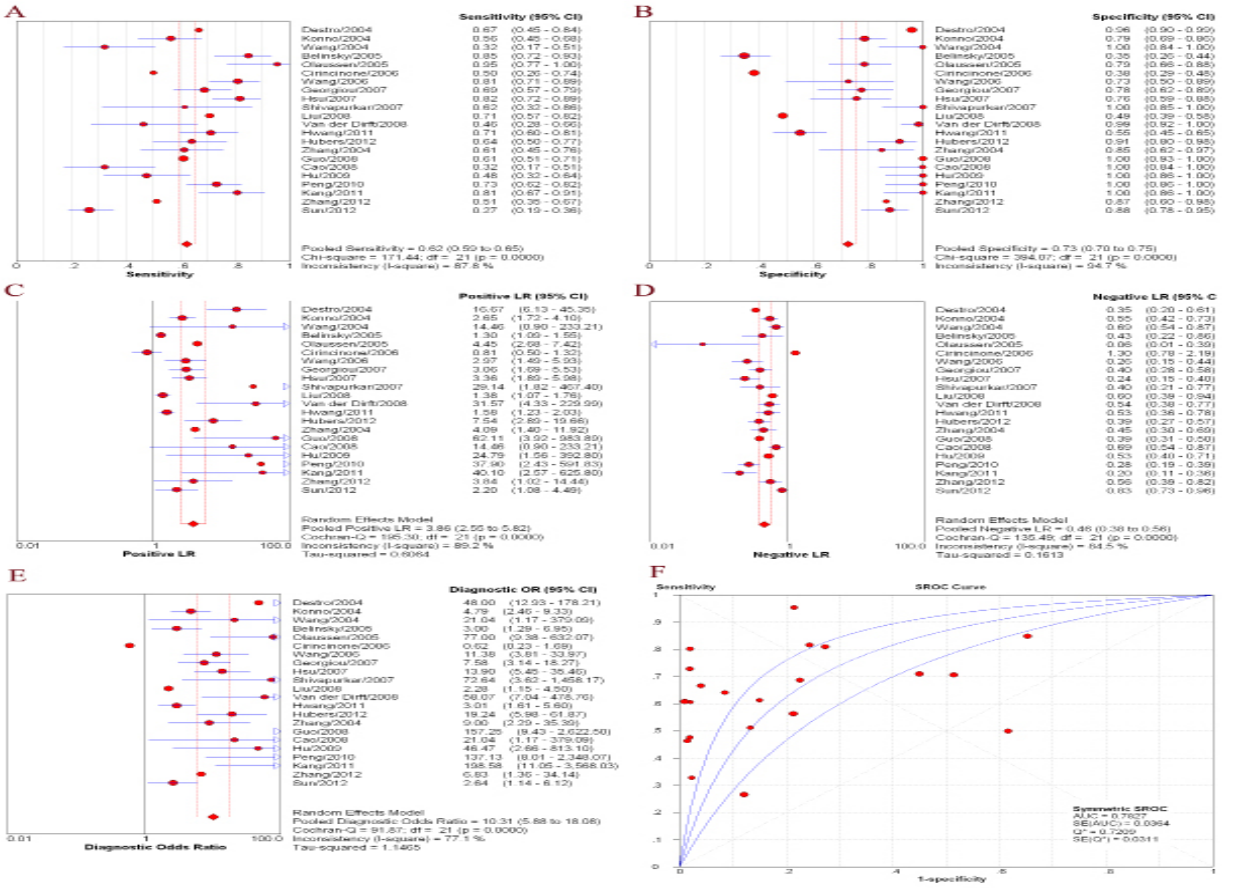

Figure 2. A) Forest plots of sensitivity of sputum DNA testing in NSCLC; **B)** Forest plots of specificity of sputum DNA testing in NSCLC; **C)** Forest plots of PLR of sputum DNA testing in NSCLC; **D)** Forest plots of NLR of sputum DNA testing in NSCLC; E) Forest plots of DOR of sputum DNA testing in NSCLC; **F)** SROC curves for sputum DNA testing for the detection of NSCLC.

Table 2. Subgroup Analysis of Sputum DNA Testing in the Detection of NSCLC

subgroup	studies, (n)	pooled sensitivity (95%CI)	pooled specificity (95%CI)	positive likelihood ratio (95%CI)	negative likelihood (95%CI)	ratiopooled DOR (95%CI)	AUC
source populations							
Asian	15	0.609(0.577-0.641)	0.763(0.730-0.794)	3.624(2.280-5.760)	0.451(0.355-0.573)	10.200(5.408-19.238)	0.806
non-Asians	7	0.669(0.609-0.725)	0.677(0.635-0.718)	4.664(1.746-12.458)	0.491(0.359-0.672)	10.269(2.962-35.608)	0.752
Control groups							
cancer-free smokers	11	0.637(0.600-0.673)	0.665(0.631-0.697)	2.681(1.709-4.208)	0.453(0.321-0.640)	6.827(3.324-14.021)	0.784
pulmonary benign disease	7	0.584(0.528-0.639)	0.978(0.950-0.993)	13.230(6.519-26.850)	0.454(0.332-0.622)	34.233(15.382-76.188)	0.885
both of all	4	0.632(0.556-0.704)	0.677(0.600-0.747)	3.439(1.212-9.759)	0.503(0.406-0.622)	6.429(2.356-17.543)	0.726
Biomarkers status							
single genes	11	0.588(0.546-0.629)	0.786(0.751-0.819)	5.294(2.661-10.533)	0.524(0.452-0.607)	11.607(5.195-25.934)	0.735
multiple genes	11	0.652(0.614-0.689)	0.671(0.633-0.708)	3.313(1.810-6.062)	0.392(0.258-0.594)	9.715(4.180-22.580)	0.817
Assay method							
qualitative method	18	0.610(0.580-0.640)	0.694(0.664-0.722)	3.231(2.124-4.913)	0.483(0.391-0.598)	8.403(4.581-15.413)	0.771
quantitative method	4	0.693(0.619-0.760)	0.920(0.869-0.956)	7.963(2.884-21.990)	0.389(0.276-0.548)	19.458(9.928-38.139)	0.861
Blind method							
blind method	7	0.599(0.537-0.660)	0.582(0.530-0.633)	3.876(1.428-10.521)	0.555(0.420-0.734)	8.072(2.267-28.734)	0.728
not mentioned	15	0.629(0.597-0.660)	0.790(0.761-0.817)	4.064(2.551-6.474)	0.424(0.328-0.548)	11.725(6.181-22.242)	0.811
Study quality							
low	4	0.682(0.626-0.734)	0.929(0.864-0.969)	8.706(1.717-44.139)	0.364(0.264-0.502)	24.250(5.117-114.92)	0.825
medium	6	0.473(0.418-0.528)	0.679(0.620-0.733)	2.762(1.191-6.401)	0.632(0.462-0.865)	5.236(1.575-17.408)	0.692
high	12	0.678(0.638-0.716)	0.715(0.683-0.746)	3.876(2.355-6.381)	0.446(0.374-0.533)	11.536(5.639-23.600)	0.793

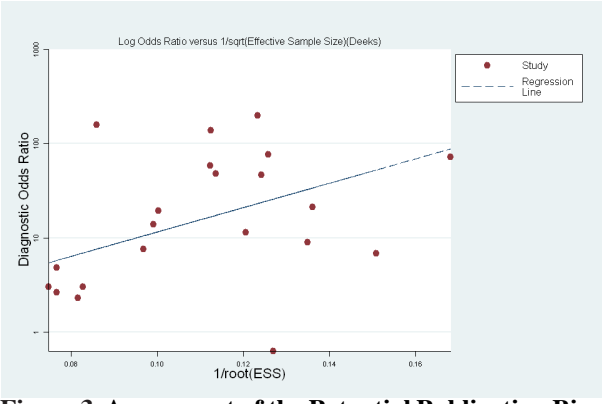

Figure 3. Assessment of the Potential Publication Bias in the Detection of NSCLC

Methodological quality of included studies

Quality assessment based on QUADAS guidelines was conducted on all 22 studies. Twelve of these studies (Destro et al., 2004; Konno et al., 2004; Belinsky et al., 2005; Olaussen et al., 2005; Georgiou et al., 2007; Shivapurkar et al., 2007; Liu et al., 2008; Vander et al., 2008; Hu et al., 2009; Kand et al., 2011 Hwang et al., 2011; Hubers et al., 2013) had QUADAS score ≥ 8, six (Wang et al., 2004; Zhang et al., 2004; Cirincione et al., 2006; Hus et al., 2007; Cao et al., 2008; Sun et al., 2013) had QUADAS score =7 and four (Wang et al., 2006; Guo et al., 2008; Peng et al., 2010; Zhang et al., 2012) had a QUADAS score ≤ 6.

Threshold effect

Computation of the spearman correction coefficient between the logit of sensitivity and that of 1-specificity of sputum DNA testing was 0.432 (p=0.045).

Diagnostic accuracy analyses

For all studies, the pooled DOR was 10.31 (95%CI: 5.88 to 18.08), Cochran Q =91.87 (p=0.000) and I^2 = 77.10%. There appeared to be heterogeneity between studies, as assessed by inspection of the forest plot (Figure 2E). Figure 2F presented the symmetrical ROC curve of sputum DNA testing and the AUC was 0.7827. The meta-analysis revealed an overall sensitivity of 0.62 (95%CI: 0.59-0.65) and a specificity of 0.73 (95%CI: 0.70-0.75) for methylated genes in sputum samples (Figure 2A-B). The PLR was 3.86 (95%CI: 2.55-5.82), the NLR was 0.46 (95%CI: 0.38-0.56) (Figure 2C-D).

Subgroup analysis of the classification of control groups

sROC curve analysis of controls of cancer-free smokers produced sensitivity of 0.637 (95%CI: 0.600 to 0.673), specificity of 0.665 (95%CI: 0.631 to 0.697) and AUC of 0.784. The corresponding values for the patients with pulmonary benign disease were 0.584 (95%CI: 0.528 to 0.639) for sensitivity, 0.978 (95%CI: 0.950 to 0.993) for specificity and 0.885 for AUC.

The results of both cancer-free smokers and patients with pulmonary benign disease showed a sensitivity of 0.632 (95%CI: 0.556 to 0.704), a specificity of 0.677 (95%CI: 0.600 to 0.747) and an AUC of 0.726 (see Table 2).

Subgroup analysis of the study quality

In lower study quality subgroup, the sensitivity of 0.682 (95%CI: 0.626-0.734), specificity of 0.929 (95%CI: 0.8640.969) and AUC of 0.825 were achieved for detection of NSCLC. The corresponding values of the subgroup with higher study quality were 0.678 (95%CI: 0.638 to 0.716) for sensitivity, 0.715 (95%CI: 683 to 0.746) for specificity and 0.793 for AUC. It showed sensitivity of 0.473 (95%CI: 0.418 to 0.528), specificity of 0.679 (95%CI: 0.620 to 0.733) and AUC of 0.692 in the medium study quality subgroup (Table 2).

Publication bias was evaluated by using the Deeks' test. It showed no significant publication bias among the studies that evaluated biomarkers in sputum samples from NSCLC patients (Figure 3).

Discussion

Lung cancer is responsible for a million cancer deaths per year worldwide, and its detrimental effects will continue to increase (Paul et al., 2008). As yet, no effective

approach for early detection was one of the important reasons for the high lung cancer mortality (Gavelli et al., 2000). In the present meta-analysis, we found that methylated genes in sputum samples for the early detection of USCLC yielded an overall sensitivity of 62% and an overall specificity of 73%. The AUC was 0.783, indicating an accuracy of middle level. Furthermore, the PLR was 3.86, NLR was 0.46 and DOR value was 10.31. Taken all together, it indicated that overall accuracy of USCLC detection utilizing sputum DNA testing was not good enough.

The results of the subgroup analysis recommended the high diagnostic value of multiple markers. Therefore, 11 studies that involved single markers appeared to have lower sensitivity and AUC. In addition, our results noted that the accuracy of quantitative method for the detection of NSCLC was higher than those routinely qualitative analysis.

Subgroup analysis of source populations showed that researches on Asians had higher specificity and AUC than Non-Asians (Europe and America). It appeared that lung cancer in Asians had unique characteristics (Federico et al., 2010). In Asian group, lung cancer often occured at an earlier age, was more common in people who had never smoked, and had a better overall prognosis (Jiang et al., 2012). To our surprise, there was a lack of studies that focused on Africans. The reasons might be that the incidence of lung cancer was low and sputum DNA testing had not been popular investigated in most African countries (Jacques et al., 2010). However, the incidence of lung cancer is increasing worldwidely, environmental exposure to asbestos, a dusty occupation, and perhaps indoor air pollution may also contribute to the development of lung cancer in African (Abdul et al., 2010). Therefore, early detection remained the key to successful outcomes (Claudia et al., 1999), and the participation of scientists around the world especially African areas was always required.

The heterogeneity had decreased when control groups were divided into cancer-free smokers, patients with pulmonary benign disease or both of them. Most of the recent diagnostic guidelines had concluded the diagnostic testing compared the results of the index test in patients with an established diagnosis of the target condition with its results in healthy controls or controls with other diagnosis (Lijmer et al., 1999). The present meta-analysis noted that control groups of only cancer-free smokers had the highest value for early detection, patients of the pulmonary benign disease were the lowest. Our results indicated that the diagnostic accuracy might be over- or under-estimated respectively in sputum DNA testing with only healthy controls or patients of pulmonary benign disease (Anne et al., 2006). Therefore, screening programs should pay more attention on the selection of the controls in order to assess diagnostic value accurately.

Another important factor that influenced the diagnostic value of sputum DNA testing was the quality index of the selected studies. Methodology checklist of diagnostic test accuracy covered participants representative, selection criteria, selection method, blind method and so on (Penny et al., 2011). The present meta-analysis concluded high quality studies had higher sensitivity than the low and medium quality studies. Therefore, these findings indicated that robust methodology design was significant for the diagnostic test (Brian et al., 2011). In addition, the presented study suggested that diagnostic accuracy had been overestimated in some low quality studies.

After systematic review of 22 studies, we identified several common limitations and insufficiency. Firstly, many studies were flawed by choosing controls in the wrong way. An ideal diagnostic test should recruited from both healthy controls and patients with pulmonary benign disease at the same time (Whiting et al., 2004). In these studies, 11 of 22 studies only selected cancer-free smokers as controls. However, defective controls would likely lead to over-estimations of specificity (Lijmer et al., 1999; Pai et al., 2003). Furthermore, many of the available studies did not report on the blind method (Jadad et al., 1996). The present meta-analysis revealed that the absence of blind method and low-quality study design would likely lead to over-estimations of diagnostic accuracy. Lastly, most studies suffered from a small sample size, as only 6 studies had a sample size greater than 150. Small sample size problem was a serious limitation when interpreting the findings and increased the potential bias of data (Gordon et al., 2011).

In conclusion, this was the first meta-analysis about sputum DNA testing and NSCLC. The current evidence suggested that the diagnostic accuracy of aberrant methylation of genes in sputum samples was not lower than single biomarkers for NSCLC, at least. However, the overall accuracy of the test was currently not strong enough to be the detection of NSCLC for clinical application. The discovery and evaluation of additional biomarkers with improved sensitivity and specificity from studies rated high quality deserved further investigation.

References

Abdul RJ, Hanaa B, Ahmet D, et al (2010). Modification and implementation of NCCN guidelines on non-small cell lung cancer in the middle east and north africa region. *J Natl Compr Canc Netw*, **8**, 16-21.

Anglim, Paul P, Todd AA, Ite ALO (2008). DNA methylation-based biomarkers for early detection of non-small cell lung cancer: an update. *Mol Cancer*, **7**, 81-89.

Anne WS, Johannes BR, Marcello DN (2006). Evidence of bias and variation in diagnostic accuracy studies. *CMAJ*, **4**, 469-76.

Bach PB, Kelley MJ, Tate RC, McCrory DC (2003). Screening for lung cancer: a review of the current literature. *Chest*, **123**, 72-82.

Belinsky SA, Klinge DM, Dekker JD, et al (2005). Gene promoter methylation in plasma and sputum increases with lung cancer risk. *Clin Cancer Res*, **11**, 6505-11.

Belinsky SA, Liechty KC, Gentry FD, et al (2006). Promoter hypermethylation of multiple genes in sputum precedes lung cancer incidence in a high-risk cohort. *Cancer Res*, **6**, 3338-44.

Bossuyt PM, Reitsma JB, Bruns DE, et al (2003). Towards complete and accurate reporting of studies of diagnostic accuracy. *BMJ*, **326**, 41-4.

Brian HW, Muireann Q (2011). The assessment of the quality of reporting of meta-analyses in diagnostic research: a systematic review. *BMC Med Res Methodol*, **11**, 163-74.

Cao AH, Wang X, Shi HJ (2008). The value of CT scan and tumor markers in sputum for early diagnosis of peripheral lung cancer. *J Chongqing Medical University*, **10**, 1193-5.

Carter D, Vazquez M, Flieder DB, et al (2007). Comparison of pathologic findings of baseline and annual repeat cancers diagnosed on CT screening. *Lung Cancer*, **56**, 193-9.

Cirincione R, Lintas C, Conte D, et al (2006). Methylation profile in tumor and sputum samples of lung cancer patients detected by spiral computed tomography: a nested case-control study. *Int J Cancer*, **5**, 1248-53.

Claudia H, Dorothy IM, David FY, et al (1999). Early Lung Cancer Action Project: overall design and findings from baseline screening. *The Lancet*, **354**, 99-105.

Destro A, Bianchi P, Alloisio M, et al (2004). K-ras and p16 (INK4A) alterations in sputum of NSCLC patients and in heavy asymptomatic chronic smokers. *Lung Cancer*, **44**, 23-32.

Deville WL, Buntinx F, Bouter LM, et al (2002). Conducting systematic reviews of diagnostic studies: didactic guidelines. *BMC Med Res Methodol*, **2**, 9-17.

Federico C, Tudor C, Lilia S, et al (2010). Erlotinib as maintenance treatment in advanced non-small-cell lung cancer: a multicentre, randomized, placebo-controlled phase 3 study. *Lancet Oncol*, **11**, 521-9.

Feller-Kopman D, Lunn W, Ernst A (2005). Autofluorescence bronchoscopy and endobronchial ultrasound: a practical review. *Ann Thorac Surg*, **80**, 2395-401.

Feng Q, Hawes SE, Stern JE, et al (2008). DNA methylation in tumor and matched normal tissues from non-small cell lung cancer patients. *Cancer Epidem Biomar*, **17**, 645-54.

Gavelli G, Giampalma E (2000). Sensitivity and specificity of chest X-ray screening for lung cancer: review article. *Cancer*, **89**, 2453-6.

Georgios N, Olaide YR, Soultana M (2012). DNA methylation biomarkers offer improved diagnostic efficiency in lung cancer. *Cancer Res*, **72**, 5692-701;

Georgiou E, Valeri R, Tzimagiorgis G, et al (2007). Aberrant p16 promoter methylation among Greek lung cancer patients and smokers: correlation with smoking. *Eur J Cancer Prev*, **16**, 396-402.

Gordon HG, Andrew DO, Regina K (2011). GRADE guidelines 6. Rating the quality of evidence-imprecision. *J Clin Epidemiol*, **64**, 1283-93.

Guo XJ, Shen L (2008). Diagnostic value of combinative detection of hypermethylation of p16 gene and cytokeratin 19 fragment and CA15-3 antigen for non-small cell lung cancer patients. *Clinical Focus*, **23**, 1067-70.

Hsu HS, Chen TP, Wen CK, et al (2007). Multiple genetic and epigenetic biomarkers for lung cancer detection in cytologically negative sputum and a nested case-control study for risk assessment. *J Pathol*, **213**, 412-9.

Hubers AJ, Prinsen CFM, Sozzi G (2013). Molecular sputum analysis for the diagnosis of lung cancer. *Brit J Cancer*, **109**, 530-7.

Hu ZJ, Liu DY, Hu HB et al (2009). Clinical significance at diagnosis of lung cancer by detecting aberrant p16 promoter methylation in suspicious lung cancer patients. *Chinese J Practical Internal Medicine*, **29**, 50-2.

Hwang SH, Kim KU, Kim JE, et al (2011). Detection of HOXA9 gene methylation in tumor tissues and induced sputum samples from primary lung cancer patients. *Clin Chem Lab Med*, **49**, 699-704.

Jacques F, Hai RS, Freddie B (2010). Estimates of worldwide burden of cancer in 2008. *Int J Cancer*, **127**, 2893-917.

Jadad AR, Moore A, Carroll D, et al (1996). Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials*, **17**, 1-12.

Jarmalaite S, Kannio A, Anttila S, Lazutka JR, Husgafvel-Pursiainen K (2003). Aberrant p16 promoter methylation in smokers and former smokers with non-small cell lung cancer. *Int J Cancer*, **106**, 913-8.

Jemal A, Siegel R, Ward E, et al (2007). Cancer statistics. *CA Cancer J Clin*, **57**, 43-66.

Jiang W, Kang Y, Shi GY et al (2012). Comparisons of multiple characteristics between young and old lung cancer patients. *Chin Med J*, **125**, 72-80.

Kang CY, Wang DD, Tang SP, et al (2011). Effect of promoter hypermethylation of FHIT, p16, MGMT, RASSF1A and APC genes in sputum specimens on diagnosis of lung cancer. *J Clin Ecp Pathol*, **27**, 869-873.

Konno S, Morishita Y, Fukasawa M, et al (2004). Anthracotic index and DNA methylation status of sputum contents can be used for identifying the population at risk of lung carcinoma. *Cancer*, **102**, 348-54.

Li R, Todd NW, Qiu Q, et al (2007). Genetic deletions in sputum as diagnostic markers for early detection of stage I non-small cell lung cancer. *Clin Cancer Res*, **13**, 482-7.

Lijmer JG, Mol BW, Heisterkamp S, et al (1999). Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA*. **282**, 1061-6.

Liu Y, Lan Q, Shen M, et al (2008). Aberrant gene promoter methylation in sputum from individuals exposed to smoky coal emissions. *Anticancer Res*, **28**, 2061-6.

Melvyn S, Tockman (2000). Advances in sputum analysis for screening and early detection of lung cancer. *Cancer Control*, **7**, 19-24.

Miozzo M, Sozzi G, Musso K, et al (1996). Microsatellite alterations in bronchial and sputum specimens of lung cancer patients. *Cancer Res*, **56**, 2285-8.

Molina JR, Yang P, Cassivi SD, Schild SE, Adjei AA (2008). Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. *Mayo Clin Proc*, **83**, 584-94.

Olaussen KA, Soria JC, Park YW, et al (2005). Assessing abnormal gene promoter methylation in paraffin-embedded sputum from patients with NSCLC. *Eur J Cancer*, **41**, 2112-9.

Pai M, Flores LL, Pai N, et al (2003). Diagnostic accuracy of nucleic acid amplification tests for tuberculous meningitis: a systematic review and meta-analysis. *Lancet Infect Dis*, **3**, 633-43.

Palmisano WA, Divine KK, Saccomanno G, et al (2000). Predicting lung cancer by detecting aberrant promoter methylation in sputum. *Cancer Res*, **60**, 5954-8.

Parkin DM, Bray FI, Devesa SS (2001). Cancer burden in the year 2000. The global picture. *Eur J Cancer*, **37**, 4-66.

Peng ZM, Shan CT, Wang HF (2010). Value of promoter methylation of RASSF1A, p16, and DAPK genes in induced sputum in diagnosing lung cancers. *J Cent South Univ (Med Sci)*, **35**, 247-53.

Penny FW, Anne WS, Westwood, et al (2011). QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*, **155**, 529-37.

Ramshankar V, Krishnamurthy A (2013). Lung cancer detection by screening-presenting circulating miRNAs as a promising next generation biomarker breakthrough. *Asian Pac J Cancer Prev*, **14**, 2167-72.

Shin KC, Lee KH, Lee CH, et al (2012). MAGE A1-A6 RT-PCR and MAGE A3 and p16 methylation analysis in induced sputum from patients with lung cancer and non-malignant lung diseases. *Oncol Rep*, **27**, 911-6.

Shivapurkar N, Stastny V, Suzuki M, et al (2007) Application of a methylation gene panel by quantitative PCR for lung cancers. *Cancer Lett*, **247**, 56-71.

Sun N, Zhang L, Liu YY, et al (2013). Methylation of p16

and RASSF1A genes in sputum samples associated with peripheral non-small cell lung cancer. *Prog Modern Biomedicine*, **12**, 2503-10.

Suzuki MM, Bird A (2008). DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet*, **9**, 465-76.

Tsou JA, Galler JS, Siegmund KD, et al (2007). Identification of a panel of sensitive and specific DNA methylation markers for lung adenocarcinoma. *Mol Cancer*, **6**, 70-9.

Van der Drift MA, Prinsen CF, Hol BE, et al (2004). Can free DNA be detected in sputum of lung cancer patients? *Lung Cancer*, **61**, 385-90.

Wang X, Cao A, Peng M, et al (2004). The value of chest CT scan and tumor markers detection in sputum for early diagnosis of peripheral lung cancer. *Chinese J Lung Cancer*, **7**, 58-63.

Wang YC, Hsu HS, Chen TP, Chen JT (2006). Molecular diagnostic markers for lung cancer in sputum and plasma. *Ann NY Acad Sci*, **1075**, 179-84.

Whiting P, Rutjes AW, Dinnes J, et al (2004). Development and validation of methods for assessing the quality of diagnostic accuracy studies. *Health Technol Asses*, **8**, 1-234.

Whiting P, Rutjes AW, Reitsma JB, et al (2004). Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med*, **140**, 189-202.

Xie GS, Hou AR, Li LY, Gao YN, Cheng SJ (2006). Aberrant p16 promoter hypermethylation in bronchial mucosae as a biomarker for the early detection of lung cancer. *Chin Med J*, **119**, 1469-72.

Zhang W, Sun YE, Lu GM (2004). The diagnostic value of determination of p16 methylation of sputum exfoliated cells for peripheral lung cancer. *Chin J Lung Cancer*, **7**, 46-9.

Zhang W, Yu CH, Xia H, et al (2012). Diagnostic value of determination of p16 methylation and k-ras mutation of sputum exfoliated cells for peripheral lung cancer. *China J Modern Medicine*, **22**, 9-13.

Ziaian B, Saberi A, Ghayyoumi MA, et al (2014). Association of high ldh and low glucose levels in pleural space with her2 expression in non-small cell lung cancer. *Asian Pac J Cancer Prev*, **15**, 1617-20.