# **[Quantitative assessment of the diagnostic role of FHIT promoter methylation in non-small cell lung cancer](https://scholar.google.com/citations?view_op=view_citation&hl=en&user=4tIViCAAAAAJ&sortby=pubdate&citation_for_view=4tIViCAAAAAJ:-jrNzM816MMC)**

Xin Geng1, Shicheng Guo2, Weilin Pu2, Lixing Tan3, Zhouyi Lu1, An Wang1, Huijun Zhang1, Sidi Chen2, Yong Xia3, Li Jin2, Xiaofeng Chen1\*, Jiucun Wang1\*

1. Department of Cardiothoracic Surgery, Huashan Hospital, Fudan University, Shanghai, China
2. State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200433, China
3. College of Acupuncture-Moxibustion and Tuina, Shanghai University of Traditional Chinese Medicine, 1200 Cailun Road, Shanghai, 201203, China
4. Fudan-Taizhou Institute of Health Sciences, 1 Yaocheng Road, Taizhou, Jiangsu 225300, China

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

Page number: 9

Word counts: 3028

Figure number: 3

Table number: 3

**Corresponding authors:**

Xiaofeng Chen, Department of Cardiothoracic Surgery, Huashan Hospital, Fudan University, Shanghai 200433, China, Phone: +86-21-55664885, Fax: +86-21-55664885, E-mail: [lijin.fudan@gmail.com](mailto:lijin.fudan@gmail.com)

Jiucun 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

## 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.

FHIT gene, whose official full name is fragile histidine triad, is a member of the histidine triad gene family, encodes Hydrolase of Ap3A [7], which is a signal in the timing of cell division[8]. The gene encompasses the common fragile site FRA3B on chromosome 3, where carcinogen-induced damage can lead to translocations and aberrant transcripts of this gene[9]. FHIT loss was observed in 64% of non-small-cell lung carcinoma patients and was significantly associated with squamous cell carcinoma and poor tumor grade[10]. In addition, aberrant transcripts from this gene have been found in about half of all gastric[11], esophageal[12], and colon carcinomas[13]. So, FHIT now is considered as a cancer suppressor gene.

In this article, we conducted a meta-analysis of the sensitivity and specificity of FHIT methylation on NSCLC diagnosis. The factors which lend heterogeneity to the sensitivity and specificity were identified with meta-regression. 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 FHIT methylation test in NSCLC. Therefore, an integrated analysis of all these existing data was conducted to come to unbiased conclusions on the relationship between FHIT methylation and NSCLC

**Results**

**Study characteristics**

The electronic search strategy identified 365 potentially relevant articles (PubMed, 229; Web of science, 549; OVID Embase, 170; 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, 12 studies with data on the relationship between FHIT gene promoter methylation and NSCLC were pooled for analysis (Table 1). All these articles were written in English. In total, 1090 lung cancer tissues/serum and 1029 normal counterpart tissues/serum were collected. The age of the subjects in the 11 studies ranged from 28 to 86 years, with mean or median ranging from 53 to 68 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 12 studies, the proportions of stage I samples differed from 0 to 67.33%, and the percentage of male individuals in the NSCLC samples has a range of 65.2 to 83.8%. For the experimental methods to explore FHIT promoter methylation status, 10 of 12 inclusions used methylation-specific polymerase chain reaction (MSP), while others used quantitative MSP (qMSP, such as Methylight, Pyrosequencing, and so on).

**Meta-analysis**

The ORs for FHIT methylation in cancer tissues/plasma compared with that in normal controls were 3.43 (95% CI: 1.85 to 6.36, z =3.92, P<0.0001) in random effects model pooled, and 2.03 (95% CI: 1.60 to 2.57, z =5.85, P<0.0001) in fixed effects model, demonstrating a statistically significant increasing in likelihood of methylation in lung cancer tissues comparing to controls (Figure 1).

Subgroup analyses were conducted for different subtypes, which included sample types (tissue or serum), age, counterpart categories (autogenous or heterogeneous), proportion of stage I, proportion of stage I and II, 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 younger (51.4, 95% CI: 12.07 to 221.80) and older (3.30, 95% CI: 1.64 to 6.64) subgroup (P < 0.0001) (Figure 2A). Both tissue and serum groups showed significant association between FHIT methylation and NSCLC (OR = 3.68 and 3.89, respectively) which suggested that FHIT methylation can be taken as a potential biomarker for NSCLC diagnosis using either tissue or serum samples. The OR in studies aiming at diagnosis (OR = 1.50) is significant less than the OR in the non-diagnosis group (OR = 6.92), which might be caused by an unbalanced distribution in the proportion of early stage samples (P =0.0008, Wilcoxon rank-sum test). The subgroup of low proportion of stage I and II had a larger OR than that of high proportion of stage I and II (OR=29.58, 2.67, respectively, p=0.0038, Wilcoxon rank-sum test) (Table 2), which support the FHIT’s role as a cancer suppressor gene. No significant difference was found between subgroups of MSP and qMSP (P = 0.657), which suggested both of the methods were equivalent in methylation detection (Table 2). In addition, ratio of male to female, counterpart categories, ratio of adenocarcinoma to squamous, and other factors are not the source of heterogeneity.

Meta-regression revealed that heterogeneity exists among 13 studies (I2 = 78.8%, P < 0.0001) (Figure 1), whereas age, aim and stage is significant heterogeneity resource. The trend in ORs was inversely correlated with age (beta = -3.92, P = 0.05), and age counted for 40.03% total variances. This result is consistent with the subgroup analysis, in which the OR of the older group (OR = 3.30) was smaller than the younger group (OR = 51.4). The aim and stage is also an important heterogeneity source (P = 0.028 and 0.006), explaining about 51.44% and 17.07% of overall heterogeneity, respectively. Other factors such as sample type, proportion of males, proportion of stage I, detection methods, and other factors fail to explain the heterogeneity counting for type I error at level of 0.05 (Table 3).

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

Pooled sensitivity and specificity were 0.357(95%CI: 0.320-0.395, P<0.0001) and 0.872 (95%CI: 0.793-0.924, P<0.0001) for all the studies based on the presupposition of the fixed effects model. The sensitivity of the tissue group was similar with that of the serum group, 0.363(95% CI: 0.326-0.402, P<0.0001) versus 0.337(95% CI: 0.270-0.412, P<0.0001) , while the specificity of the tissue group was slightly lower than that of the serum group, 0.883(95% CI: 0.791-0.937, P<0.0001) versus 0.902(95% CI: 0.694-0.902, P<0.0001), which suggested the biomarker would have adequate ability both in tissues and 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 0.44, suggesting a median ability for NSCLC diagnosis (Figure 2F). Meanwhile, the AUC of the SROC for the serum and the tissue group was 0.364 and 0.437 respectively, showing slightly better performances for the FHIT 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 based on Linear regression test of funnel plot asymmetry showed significant publication bias (Egger test, z =2.7571, P=0.01865) and 7 studies 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 2.09 (95% CI: 1.10 to 3.96, P =0.0242) in the random effects model and 1.83 (95% CI: 1.45 to 2.32, P < 0.0001) in the fixed effects model. Both results demonstrate a significantly positive association between FHIT 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 2.97 (95% CI: 1.64 to 5.37) and 4.10 (95% CI: 2.17 to 7.76) in the random effects method, which suggested that combined OR was consistent and reliable (Additional file 2: Figure S3).

**Discussion**

The FHIT gene has been reported as an important tumor suppressor in colorectal cancer and the aberrant of FHIT methylation had been reported in numeric for cancers, such as bladder, prostate, breast and lung cancer. However, the diagnostic role of the methylation status of the FHIT gene in lung cancer lacks quantitative assessment. We therefore performed an integrated analysis to quantify the ability for the FHIT promoter methylation test in NSCLC diagnosis, and a significant association was identified between FHIT methylation and NSCLC (OR = 3.43, P < ？？？). Four imputed studies were filled when trim and fill tests were performed to eliminate the influence of publication bias on the random effects model, and the overall OR (2.09, 95% CI: 1.10 to 3.96) was still significant, although it was slightly smaller than that in the crude meta-analysis (3.43, 95% CI: 1.85 to 6.36), indicating the existence of a strong association between FHIT promoter methylation and lung cancer. The pooled sensitivity, specificity of the FHIT methylation test in the present meta-analysis were 0.357 and 0.872 respectively, which revealed that FHIT methylation status is a good biomarker in NSCLC diagnosis.

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

**Conclusion**  
In conclusion, this integrated analysis of the pooled data provides strong evidence that the methylation status of the FHIT promoter is strongly associated with NSCLC, especially for adenocarcinoma. Therefore, the FHIT 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 by December 2014. The study used a subject and text word strategy with (FHIT OR AP3Aase OR FRA3B) AND (lung cancer) 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 FHIT 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 WP, ZL and AW. 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 FHIT 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. 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. 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. 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. Thus, SROC analysis was applied to meta-analysis of diagnostic tests. The SROC curve shows the performance of the diagnostic ability of FHIT 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

**Reference**

[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] Barnes, L.D. et al . (1996) FHIT, a putative tumour suppressor in humans, is a dinucleotide 5’,5’’’-P1,P3-triphosphate hydrolase. Biochemistry 35, 11529–11535

[8]. Nishimura, A., The timing of cell division: Ap4A as a signal. Trends Biochem Sci, 1998. 23(5): p. 157-9.

[9]. Huebner, K., et al., The role of the FHIT/FRA3B locus in cancer. Annu Rev Genet, 1998. 32: p. 7-31.

[10]. Younes, S.F., et al., FHIT, EGFR, and MSH2: possible etiopathologic, prognostic, and predictive role in non-small cell lung carcinoma in Egyptian patients. Appl Immunohistochem Mol Morphol, 2014. 22(4): p. 275-83.

[11]. Wang, H.L., et al., Abnormal FHIT protein expression may be correlated with poor prognosis in gastric cancer: a meta-analysis. Tumour Biol, 2014. 35(7): p. 6815-21.

[12]. Guo, X.Q., et al., DNA methylation and loss of protein expression in esophageal squamous cell carcinogenesis of high-risk area. J Exp Clin Cancer Res, 2007. 26(4): p. 587-94.

[13]. Kapitanovic, S., et al., Reduced FHIT expression is associated with tumor progression in sporadic colon adenocarcinoma. Exp Mol Pathol, 2014. 96(1): p. 92-7.

## Figure legends

## Tables

Table 1.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| group | author | sampletype | age | stage1 | stage2 | genderatio | event.e | n.e | event.c | n.c | methods | aim | Multipletarget | controldesign | year | Ad2Sc | smoking |
| 1 | Haroun et al (2014, Egypt) | Tissue | 53 | 0.178571 | 0.571429 | 0.714286 | 15 | 28 | 1 | 28 | qMSP | Non-Diagnosis | Multi | hom | 2014 | 1.777778 | 2.11 |
| 2 | Li et al (2014, China) | Non-tissue | 53.15 | 0.267857 | 0.392857 | NA | 19 | 56 | 0 | 56 | MSP | Non-Diagnosis | Multi | heter | 2014 | 0.586207 | NA |
| 3 | Li et al (2010, China) | Non-tissue | 55.03 | NA | NA | 0.710526 | 42 | 123 | 0 | 105 | MSP | Non-Diagnosis | Single | heter | 2010 | 0.602941 | 1 |
| 4 | Zhang et al (2011, China) | Tissue | 59 | 0.320513 | 0.74359 | 0.74359 | 1 | 40 | 1 | 40 | MSP | Diagnosis | Multi | hom | 2011 | 0.833333 | 1.68 |
| 5 | Fischer et al (2007, Germany) | Non-tissue | 60.9 | 0 | 0 | 0.652174 | 43 | 92 | 0 | 7 | MSP | Non-Diagnosis | Multi | heter | 2007 | 1.714286 | NA |
| 6 | Zochbauer et al (2001, Multi） | Tissue | 61 | 0.570093 | 0.766355 | 0.71028 | 40 | 107 | 9 | 104 | MSP | Non-Diagnosis | Single | hom | 2001 | 1.046512 | 10.88 |
| 7 | Kim.D et al (2007, Korea) | Tissue | 63 | 0.565657 | 0.747475 | 0.808081 | 34 | 99 | 17 | 99 | MSP | Non-Diagnosis | Multi | hom | 2007 | 0.622951 | 3.95 |
| 8 | Verri et al (2009, Multi) | Tissue | 63.9 | 0.64624 | NA | 0.83844 | 84 | 229 | 68 | 208 | MSP | Non-Diagnosis | Single | hom | 2009 | 1.109091 | 20.75 |
| 9 | Yanagawa et al (2007, Japan) | Tissue | 68.1 | 0.673267 | 0.742574 | 0.712871 | 34 | 101 | 7 | 101 | MSP | Non-Diagnosis | Multi | hom | 2007 | 1.589744 | 2.6 |
| 10 | Hsu et al (2007, Taiwan) | Tissue | 69 | NA | 0.650794 | 0.714286 | 22 | 57 | 9 | 63 | qMSP | Diagnosis | Multi | hom | 2007 | 0.759259 | 1.94 |
| 11 | Fraipont et al (2005, France) | Non-tissue | NA | NA | NA | NA | 6 | 16 | 18 | 56 | MSP | Diagnosis | Multi | heter | 2005 | NA | NA |
| 12 | Hsu et al (2007, Taiwan) | Non-tissue | NA | NA | 0.650794 | 0.707071 | 18 | 57 | 7 | 35 | qMSP | Diagnosis | Multi | heter | 2007 | 0.759259 | 1.4 |
| 13 | Kim.H et al (2004, Korea) | Non-tissue | NA | 0.590909 | 1 | 0.665094 | 19 | 85 | 36 | 127 | MSP | Diagnosis | Multi | heter | 2004 | 0.72093 | NA |

Table 2.

Table 3.

Table 4.

Figure 1.

Figure 2.

Figure 3.