**Intruduction**

Lung cancer is a complicated disease involving genetic and epigenetic variation, and is the leading cause of cancer death all over the world [1]. Lung cancer is often silent in its early stages and difficult to diagnose which is the stage when it could be treated effectively. Approximately eighty percent of primary lung cancer is non-small cell lung cancer (NSCLC), which often progresses slower than small cell lung cancer (SCLC), so we have just long enough time to detect and treat it. The overall five-year survival rates for late stage III and IV of NSCLC patients were just 5%-14% and 1% respectively, the rate could come up to 63% for the early stage Ia of the NSCLC patients who are treated with surgery properly[2, 3]

DNA methylation is one of the epigenetic modifications in eukaryote, which regulates genes and microRNAs expression[4], gene alternative splicing[5], playing important role in the developing of cancer, thus can be used as a biomarker to detect lung cancer in the early stage. In addition, DNA methylation has the advantage of stable chemical property, detection ability in remote patient media, quantitative signal, relatively low cost in detection and so on[6]. So it can be a useful clinical tool to achieve early diagnosis of lung cancer.

The *CDH13* gene, an atypical member of the cadherin superfamily, was isolated recently and has been mapped to 16q24[7], which was devoid of a transmembrane domain and anchored to the exterior surface of the plasma membrane via a glycosylphosphatidylinositol anchor[8]. Evidence showed that promoter methylation, which inhibits CDH13 gene expression, is mediated by DNA methyltransferases Dnmt3A[9]. Researchers have reported that hypermethylation and loss of function of CDH13 was detected in in breast[10] and lung cancers[11-13], in pituitary adenomas[14], diffuse large B cell lymphoma[15], and nasopharyngeal carcinoma[16]. Furthermore, CDH13 gene has been suggested as an early detecting marker for lung cancers [17].

In this article, we conducted a meta-analysis of the sensitivity and specificity of CDH13 methylation on NSCLC diagnosis. The homogeneity of these studies is very well since I−squared is low. We also found that The Cancer Genome Atlas project (TCGA) had collected hundreds of whole genome DNA methylation microarray datasets of NSCLC samples which included comprehensive clinical and demographic information, providing an additional resource that may be without publication bias. In our work, we innovatively integrated these TCGA data (Additional file 1: Table S1) and the data from published articles to evaluate the diagnostic ability of the CDH13 methylation test in NSCLC. Therefore, an integrated analysis of all these existing data was conducted to come to unbiased conclusions on the relationship between CDH13 methylation and NSCLC

**Results**

**Study characteristics**

The electronic search strategy identified 365 potentially relevant articles (PubMed, 73; Web of science, 177; OVID Embase, 115; Cochrane Library, 0), which were further screened for inclusion on the basis of their titles, abstracts, full texts, or a combination of these terms. The electronic search was supplemented from reference lists of relevant articles including reviews. Finally, 11 studies with data on the relationship between CDH13 gene promoter methylation and NSCLC were pooled for analysis (Table 1) [10,12-27]. All these articles were written in English. In total, 1261 lung cancer tissues/serum and 652 normal counterpart tissues/serum were collected. The age of the subjects in the 11 studies ranged from 26 to 87 years, with mean or median ranging from 62 to 70 years. As for the study aim, 4 articles were especially aiming at diagnosis, while the others were for prognosis, survival research, and so on. Among 11 studies, the proportions of stage I samples differed from 9.52 to 57.95%, and the percentage of male individuals in the NSCLC samples has a range of 53 to 80%. For the experimental methods to explore CDH13 promoter methylation status, 7 of 11 inclusions used methylation-specific polymerase chain reaction (MSP), while others used quantitative MSP (qMSP, such as Methylight, Prosequencing, and so on).

**Meta-analysis**

The ORs for CDH13 methylation in cancer samples(tissue or plasma) compared with that in normal controls were 6.49 (95% CI: 4.18 to 10.09, z =8.32, P<0.00001) in random effects model pooled using Inverse Variance method, and 5.85 (95% CI: 4.25 to 8.04, z =10.86, P<0.00001) in fixed effects model using Inverse Variance method(为什么这两个都不用Mantel-Haenszel，不过好像并无所谓？), demonstrating a statistically significant increasing in likelihood of methylation in lung cancer tissues comparing to controls (Figure 1). The homogeneity of these studies is very well since I−squared is low(I2=40%, Chi2=20.11).

Subgroup analyses were conducted for different subtypes, which included sample types (tissue or serum), age, counterpart categories (autogenous or heterogeneous与组织/血清亚组相同), proportion of early stage, aim of the study (for diagnosis or non-diagnosis), ratio of adenocarcinoma to squamous (Ad2Sc) and other possible confounding factors (Table 2). Significant differences were found between the ORs of the diagnosis (3.38, 95% CI: 2.10 to 5.43) and non-diagnosis (10.62, 95% CI: 6.97 to 16.18) subgroup (P = 0.0004) (Figure 2A). Both tissue and serum groups showed significant association between FHIT methylation and NSCLC (OR = 6.73 and 8.17, respectively) (Figure 2B) which suggested that FHIT methylation can be taken as a potential biomarker for NSCLC diagnosis using either tissue or serum samples. No significant difference was found between subgroups of MSP and qMSP (OR = 6.70 and 7.01, respectively; P = 0.62), which suggested both of the methods were equivalent in methylation detection. In addition, ratio of male to female, counterpart categories, ratio of adenocarcinoma to squamous, the primer set and other factors are not the source of heterogeneity (Table 2).

因为是否以诊断为目的而区分的亚组间有明显差异，我们在非诊断组中继续深入研究。However, when we focus on the studies not aiming at diagnosis, significant difference was found between subgroups of the proportion of adenocarcinoma <50% (7.22, 95% CI: 4.29 to 12.16) and >50% (23.86, 95% CI: 9.97 to 57.12) (p=0.02).

**Summary receiver operating characteristic curve for diagnostic capacity of CDH13 methylation**

**（此段是否需要）**

Pooled sensitivity and specificity were ？？and ？？ for all the studies based on the presupposition of the fixed effects model. The sensitivity of the tissue group was higher than that of the serum group, ？ (？ to ？) versus ？ (？ to ？), while the specificity of the serum group was higher than that of the tissue group, ？ (0.86 to 0.96) versus ？ (0.49 to 0.83), which suggested the advantage of this biomarker for its higher diagnostic ability using remote non-invasive media.

Although sensitivity and specificity were two of most important features of a diagnostic test, in some occasions, pooling sensitivity or specificity could be misleading as mentioned in the Methods section. Therefore, we constructed the summery receiver operating characteristic (SROC) curve to depict the stability and accuracy of the methylation test’s diagnostic ability. The area under the curve (AUC) of the SROC was ???, suggesting a fair ability for NSCLC diagnosis (Figure 2F). Meanwhile, the AUC of the SROC for the serum and the tissue group was ? and ? respectively, showing slightly different performances for the CDH13 methylation test in serum and tissue samples.

**Bias analysis and robust estimation of pooled OR**

A funnel plot of methylation status of lung cancer tissue versus normal tissue showed significant publication bias (Egger test, z =??, P <???) and 1 study exceeded the 95% confidence limits (Additional file 2: Figure S1). In order to eliminate the effect of publication bias, trim and fill analysis was performed with the random effects model. The adjusted pooled OR were 5.04 (95% CI: 3.20 to 7.96, P =???) in the random effects model and 4.93 (95% CI: 3.69 to 6.60, P < ????) in the fixed effects model. Both results demonstrate a significantly positive association between CDH13 methylation and NSCLC (Additional file 2: Figure S2).

In sensitivity analysis to determine the effect of omitting a single study on the overall effect, the overall ORs were between 5.74 (95% CI: 4.19 to 7.85) and 6.88 (95% CI: 4.93 to 9.61) in the fixed effects method, which suggested that combined OR was consistent and reliable (Additional file 2: Figure S3).

**Validation by independent TCGA lung cancer dataset（濮）**

**在数据库验证中，我们发现了与非诊断组相同的结论，即CDH13的甲基化在腺癌与非癌样本中存在差异，而在鳞癌中则无明显区别。**

**Discussion**

The CDH13 gene has been reported as an important tumor suppressor in colorectal cancer [30], and the aberrant of CDH13 methylation had been reported in numerics for cancers, such as bladder [31], prostate [32], breast and lung cancer [24]. However, the diagnostic role of the methylation status of the CDH13 gene in lung cancer lacks quantitative assessment. We therefore performed an integrated analysis to quantify the ability for the CDH13 promoter methylation test in NSCLC diagnosis, and a significant association was identified between CDH13 methylation and NSCLC (OR = 5.86(使用固定效应模型), P < ？？？). Five imputed studies were filled when trim and fill tests were performed to eliminate the influence of publication bias on the fixed effects model, and the overall OR (4.93, 95% CI: 3.69 to 6.60) was still significant, although it was slightly smaller than that in the crude meta-analysis (5.86, 95% CI: 4.26 to 8.06), indicating the existence of a strong association between CDH13 promoter methylation and lung cancer.

**CDH13的甲基化在腺癌与非癌样本中存在差异，而在鳞癌中则无明显区别。因此CDH13未来有潜力被用作肺腺癌的联合诊断指标之一。**

**Conclusion**  
In conclusion, this integrated analysis of the pooled data provides strong evidence that the methylation status of the CDH13 promoter is strongly associated with NSCLC, especially for adenocarcinoma. Therefore, the CDH13 methylation test could be a promising diagnostic biomarker which could be applied in the clinical diagnosis of lung adenocarcinoma with remote non-invasive media detection.

**Methods**

**Search strategy, selection of studies and data extraction**

This pooled study involved searching a range of computerized databases, including PubMed, Cochrane Library, OVID Medline and Web of Science for articles published in English or Chinese by December 2014. The study used a subject and text word strategy with (CDH13 OR CDHH OR P105 OR H-cadherin OR Cdht OR T-cadherin OR Tcad OR CH211-122A20.1 OR BOS\_16969 OR cdhh) AND (lung or non-small) as the primary search terms. Wildcard character of star, dollar or some other truncations were applied according to the rules of the databases to allow effective article collection.

Two independent reviewers (Geng, Guo) screened the titles and abstracts derived from the literature search to identify relevant studies. The following types of studies were excluded: animal and cell experiments, case reports, reviews or meta-analyses and studies of non-case-control studies or studies with insufficient data or those proving inaccessible after making contact with the authors. The remaining articles were further examined to see if they met the inclusion criteria: 1) the patients had to be diagnosed with NSCLC (Ad and Sc), 2) the studies had to contain CDH13 gene promoter methylation data from tissue, blood or serum, 3) the studies had to be case-control studies which included tissue-tissue, blood-blood or serum-serum in case and controls respectively. The reference sections of all retrieved articles were searched to identify further relevant articles. Potentially relevant papers were obtained and the full text articles were screened for inclusion by two independent reviewers (Geng, Guo). Disagreements were resolved by discussion with KX, JJW, and JHW. Included studies were summarized in data extraction forms. Authors were contacted when relevant data were missing. The name of the first author, year of publication, sample size, age (mean or median), gender proportion (male/female, M2F), the proportion of TNM stage I samples (proportion of early stage of NSCLC samples), publication aim (for diagnosis or not), analyzing multiple genes or not (one or more genes detected simultaneously in studies design), control type (autogenous or heterogeneous counterpart) and methylation status of the CDH13 promoter in human NSCLC and normal or control tissues were extracted.

**Meta-analysis and SROC analysis**

Data were analyzed and visualized mainly using R Software (R version 2.15.3) including meta, metefor and mada packages. The strength of association was expressed as pooled odds ratio (OR) with corresponding 95% confidence intervals (95% CI). Data were extracted from the original studies and recalculated if necessary. Heterogeneity was tested using the I2 statistic with values over 50% and Chi-squared test with P ≤ 0.1 indicating strong heterogeneity between the studies [36]. Tau-squared (τ2) was used to determine how much heterogeneity was explained by subgroup differences. The data were pooled using the DerSimonian and Laird random effects model (I2 > 50%, P ≤ 0.1) or fixed effects model (I2 < 50%) according to heterogeneity statistic I2 [37]. A two-sided P ≤ 0.05 was considered significant without special annotation. Random effects meta-regression, was employed to determine how much of the heterogeneity (between-study variance) is explained by the explanatory variables when the heterogeneity was significant [38]. Nine variables were analyzed in meta-regression, including control types (autogenous and heterogeneous), gender proportion, proportion of TNM stage I samples, mean or median age (> 65 or ≤ 65), single or multiple target detection, sample types (serum or tissue), methylation detection methods (MSP, qMSP), study designs (diagnosis or non-diagnosis) and primer sets. Sensitivity analyses were performed to assess the contributions of single studies to the final results with the abandonment of one article each time. Publication bias was analyzed by funnel plot with mixed-effects version of the Egger test. If bias was suspected, the conventional meta-trim method was used to re-estimate the effect size.

Compared with traditional SNP association studies, methylation-associated research might be involved with different methylation-definition thresholds. In these cases, traditional weighted averages (pooled sensitivity and specificity) would not reflect the overall accuracy of the test, because the extremes of threshold criteria could skew the distribution, known as the threshold effect [39]. Thus, SROC analysis was applied to meta-analysis of diagnostic tests[39, 40]. The SROC curve shows the performance of the diagnostic ability of CDH13 methylation to NSCLC. Each study produces values for sensitivity, specificity and therefore true positive rate (TPR) and false positive rate (FPR), and the plots were placed over the TPR and FPR points to form a smooth curve. A linear regression model was selected to fit the SROC curve where sensitivity and (1-specificity) are transformed into complex logarithmic variables. The exact AUC for the SROC function was used to assess the accuracy of the test [39].

**TCGA data extraction and analysis**

**References:**

[1]. Siegel, R., D. Naishadham and A. Jemal, Cancer statistics, 2012. CA Cancer J Clin, 2012. 62(1): p. 10-29.

[2]. Hankey, B.F., L.A. Ries and B.K. Edwards, The surveillance, epidemiology, and end results program: a national resource. Cancer Epidemiol Biomarkers Prev, 1999. 8(12): p. 1117-21.

[3]. van Rens, M.T., et al., Prognostic assessment of 2,361 patients who underwent pulmonary resection for non-small cell lung cancer, stage I, II, and IIIA. Chest, 2000. 117(2): p. 374-9.

[4]. He, Y., et al., Hypomethylation of the hsa-miR-191 locus causes high expression of hsa-mir-191 and promotes the epithelial-to-mesenchymal transition in hepatocellular carcinoma. Neoplasia, 2011. 13(9): p. 841-53.

[5]. Flores, K., et al., Genome-wide association between DNA methylation and alternative splicing in an invertebrate. BMC Genomics, 2012. 13: p. 480.

[6]. Gokul, G. and S. Khosla, DNA methylation and cancer. Subcell Biochem, 2013. 61: p. 597-625.

[7]. Flores, K., et al., Genome-wide association between DNA methylation and alternative splicing in an invertebrate. BMC Genomics, 2012. 13: p. 480.

[8]. Hulpiau, P. and F. van Roy, Molecular evolution of the cadherin superfamily. Int J Biochem Cell Biol, 2009. 41(2): p. 349-69.

[9]. Shamay, M., et al., Recruitment of the de novo DNA methyltransferase Dnmt3a by Kaposi's sarcoma-associated herpesvirus LANA. Proc Natl Acad Sci U S A, 2006. 103(39): p. 14554-9.

[10]. Riener, M.O., et al., Microarray comparative genomic hybridization analysis of tubular breast carcinoma shows recurrent loss of the CDH13 locus on 16q. Hum Pathol, 2008. 39(11): p. 1621-9.

[11]. Sato, M., et al., The H-cadherin (CDH13) gene is inactivated in human lung cancer. Hum Genet, 1998. 103(1): p. 96-101.

[12]. Sato, M., et al., Identification of a 910-kb region of common allelic loss in chromosome bands 16q24.1-q24.2 in human lung cancer. Genes Chromosomes Cancer, 1998. 22(1): p. 1-8.

[13]. Toyooka, K.O., et al., Loss of expression and aberrant methylation of the CDH13 (H-cadherin) gene in breast and lung carcinomas. Cancer Res, 2001. 61(11): p. 4556-60.

[14]. Qian, Z.R., et al., Tumor-specific downregulation and methylation of the CDH13 (H-cadherin) and CDH1 (E-cadherin) genes correlate with aggressiveness of human pituitary adenomas. Mod Pathol, 2007. 20(12): p. 1269-77.

[15]. Ogama, Y., et al., Prevalent hyper-methylation of the CDH13 gene promoter in malignant B cell lymphomas. Int J Oncol, 2004. 25(3): p. 685-91.

[16]. Sun, D., et al., Aberrant methylation of CDH13 gene in nasopharyngeal carcinoma could serve as a potential diagnostic biomarker. Oral Oncol, 2007. 43(1): p. 82-7.

[17]. Kim, D.S., et al., Aberrant methylation of E-cadherin and H-cadherin genes in nonsmall cell lung cancer and its relation to clinicopathologic features. Cancer, 2007. 110(12): p. 2785-92.