# Title page

**Title:** Quantitative assessment of the Diagnostic role of APC promoter methylation in non-small cell lung cancer

Shicheng Guo1, Lixing Tan1, Weilin Pu1, Junjie Wu1,2, Kuan Xu3, Jinhui Wu4, Qiang Li2, Yanyun Ma1, Jibin Xu5, Li Jin1,6\*, Jiu-Cun Wang1,6\*

Author’s Affiliations:

1State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200433 China. 2Department of Pneumology, Changhai Hospital of Shanghai, Second Military Medical University, Shanghai 200433, China;

3Department of Head and Neck Surgery, Cancer Hospital, Fudan University, Shanghai, 200032, China;

4Department of General Surgery University of Qingdao Affiliated Hospital of Medical College, Qingdao Univesity, No.1677 Wutaishan Street, Qingdao City 266071, China

5Department of Cardiothoracic Surgery, Changhai Hospital of Shanghai, Second Military Medical University, Shanghai, China

6Fudan-Taizhou Institute of Health Sciences, 1 Yaocheng Road, Taizhou, Jiangsu 225300, China

**Running title:** Diagnostic Role of *APC* Promoter Methylation in NSCLC

**Key words:** *APC*, DNA methylation, Diagnosis, Meta-analysis, TCGA, NSCLC, Biomarker, Adenocarcinoma

**Grant support:** This research was supported by National Science Foundation of China (NSFC, 81172228), National S&T Major Special Project (2011ZX09102-010-01), National High-Tech Research and Development Program (2012AA021802), and Shanghai Postdoctoral Sustentation Fund (12R21411500).

**Corresponding authors:**  Li Jin, National Ministry of Education Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200433, China, Phone: +86-21-55664885, Fax: +86-21-55664885, E-mail: [lijin.fudan@gmail.com](mailto:lijin.fudan@gmail.com) and Jiu-Cun Wang, National Ministry of Education Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200433, China, Phone: +86-21-55665499, Fax: +86-21-556648845, E-mail: [jcwang@fudan.edu.cn](mailto:jcwang@fudan.edu.cn).

**Abstract**

Adenomatous polyposis coli (*APC*) has been reported to be a candidate tumor suppressor in many cancers. However, the diagnostic role of *APC* promoter methylation in NSCLC remains unclear. In present study, 2259 NSCLC and 1039 controls were collected from 17 published studies and TCGA NSCLC data to discover the diagnosis performance of *APC* methylation test for NSCLC. A significant association was observed between *APC* promoter hypermethylation and NSCLC, with an aggregated odds ratio (OR) of 3.79 (95% CI: 2.22-6.45) in random effect model. Pooled sensitivity and specificity were 0.548 (95% CI: 0.42-0.67, P<0.0001) and 0.776 (95% CI: 0.62-0.88, P<0.0001), respectively. Comprehensive microarray analysis showed

in lung adenocarcinoma (Ad) was much better in prediction (AUC: 0.71-0.73) than that in lung squamous cell carcinoma (Sc) (AUC: 0.45-0.61). The AUCs of the logistic prediction model based on these 5 CpGs were 0.73 and 0.60 for Ad and Sc, respectively. Integrated analysis showed CpG site location, heterogeneous or autogenous controls and the proportion of adenocarcinoma in samples were most significant heterogeneity sources. The methylation status of *APC* promoter was strongly associated with NSCLC, especially adenocarcinoma.

**Introduction**

Non-small cell Lung cancer(NSCLC), including adenocarcinoma(Ad) and squamous cell carcinoma (Sc), is the leading cause of cancer death in men and women in the United States([1](#_ENREF_1)). Over 159,480 Americans die of this disease every year in U.S.A ([1](#_ENREF_1)). The 5-year relative survival rate varies markedly depending on the stage at diagnosis, from 49% to 16% to 2% for patients with local, regional, and distant stage disease, respectively (SEER Cancer Statistics Review 1975-2002). Early detection is a key bottleneck in increasing lung cancer patient survival([2](#_ENREF_2)). DNA hypermethylation has been recognized as an important mechanism for tumor suppressor gene inactivation in cancer and could yield powerful biomarkers for early detection of lung cancer and own incomparable advantage than other traditional markers for its stable chemical property, detect ability in remote patient media, quantitative signal, convenient low cost in detection, etc. ([3](#_ENREF_3)). Several revolutionary steps has been made to push methylation biomarkers into cancer screening ([4](#_ENREF_4), [5](#_ENREF_5)) which indicated DNA methylation would become a powerful tools for lung cancer diagnosis.

*APC* gene encodes a tumor suppressor protein that acts as an antagonist of the Wnt signaling pathway. It is also involved in other processes including cell migration and adhesion, transcriptional activation, and apoptosis ([6](#_ENREF_6)). Defects in this gene cause familial adenomatous polyposis (FAP), an autosomal dominant pre-malignant disease that usually progresses to malignancy which suggests it is a potential predictor for cancer initial or development. Promoter methylation inhibits *APC* gene expression mediated by changes of chromatin conformation and aberrant binding of CCAAT-box binding transcription factors ([7](#_ENREF_7)).

Following P16INK4A([8](#_ENREF_8)), the relationship between hypermethylation of *APC* with cancers was extensively estimated([9](#_ENREF_9)) and *APC* promoter hypermethylation in NSCLC was believed as an effective biomarker for diagnosis([10](#_ENREF_10), [11](#_ENREF_11)). However, the results were dramatically different among all researches which may be caused by the difference of gender proportion, age distribution, race source, some other epidemiological characteristics in samples, and detection methods, etc. In addition, there was not any quantitative assessment of the relationship between the hypermethylation in the promoter region of *APC* gene and NSCLC yet.

In this article, we conducted a meta-analysis of the sensitivity and specificity of *APC* methylation on NSCLC diagnosis. The factors which led heterogeneity to the sensitivity and specificity were discovered with meta-regression. We also found that the cancer genome atlas project (TCGA) had provided hundreds of whole genome DNA methylation microarray data of NSCLC samples with comprehensive clinical and demographic information, which could be integrated with the data in published articles to evaluate the diagnosis ability of *APC* methylation test in NSCLC. Therefore, an integrated analysis of all these existing data was conducted to make unbiased conclusions on the relationship between *APC* methylation and NSCLC.

**Materials and 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 TMC ProSearch for articles published in English or Chinese to March 2013. The study used a subject and text word strategy with (*APC* OR *BTPS2* OR *DP2* OR *DP2.5* OR *DP3* OR *PPP1R461*) AND ((Lung OR NSCLC) AND (cancer OR neoplasm)) as the primary search terms. Wildcard character of star, dollar or some other truncations were applied according to the rule of the databases to make effective collection of the articles.

Two independent reviewers (Guo, Tan) screened the titles and abstracts derived from the literature search to identify relevant studies. The following types of studies were excluded: animal experiments, case reports, reviews or meta-analyses and studies of non-case-control study or with insufficient data or be inaccessible after the contact with the authors. The remained 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 have *APC* gene promoter methylation data from tissue, blood or serum, 3) case-control study, 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 (Guo, Tan). Disagreements were resolved by discussion with KX, JJW, 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 heterogenous counterpart) and methylation status of the *APC* promoter in human NSCLC and normal or control tissues were extracted.

**Meta-analysis and summary receiver operating characteristics 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 ([12](#_ENREF_12)). 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([13](#_ENREF_13)). A two-sided P ≤ 0.05 was considered significant without special annotation. Random-effects meta-regression, were employed to determine how much of the heterogeneity (between-study variance) is explained by the explanatory variables when the heterogeneity was significant ([14](#_ENREF_14)). 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 type (serum or tissue), methylation detection method (MSP, qMSP), study design (diagnosis or non-diagnosis) and primer sets. Sensitivity analyses were performed to assess the contributions of single studies to the final results with the abandon 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 study, 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, as the extremes of threshold criteria could skew the distribution, that is so called threshold effect ([15](#_ENREF_15)). Thus, summary receiver operating characteristics (SROC) analysis can be applied to meta-analysis of diagnostic tests ([15](#_ENREF_15), [16](#_ENREF_16)). It was plotted to show the performance of the diagnosis ability of *APC* methylation to NSCLC. Each study produces values for sensitivity, specificity and therefore true positive rate (TPR) and false positive rate (FPR). The SROC curve is placed over the (TPR, FPR) points to form a smooth curve. Linear regression model were selected to fit the SROC curve where sensitivity and (1-specificity) are transformed into complex logarithmic variables. The exact area under the curve (AUC) for the SROC function was used to assess the accuracy of the test ([15](#_ENREF_15)).

**TCGA data extraction and analysis**

DNA methylation information for NSCLC which included two sets of samples (535 Ad and 50 Control, and 385 Sc and 67 control) were collected from TCGA Project including methylation 27K and 450K dataset ([http://cancergenome.nih.gov/). The](http://cancergenome.nih.gov/).%20The)estimate of methylation for each CG probe was calculated with the traditional function: beta. M and U represent the mean signal intensities for about 30 replicate methylated (M) and unmethylated (U) probes on the array. The methylation signals of the 25978 shared CpG sites by 27K and 450K datasets were extracted and the methylation status of each probe was defined according to the beta value. The CpG site will be considered methylated when the beta value is greater than the empirical threshold of 0.3 for tissue data ([17](#_ENREF_17)). Six CpG sites located in the promoter region of *APC* gene (cg01240931, cg15020645, cg16970232, cg20311501, cg21634602 and cg24332422) were taken as the object of study (**Supplementary Table S1**). Adjustment for multiple testing of differential methylation was conducted with the method of Benjamini and Hochberg at the 5% FDR level.

**Results**

**Study characteristics**

The electronic search strategy identified 506 potentially relevant articles (Pubmed, 315; Scopus, 112; Cochrane Library,3; OVID Medline, 53; TMC ProSearch, 23), 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, 17 studies included data on the relationship between *APC* gene promoter methylation and NSCLC were pooled for analysis (**Table 1**) ([10](#_ENREF_10), [18-33](#_ENREF_18)). All the included articles were written in English. Totally, 1338 lung cancer tissues /serum and 913 normal counterpart tissues/serum were collected. The age of the subjects in the 17 studies ranged from25 to 86 years while mean or median age was 53-67. Among the 17retrieved studies (13 articles were especially for diagnosis, while the others were for prognosis, survival research, etc.), 7 observations used methylation-specific polymerase chain reaction (MSP) while others used quantitative MSP (qMSP, such as Methylight, Prosequencing, etc.) to explore *APC* promoter methylation status. The proportions of the samples in stage I were counted and the ranges were 32.1-100% and 70-100% respectively. The percentage of male individuals in the NSCLC samples ranged from 53% to 81%. Two kinds of methylation detection primers or probes were found to be utilized for most of the 17 studies. The information of the two sets of primers (set I: chr5:112073421-112073518, 7 studies; and set II: chr5:112101379-112101452, 7 studies) was listed in **Supplementary Table S2**. In addition, no CpG sites from the methylation microarrays was found located in the above primers, while cg20311501 is covered by the replication region of Set II primers.

**Meta-analysis, subgroup analysis and meta-regression**

The ORs for *APC* methylation in cancer tissues compared with that in normal controls were 4.67 (95% CI: 2.66-8.22, z=5.35, P < 0.0001), and 2.74 (95% CI: 1.99-3.23, z =8.10, P < 0.0001) in random effect model pooled and fixed effect model, respectively, indicating an increased likelihood of methylation in Lung cancer tissues **(Fig. 1)**.

Subgroup analysis were conducted for different subtypes, which included sample type( tissue or serum),counterparts category (autogenous or heterogeneous), proportion of stage I, aim of the study (for diagnosis or non-diagnosis), ratio of adenocarcinoma to squamous(Ad2Sc), primer categories (set I and II) and other possible confounding factors **(Table 2)**. Significant difference were found between the OR of younger (5.03, 95% CI: 2.53-10.0) and older group (0.91, 95% CI: 0.57-1.41) subgroup (P <0.0001) (**Fig. 2A**). High proportion of adenocarcinoma group had a significantly bigger OR than that of low subgroup (P=0.0077), which suggested that *APC* methylation might have subtype specificity in NSCLC (**Fig. 2C**). Significant difference was found between primer set I and II (P=0.0137), which supported primers were one of the most important heterogeneity sources in the *APC* methylation test (**Fig. 2D**). When we see the effects of sample type and control type on the OR of *APC* methylation, it can be found that both tissue and serum groups had showed significant association between *APC* methylation and NSCLC(OR=3.72, 11.54, respectively) which suggested *APC* methylation can be taken as a potential biomarker for NSCLC diagnosis using either tissue or serum samples. In addition, significant difference were found between the ORs of heterogeneous (ORh = 8.33, 95% CI: 3.77-18.39) and autogenous (ORa = 2.25, 95% CI: 1.06-4.77) subgroups (P = 0.0187) (**Fig. 2B**). One possible reason might be the impure composition of the adjacent normal specimens which might have been slightly contaminated by cancer cells or it have been transformed to precancerous status, while normal serum samples came from healthy individuals in general. The subgroup of high Ad2Sc group had a larger OR than that of low Ad2Sc **(Table 2)**, indicating that methylation of *APC* might occurred or functioned at the early stage of the tumorigenesis, which had been founded in endometrial cancer ([34](#_ENREF_34)). Differences in the OR of diagnosis (OR=6.79) or non-diagnosis group (OR = 2.59) was very large, which might be caused by unbalanced distribution of proportion of early stage samples (P = 0.0218, Wilcoxon sum-rank test) . No significant difference was found between the subgroups of MSP and qMSP (P = 0.77), which suggested both of the methods were equivalent in methylation detection **(Table 2)** and the result was consistent with Wu’s conclusion ([35](#_ENREF_35)).

Further analyses were performed using meta-regression method with Knapp-Hartung modification to determine the sources of heterogeneity, which significantly existed among all studies (I2= 79.2%, Q = 52.78, P < 0.0001) (**Fig. 1**). It showed that the trend in ORs was inversely correlated with age, suggesting that the age accounted for some of the heterogeneity (beta = -0.3, P = 2.0×10-5), which could count for 83.8% total variances. It was consistent with the subgroup analysis in which the OR of older group (OR=2.24) was smaller than younger group (OR=4.65). The variable of primer set was also an important heterogeneity source (P = 0.05) which could explain about 68% of overall heterogeneity. Other factors such as sample type, proportion of males, proportion of stage I and detection methods could not explain the heterogeneity significantly (**Table 3**).

**Summary receiver operating characteristic curve for diagnosis capacity of *APC* methylation**

Pooled sensitivity and specificity were 0.548 (95% CI: 0.42-0.67, P<0.0001) and 0.78 (95% CI: 0.62-0.88, P<0.0001) for the whole studies based on the presupposition of fixed effect model. The sensitivity of tissue subgroup was higher than that of serum subgroup, 0.61 (0.45-0.75) versus 0.396 (0.26-0.56), while the specificity of serum subgroup was higher than that of tissue subgroup, 0.92(0.86-0.96) versus 0.68 (0.49-0.83), which suggested the advantage of this biomarker for its high ability in diagnosis using remote non-invasive media.

However, although sensitivity and specificity were two of most important features of a diagnosis test, in some occasions, pooling sensitivity or specificity could be a misleading event as mentioned in the method section. Therefore, the ability in diagnosis of the methylation test was assessed with SROC curve to depict its stability and accuracy. The AUC of the SROC was 0.64, suggesting the potential ability for NSCLC diagnosis (**Fig. 2F**). Meanwhile, the AUC of the SROC for serum and tissue group were 0.67 and 0.64 respectively which indicated the different performance for *APC* 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 = 4.3, P<0 .0001) and eight studies exceeded the 95% confidence limits **(S Fig. S1)**. In order to eliminate the effect of publication bias, trim and fill analysis was performed using the random effects model. The adjusted pooled OR were 2.50 (95% CI: 1.43-4.38, P=0.0013) in random effect model and 2.19(95% CI: 1.74-2.77, P<0.0001) in fixed effect model, respectively, indicating a significantly positive association between *APC* methylation and NSCLC **(S Fig. S2)**.

In sensitivity analysis to determine the effect of omitting a single study on the overall effect, it showed the overall ORs were between 4.3(95% CI: 2.46-7.52) and 5.27 (95% CI: 2.92-9.53) in the random effect method, which suggested that combined OR was consistent and reliable **(S Fig. S3)**.

A cumulative meta-analysis by the time of the published literature were also conducted, and we found the OR was tending to be stable (**Fig. 2E**), which was suggested the result of the meta-analysis might be more credible when add more incoming researches.

Using similar methodology, the influence on meta-regression was analyzed by omitting one study each time to explore heterogeneity sources. The sample type of tissue or serum would be one of the heterogeneity sources when Begum et al (2011, USA) were removed from the meta studies(P< 0.026), likewise, the proportion of stage I and aim of the study would become heterogeneity source when Lin et al (2009, China), Zhang et al (2011, China) or Yanagawa et al (2003, Japan) was removed(P-value were 0.0046, 0.029 and 0.039 respectively), which suggested these factors should be considered in the future case-control association study.

**Validation by independent TCGA lung cancer dataset**

In order to make independent validation of the above results, we collected the data of the methylation status of 6 CpG sites located in the promoter region of *APC* gene from the lung cancer samples of TCGA Project (**Supplementary Table S1**). Pairwise methylation Pearson correlation analysis showed that the methylation status was highly correlated among these CpG sites (R2> 0.90 for all) except cg01240931 (R2 < 0.45 for all), which suggested that cg01240931 was beyond of the “methylation block” composed by the other 5 CpG sites. Meanwhile, cg01240931 were hypermethylated in both the cancer and normal specimens. Therefore, this CpG site were excluded in the following analysis.

The Clinical Characteristics of the NSCLC samples were extracted from TCGA Project. There is no significant difference in age and gender between the cases and controls (data not shown).The methylation percentage of cg15020645, cg16970232, cg20311501, cg21634602 and cg24332422 were dramatically different between the two groups, especially in adenocarcinoma. The methylation of all these 5 CpG sites were significantly different between Ad and its counterparts based on t-test after FDR adjustment (P<10-17), whereas, only two CpG sites (cg16970232, cg20311501) were significantly different between Sc and its counterparts (P=1.6×10-6and 3.9×10-3), and the significant level in Sc was dramatically lower than that in Ad (**Table 4**). In addition, logistic regression analysis also supported the above results: the ORs in Ad were from 23.3 to 1.2×103, while those were from 0.15 to 7.54 in Sc **(Table 4)**. The AUCs of the 5 CpG methylation test were calculated to assess their prediction ability. As shown in Table 4, each of the CpG site in Ad was much better in prediction (AUC: 0.71-0.73) than that in Sc (AUC: 0.45-0.61)**.** In addition, the AUCs of the logistic prediction model based on all the 5 CpG sites were 0.73 and 0.60 for Ad and Sc, respectively. It revealed from all the above evidences that *APC* methylation test would have better performance in adenocarcinoma than that in Sc, and therefore, the different proportion of Ad and Sc in the samples might bring certain bias for the association between *APC* methylation and NSCLC. Generally, 25% to 30% of lung cancers were Sc while 40% were Ad. Thus, we resampled the Ad and Sc from TCGA data to simulate the effect of different ratio of Ad versus Sc (Ad2Sc) at 2:1, 4:3, 3:4 and 1:2 on the odds ratio of *APC* methylation for NSCLC. The ORs were dramatically varied within group and between groups of the 5 CpGs by 10000 times of resampling simulations **(Supplementary Table S3**). As expected, cg16970232 and cg20311501, the significant sites in both Ad and Sc, were consistently significant risk factors for NSCLC, while the other three CpGs would loss association with NSCLC in certain vignettes (**Supplementary Table S3**). Moreover, ORs from logistic regression based on heterogeneous samples were significantly greater than those on the autogenous samples in the condition of Ad2Sc of 4:3, which was concordant with the above subgroup meta-analysis (**Supplementary Table S4**). Logistic based interaction analysis among age, gender with NSCLC showed there were no significant interaction in NSCLC risk between *APC* methylation and these covariates (**Supplementary Table S5**).

**Discussion**

*APC* gene has been supposed as an important tumor suppressor gene in colorectal cancer([36](#_ENREF_36)), and the aberrant of *APC* methylation had been reported in numeric of cancers, such as bladder([37](#_ENREF_37)), prostate ([38](#_ENREF_38)),breast and lung cancer([30](#_ENREF_30)). However, the diagnostic role of the methylation status of *APC* gene in Lung cancer lacks for quantitative assessment. We therefore performed an integrated analysis to quantify the ability for *APC* promoter methylation test in NSCLC diagnosis, and a significant association was identified between *APC* methylation and NSCLC (OR = 4.67, P < 0.0001). Seven virtual studies were filled when trim and fill tests were performed to eliminate the influence of the publication bias in the random effects model, and the overall OR (2.49, 95% CI: 1.18-5.26) was still significant, although it was slightly smaller than that in the crude meta-analysis (4.67, 95% CI: 2.66-8.22), indicating the existence of a strong association between *APC* promoter methylation and Lung cancer. The pooled sensitivity, specificity and AUC of *APC* methylation test in the present meta-analysis were 0.548, 0.78 and 0.64, respectively, which revealed that *APC* methylation status is a good biomarker in NSCLC diagnosis.

Integrated analysis showed that the age at the diagnosis, autogenous or heterogeneous control, the ratio of the adenocarcinoma to squamous cell carcinoma, and primer set of CpG sites were the most important heterogeneity sources, while sample type (tissue or serum), proportion of males, proportion of stage I, and detection methods could not explain the heterogeneity.

Age was one of the most important heterogeneity sources from meta-regression analysis (beta = -0.3, P = 2.0×10-5), meanwhile, the OR in the younger subgroup (OR = 4.65) was greater than that in older subgroup (OR = 2.24). However, the same conclusion has not been obtained in TCGA NSCLC dataset. Furthermore, neither Ad nor Sc data supported that age would affect the odds ratio of the *APC* methylation to the risk of NSCLC in logistic regression model (P > 0.05), and thus much more evidence should be collected to make an eventual decision.

As to the contribution of Ad2Sc, both subgroup analysis and TCGA analysis showed significantly greater OR in high Ad2Sc than that in low Ad2Sc group, which suggested *APC* methylation test have better diagnosis performance for adenocarcinoma.

Since the late1980s, various studies showed that the same genetic/epigenetic alterations, such as DNA methylation, in the primitive tumors were also found in the circulating DNA of the patients affected with tumors ([39-41](#_ENREF_39)). Interestingly, in the present study, the odds ratio of the serum subgroup was greater than that of the tissue group and the AUC of *APC* methylation test for serum was greater than that for tissue in both meta- and microarray analysis, which indicated *APC* methylation test would be a promising serum biomarker for NSCLC diagnosis.

Meta-analysis has been widely applied in SNP-disease risk association study because SNPs had specific genome location, meanwhile, it is also booming in the realm of DNA methylation gradually. The primers for methylation detection have been considered when extracting information from studies, however, it was difficult to be analyzed in the following subgroup or meta-regression analysis sometimes since the high diversity of the primers used in each individual article. For example, at least 3 different primer sets were observed in the 17 studies we selected for meta-analysis (**Supplementary table S2**). Moreover, in order to expatiate on the divergence of different CpG sites, we collected the methylation signals of 5 CpGs from methylation 27K and 450K microarray dataset from TCGA project (Ad and Sc). It was found that the ORs of the five CpG sites were dramatically different **(Table 4)**. Subgroup analysis further showed significantly different ORs in different primer set. It reminded that future DNA methylation detection in case-control studies should be designed more accurately and comprehensively to some certain CpG site or blocks and the location information should be clearly noted when published to facilitate the re-analysis of the published data.

In conclusion, this integrated analysis of the pooled data provided strong evidence that the *APC* promoter hypermethylation is significantly associated with NSCLC, especially with adenocarcinoma, and it would be a promising diagnosis biomarker for lung adenocarcinoma with remote non-invasive media detection.

**Conflict of interest statement**

No potential conflicts of interest were disclosed for all the authors

**Reference**

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. CA: a cancer journal for clinicians. 2013;63:11-30.

2. Nesbitt JC, Putnam JB, Jr., Walsh GL, Roth JA, Mountain CF. Survival in early-stage non-small cell lung cancer. The Annals of thoracic surgery. 1995;60:466-72.

3. Gokul G, Khosla S. DNA methylation and cancer. Sub-cellular biochemistry. 2012;61:597-625.

4. Dietrich D, Hasinger O, Liebenberg V, Field JK, Kristiansen G, Soltermann A. DNA methylation of the homeobox genes PITX2 and SHOX2 predicts outcome in non-small-cell lung cancer patients. Diagnostic molecular pathology : the American journal of surgical pathology, part B. 2012;21:93-104.

5. Dietrich D, Kneip C, Raji O, Liloglou T, Seegebarth A, Schlegel T, et al. Performance evaluation of the DNA methylation biomarker SHOX2 for the aid in diagnosis of lung cancer based on the analysis of bronchial aspirates. International journal of oncology. 2012;40:825-32.

6. Fodde R, Kuipers J, Rosenberg C, Smits R, Kielman M, Gaspar C, et al. Mutations in the APC tumour suppressor gene cause chromosomal instability. Nature cell biology. 2001;3:433-8.

7. Deng G, Song GA, Pong E, Sleisenger M, Kim YS. Promoter methylation inhibits APC gene expression by causing changes in chromatin conformation and interfering with the binding of transcription factor CCAAT-binding factor. Cancer research. 2004;64:2692-8.

8. Gu J, Wen Y, Zhu S, Hua F, Zhao H, Xu H, et al. Association between P(16INK4a) promoter methylation and non-small cell lung cancer: a meta-analysis. PloS one. 2013;8:e60107.

9. Chen Y, Li J, Yu X, Li S, Zhang X, Mo Z, et al. APC gene hypermethylation and prostate cancer: a systematic review and meta-analysis. European journal of human genetics : EJHG. 2013.

10. Usadel H, Brabender J, Danenberg KD, Jeronimo C, Harden S, Engles J, et al. Quantitative adenomatous polyposis coli promoter methylation analysis in tumor tissue, serum, and plasma DNA of patients with lung cancer. Cancer research. 2002;62:371-5.

11. Tsou JA, Hagen JA, Carpenter CL, Laird-Offringa IA. DNA methylation analysis: a powerful new tool for lung cancer diagnosis. Oncogene. 2002;21:5450-61.

12. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003;327:557-60.

13. DerSimonian R, Laird N. Meta-analysis in clinical trials. Controlled clinical trials. 1986;7:177-88.

14. Huizenga HM, Visser I, Dolan CV. Testing overall and moderator effects in random effects meta-regression. The British journal of mathematical and statistical psychology. 2011;64:1-19.

15. Midgette AS, Stukel TA, Littenberg B. A meta-analytic method for summarizing diagnostic test performances: receiver-operating-characteristic-summary point estimates. Medical decision making : an international journal of the Society for Medical Decision Making. 1993;13:253-7.

16. Jones CM, Athanasiou T. Summary receiver operating characteristic curve analysis techniques in the evaluation of diagnostic tests. The Annals of thoracic surgery. 2005;79:16-20.

17. Sproul D, Nestor C, Culley J, Dickson JH, Dixon JM, Harrison DJ, et al. Transcriptionally repressed genes become aberrantly methylated and distinguish tumors of different lineages in breast cancer. Proceedings of the National Academy of Sciences of the United States of America. 2011;108:4364-9.

18. Begum S, Brait M, Dasgupta S, Ostrow KL, Zahurak M, Carvalho AL, et al. An epigenetic marker panel for detection of lung cancer using cell-free serum DNA. Clin Cancer Res. 2011;17:4494-503.

19. Brabender J, Usadel H, Danenberg KD, Metzger R, Schneider PM, Lord RV, et al. Adenomatous polyposis coli gene promoter hypermethylation in non-small cell lung cancer is associated with survival. Oncogene. 2001;20:3528-32.

20. Feng Q, Hawes SE, Stern JE, Wiens L, Lu H, Dong ZM, et al. DNA methylation in tumor and matched normal tissues from non-small cell lung cancer patients. Cancer Epidemiol Biomarkers Prev. 2008;17:645-54.

21. Jin M, Kawakami K, Fukui Y, Tsukioka S, Oda M, Watanabe G, et al. Different histological types of non-small cell lung cancer have distinct folate and DNA methylation levels. Cancer Sci. 2009;100:2325-30.

22. Kim DS, Cha SI, Lee JH, Lee YM, Choi JE, Kim MJ, et al. Aberrant DNA methylation profiles of non-small cell lung cancers in a Korean population. Lung Cancer. 2007;58:1-6.

23. Lin Q, Geng J, Ma K, Yu J, Sun J, Shen Z, et al. RASSF1A, APC, ESR1, ABCB1 and HOXC9, but not p16INK4A, DAPK1, PTEN and MT1G genes were frequently methylated in the stage i non-small cell lung cancer in China. J Cancer Res Clin Oncol. 2009;135:1675-84.

24. Pan SY, Xie EF, Shu YQ, Gao L, Zhang LX, Chen D, et al. [Methylation quantification of adenomatous polyposis coli (APC) gene promoter in plasma of lung cancer patients]. Ai zheng = Aizheng = Chinese journal of cancer. 2009;28:384-9.

25. Rykova EY, Skvortsova TE, Laktionov PP, Tamkovich SN, Bryzgunova OE, Starikov AV, et al. Investigation of tumor-derived extracellular DNA in blood of cancer patients by methylation-specific PCR. Nucleosides Nucleotides Nucleic Acids. 2004;23:855-9.

26. Shivapurkar N, Stastny V, Suzuki M, Wistuba, II, Li L, Zheng Y, et al. Application of a methylation gene panel by quantitative PCR for lung cancers. Cancer Letters. 2007;247:56-71.

27. Suzuki M, Shigematsu H, Iizasa T, Hiroshima K, Nakatani Y, Minna JD, et al. Exclusive mutation in epidermal growth factor receptor gene, HER-2, and KRAS, and synchronous methylation of nonsmall cell lung cancer. Cancer. 2006;106:2200-7.

28. Topaloglu O, Hoque MO, Tokumaru Y, Lee J, Ratovitski E, Sidransky D, et al. Detection of promoter hypermethylation of multiple genes in the tumor and bronchoalveolar lavage of patients with lung cancer. Clinical Cancer Research. 2004;10:2284-8.

29. Vallbohmer D, Brabender J, Yang D, Schneider PM, Metzger R, Danenberg KD, et al. DNA methyltransferases messenger RNA expression and aberrant methylation of CpG islands in non-small-cell lung cancer: association and prognostic value. Clin Lung Cancer. 2006;8:39-44.

30. Virmani AK, Rathi A, Sathyanarayana UG, Padar A, Huang CX, Cunnigham HT, et al. Aberrant methylation of the adenomatous polyposis coli (APC) gene promoter 1A in breast and lung carcinomas. Clinical cancer research : an official journal of the American Association for Cancer Research. 2001;7:1998-2004.

31. Wang Y, Zhang D, Zheng W, Luo J, Bai Y, Lu Z. Multiple gene methylation of nonsmall cell lung cancers evaluated with 3-dimensional microarray. Cancer. 2008;112:1325-36.

32. Yanagawa N, Tamura G, Oizumi H, Takahashi N, Shimazaki Y, Motoyama T. Promoter hypermethylation of tumor suppressor and tumor-related genes in non-small cell lung cancers. Cancer Science. 2003;94:589-92.

33. Zhang Y, Wang R, Song H, Huang G, Yi J, Zheng Y, et al. Methylation of multiple genes as a candidate biomarker in non-small cell lung cancer. Cancer letters. 2011;303:21-8.

34. Ignatov A, Bischoff J, Ignatov T, Schwarzenau C, Krebs T, Kuester D, et al. APC promoter hypermethylation is an early event in endometrial tumorigenesis. Cancer Sci. 2010;101:321-7.

35. Wu T, Giovannucci E, Welge J, Mallick P, Tang WY, Ho SM. Measurement of GSTP1 promoter methylation in body fluids may complement PSA screening: a meta-analysis. British journal of cancer. 2011;105:65-73.

36. Purnak T, Ozaslan E, Efe C. Molecular basis of colorectal cancer. The New England journal of medicine. 2010;362:1246; author reply -7.

37. Eissa S, Swellam M, El-Khouly IM, Kassim SK, Shehata H, Mansour A, et al. Aberrant methylation of RARbeta2 and APC genes in voided urine as molecular markers for early detection of bilharzial and nonbilharzial bladder cancer. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2011;20:1657-64.

38. Trock BJ, Brotzman MJ, Mangold LA, Bigley JW, Epstein JI, McLeod D, et al. Evaluation of GSTP1 and APC methylation as indicators for repeat biopsy in a high-risk cohort of men with negative initial prostate biopsies. BJU international. 2012;110:56-62.

39. Sozzi G, Conte D, Mariani L, Lo Vullo S, Roz L, Lombardo C, et al. Analysis of circulating tumor DNA in plasma at diagnosis and during follow-up of lung cancer patients. Cancer research. 2001;61:4675-8.

40. Esteller M, Sanchez-Cespedes M, Rosell R, Sidransky D, Baylin SB, Herman JG. Detection of aberrant promoter hypermethylation of tumor suppressor genes in serum DNA from non-small cell lung cancer patients. Cancer research. 1999;59:67-70.

41. Jahr S, Hentze H, Englisch S, Hardt D, Fackelmayer FO, Hesch RD, et al. DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. Cancer research. 2001;61:1659-65.

Table 1, Characteristics of eligible studies considered in the report

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Author (Published Year) | Sample Type | Agea (years) | Stages I % | Gender (M/F) | Patients  (M+/M-) | Control (M+/M-) | Methods | Aim | Multiple Target | Ad2Sc | Control  Design | ref. |
| Zhanget al (2011,China)b | tissue | 59 | 32.05 | 29/39 | 44/34 | 10/68 | MSP | Diagnose | Yes | 0.83 | hom | ([33](#_ENREF_33)) |
| Wang et al (2008, China) | tissue | - | - | 17/28 | 19/9 | 1/11 | qMSP | Diagnose | Yes | 2.14 | heter | ([31](#_ENREF_31)) |
| Jin et al (2009, Japan) | tissue | 66.7 | - | 17/24 | 27/45 | 22/41 | qMSP | Non-diagnose | Yes | 1.87 | heter | ([21](#_ENREF_21)) |
| Feng et al (2008, USA) | tissue | 64.3 | 42.86 | 26/49 | 26/23 | 21/28 | qMSP | Diagnose | Yes | 1.43 | hom | ([20](#_ENREF_20)) |
| Brabender et al (2001, USA) | tissue | 63.3 | 49.45 | 69/91 | 86/5 | 80/11 | qMSP | Non-diagnose | Single | 0.77 | hom | ([19](#_ENREF_19)) |
| Virmani et al (2001, USA) | tissue | - | - | - | 22/26 | 0/18 | MSP | Diagnose | Yes | NA | heter | ([30](#_ENREF_30)) |
| Yanagawa et al (2003, Japan) | tissue | 67.3 | 66.67 | 18/25 | 28/47 | 36/39 | MSP | Diagnose | Yes | 1.48 | hom | ([32](#_ENREF_32)) |
| Topaloglu et al (2004, USA) | tissue | - | 54.84 | - | 17/14 | 5/17 | qMSP | Diagnose | Yes | 3.00 | heter | ([28](#_ENREF_28)) |
| Kim et al (2007, Korea) | tissue | 63 | 56.57 | 64/79 | 48/41 | 33/66 | MSP | Non-diagnose | Yes | 0.62 | hom | ([22](#_ENREF_22)) |
| Vallbohmer et al (2006, USA) | tissue | 63 | 49.45 | 69/91 | 86/5 | 80/3 | PCR | Non-diagnose | Yes | 0.77 | hom | ([29](#_ENREF_29)) |
| Lin et al (2009, China) | tissue | 61.1 | 100.00 | 20/31 | 49/75 | 2/24 | MSP | Diagnose | Yes | 1.84 | heter | ([23](#_ENREF_23)) |
| Shivapurkar et al (2007, USA) | tissue | - | - | - | 35/5 | 23/17 | qMSP | Diagnose | Yes | 1.22 | heter | ([26](#_ENREF_26)) |
| Suzuki et al (2006, Japan) | tissue | 64 | 34.00 | 33/49 | 53/97 | 3/57 | MSP | Non-diagnose | Yes | NA | heter | ([27](#_ENREF_27)) |
| Zhang et al (2011, China) b | serum | - | - | - | 54/56 | 5/45 | MSP | Diagnose | Yes | NA | heter | ([33](#_ENREF_33)) |
| Pan et al (2009,China) | serum | 53 | - | 17/26 | 40/38 | 0/31 | qMSP | Diagnose | Single | NA | heter | ([24](#_ENREF_24)) |
| Begum et al (2011, USA) | serum | 65 | - | 10/19 | 12/64 | 3/27 | qMSP | Diagnose | Yes | NA | heter | ([18](#_ENREF_18)) |
| Rykova et al (2004, Russia) | serum | NA | - | - | 3/6 | 0/16 | MSP | Diagnose | Yes | NA | heter | ([25](#_ENREF_25)) |
| Usadel et al (2002, USA) | serum | 64.2 | - | - | 42/47 | 0/50 | qMSP | Diagnose | Single | NA | heter | ([10](#_ENREF_10)) |

a mean or median age from articles;

b with two records since there are tissue and serum data simultaneously in this article.

Abbreviations: Ad2Sc, the ratio of the adenocarcinoma to squamous cell carcinoma; Sample type represents the samples analyzed; hom, homogenous control; heter, heterogeneous control. MSP, qualitative methylation detection method; qMSP, quantitative detection method.

Table 2. Subgroup analysis for the main potential confounding factors with random effect model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Subgroup | No. of Study | OR | 95% CI | Q | I2 | P-value |
| Overall | 18 | 4.67 | 2.65-8.21 | 73.99 | 77.00% |  |
| Age≤ 65 | 9 | 5.03 | 2.53-10.0 | 27.96 | 71.40% |  |
| Age>65 | 3 | 0.91 | 0.57-1.41 | 2.21 | 9.400% | **<0.0001** |
| Stage I>49.5% | 5 | 4.11 | 1.90-8.91 | 12.76 | 68.60% |  |
| Stage I≤ 49.5% | 4 | 2.81 | 0.87-9.09 | 19.42 | 84.60% | 0.5944 |
| M2F≤ 69% | 6 | 5.98 | 2.04-17.53 | 16.66 | 70.00% |  |
| M2F> 69% | 6 | 2.13 | 0.99-4.55 | 29.05 | 82.80% | 0.1246 |
| MSP | 8 | 5.16 | 2.01-13.26 | 44.61 | 84.30% |  |
| qMSP | 10 | 4.32 | 2.08-8.94 | 29.28 | 69.30% | 0.7685 |
| Diagnose | 13 | 6.79 | 2.99-15.44 | 59.54 | 79.80% |  |
| Non-diagnose | 5 | 2.59 | 1.33-5.05 | 11.56 | 65.40% | 0.0745 |
| Multiple targets | 15 | 4.08 | 2.28-7.34 | 62.99 | 77.80% |  |
| Single target | 3 | 18.72 | 1.23-283 | 9.03 | 77.80% | 0.2836 |
| Heterogeneous | 12 | 8.33 | 3.77-18.39 | 35.71 | 69.20% |  |
| Autogenous | 6 | 2.25 | 1.06-4.77 | 27.19 | 81.60% | **0.0187** |
| Serum | 5 | 11.54 | 2.87-46.40 | 10.4 | 61.50% |  |
| Tissue | 13 | 3.72 | 2.03-6.78 | 55.18 | 78.30% | 0.14 |
| Ad2Sc < 2 | 9 | 2.46 | 1.35-4.48 | 35.79 | 77.00% |  |
| Ad2Sc >= 2 | 2 | 17.1 | 4.68-62.7 | 0.11 | 0.000% | **0.0077** |
| Primer Set I | 5 | 5.41 | 2.43-12.04 | 13.71 | 70.80% |  |
| Primer Set II | 4 | 1.82 | 1.05-3.13 | 4.57 | 34.30% | **0.0137$** |

Bold P-values lower than 0.05 indicate significant differences between groups (random effect model, d.f. = 1).

$the serum groups and the studies with less than 50 samples were removed.

Table 3. Meta-regression analysis for the main potential interference factors with random-effects model

|  |  |  |  |
| --- | --- | --- | --- |
| Subgroup | Coef. (95% CI) | P-value | τ2 |
| Sample type | -1.03 (-2.4, 0.34) | 0.14 | 0.90 |
| Age | -0.3 (-0.44, -0.16) | **2.0×10-5** | 0.18 |
| Proportion of Stage I | -0.01 (-0.05, 0.03) | 0.608 | 0.79 |
| Ratio of Male to Female | -0.69 (-8.1, 6.71) | 0.855 | 0.98 |
| Detection Methods | -0.09 (-1.28, 1.1) | 0.88 | 1.11 |
| Study Aim | -0.82 (-2.05, 0.41) | 0.19 | 1.07 |
| Single/Multiple Targets | 1.05 (-0.71, 2.81) | 0.243 | 1.01 |
| Hetero/Autogeous Control | -1.25 (-2.35, -0.15) | 0.026 | 0.89 |
| Ad2Sc | 0.44 (-0.56 , 1.44 ) | 0.387 | 0.89 |
| Primer Set | -1.02 (-1.02, -2.02) | **0.05** | 0.35 |

Bold P-values lower than 0.05 indicate the item would be a significant heterogeneity.

Table 4. Differential APC methylation, odds ratio, AUC between adenocarcinoma, squamous cell carcinoma and their counterparts

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Adenocarcinoma | | | | | | | | Squamous Cell Carcinoma | | | | | | | |
| CpG Site | MCaM | MCoM | P-value¥ | FDR¥ | OR† | P-value† | 95%CI† | AUC† | MCaM | MCoM | P-value¥ | FDR¥ | OR† | P-value† | 95%CI† | AUC† |
| (N=535) | (N=56) | (N=386) | (N=70) |
| cg15020645 | 0.26(40.7%) | 0.13(0%) | 3.5×10-32 | **1.0×10-31** | 190.6 | **7.7×10-6** | 22.65-2321 | 0.72 | 0.13(14.77%) | 0.11(0%) | 0.087466 | 0.131199 | 3.16 | 0.406 | 0.28-68.72 | 0.61 |
| cg16970232 | 0.3(45.2%) | 0.11(0%) | 5.0×10-38 | **3.0×10-37** | 108.9 | **5.1×10-6** | 17.64-1043 | 0.73 | 0.15(18.91%) | 0.09(0%) | 2.7×10-7 | **1.6×10-6** | 7.54 | **0.035** | 1.39-64.07 | 0.45 |
| cg20311501 | 0.33(48.4%) | 0.16(5.3%) | 1.4×10-22 | **2.1×10-22** | 61.56 | **4.96×10-6** | 11.94-420 | 0.73 | 0.18(19.95%) | 0.14(0%) | 0.001955 | **0.003909** | 2.48 | 0.257 | 0.57-13.74 | 0.49 |
| cg21634602 | 0.33(47.4%) | 0.16(7.1%) | 3.6×10-17 | **4.3×10-17** | 23.34 | **3.6×10-5** | 5.75-116.0 | 0.71 | 0.16(20.47%) | 0.14(7.14%) | 0.222306 | 0.266767 | 1.27 | 0.726 | 0.35-5.42 | 0.53 |
| cg24332422 | 0.26(40.5%) | 0.16(0%) | 1.0×10-26 | **2.0×10-26** | 223.6 | **2.81×10-5** | 21.11-3463 | 0.71 | 0.16(17.36%) | 0.15(0%) | 0.338755 | 0.338755 | 1.6 | 0.656 | 0.23-14.30 | 0.52 |

MCaM, MCoM represent the mean of case methylation (Beta) and mean of control methylation (Beta). Methylation level are calculated with formula: Beta= (M/M+U).

¥ P-value from t-test or Wilcoxon sum-rank test and P-values after false discovery rate (FDR adjustment). Significant P-values after FDR are bolded which indicate significant different methylation between case and control

† OR and its 95% CI, P-value and AUC from logistic regression analysis or prediction model

**Figure Legends**

Figure 1. Combined estimates for the association between APC promoter hypermethylation and NSCLC with forest plot.

Author, year, country of the studies and methylated (M) and total number of the sample (T) in case and control, combined odds ratio (OR) with 95% confidence region were labeled in the left column of the figure. The DerSimonian-Laird estimator and Mantel-Haenszel method were selected to conduct combination estimation for random effects model and fixed effects model, respectively.

Figure 2. Subgroup meta-analysis, cumulative analysis and SROC estimation for the relationship between APC promoter hypermethylation and NSCLC. A-D, subgroup meta-analysis based on age, control type, percentage of adenocarcinoma in total samples and primer set, respectively. E, cumulative meta-analysis of studies ordered chronologically by publication year with random effect model. F, Summary receiver operating characteristics (SROC) of APC methylation test in NSCLC

**Supplementary Figure 1-3.**

Supplementary Fig. 1. Funnel plot to diagnosis of the publication bias

Supplementary Fig. 2. Combined estimates for the association between APC methylation and NSCLC after trim-fill treatment

Supplementary Fig. 3. Sensitivity analyses of the overall effect by omitting a single study

**Supplementary tables 1-5.**

Supplementary Table S1. TCGA probe information in this study.

Supplementary Table S2. Three kinds of primers of present 17 studies.

Supplementary Table S3. The fluctuation of odds ratio in vignettes of different proportion of Ad.

Supplementary Table S4. Odds ratio difference between heterogeneous and autogenous samples in vignettes of different proportion of Ad.

Supplementary Table S5. Interaction estimation between CpG methylation and age, gender, TNM in Ad and Sc.

**Supplementary tables for reviewers only**

Special Table 1. Clinical characteristic of the NSCLC samples from TCGA project

Special Table 2. Epidemiological characteristic of the NSCLC samples from TCGA project