

Aberrant methylation of FHIT can be a diagnostic biomarker for NSCLC in Asian population

Xin Geng^{1†}, Weilin Pu^{2†}, Yulong Tan¹, Zhouyi Lu³, An Wang³, Lixing Tan², Sidi Chen², Shicheng Guo^{4*}, Jiucun Wang^{2*}, Xiaofeng Chen^{1*}

1. Department of Cardiothoracic Surgery, Huashan Hospital, Fudan University, Shanghai 200032, 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. Department of Chest Surgery, Shanghai Pulmonary Hospital, Shanghai 200433, China
4. Department of Bioengineering, University of California at San Diego, 9500 Gilman Drive, MC0412, La Jolla, CA 92093-0412

Corresponding authors:

Xiaofeng Chen, Department of Cardiothoracic Surgery, Huashan Hospital, Fudan University, Shanghai 200032, China, Phone: +86-21-52888299, E-mail: cxf3166@126.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.

Key words

FHIT, DNA methylation, Non-small cell lung cancer, NSCLC, Diagnosis

Abstract

Aberrant methylation of CpG islands acquired in promoter regions plays an important role in carcinogenesis. Accumulated evidence demonstrates *FHIT* gene promoter hypermethylation is involved in non-small cell lung carcinoma (NSCLC), indicating it may be a potential biomarker for NSCLC diagnosis. The pooled odds of *FHIT* promoter methylation in lung cancer tissues versus normal controls were calculated by meta-analysis method. Simultaneously, four independent DNA methylation datasets of NSCLC from TCGA and GEO database were downloaded and analyzed to validate the results. Thirteen studies, including 2119 samples were included in this meta-analysis. The pooled odds ratio of *FHIT* promoter methylation in cancer samples was 3.43 (95% CI: 1.85 to 6.36) compared with that in controls. In subgroup analysis, significant difference of *FHIT* gene promoter methylation status in NSCLC and controls was found in Asian populations but not in white population. In validation stage, 126 paired samples from TCGA, 568 cancer tissues and 256 normal controls from GEO database were analyzed and all 8 CpG sites near the promoter region of *FHIT* gene were not significantly differentially methylated. Thus the diagnostic role of *FHIT* gene in the lung cancer may be relatively limited in the white population but useful in the Asians.

Introduction

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 the five-year survival rate can be increased from 5% to 63% with the early stage of NSCLC thus showing the importance of early diagnosis of NSCLC[2, 3]. DNA methylation is one of the epigenetic modifications in eukaryote, which regulates genes and microRNAs expression[4] and alternative splicing[5]. It has been observed and confirmed that DNA methylation change is widespread in tumor tissues. Hence, with the numerous advantages like stable chemical property, detection ability in remote patient media, quantitative signal, relatively low cost in detection and so on[6], DNA methylation could be a promising biomarker in early cancer detection.

FHIT (fragile histidine triad) is a member of the histidine triad gene family, which encodes Hydrolase of Ap3A[7], and the *Fhit*-Ap3A enzyme-substrate complex appears to be the tumor suppressor signal[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]. In summary, FHIT is now considered as a cancer suppressor gene and the loss or aberrant transcripts of FHIT may be associated with carcinogenesis.

In this article, we firstly conducted a meta-analysis to evaluate the ability to use FHIT methylation level for early lung cancer diagnosis. The factors which lead heterogeneity to the ORs were identified with subgroup analyses and meta-regression. We also found that The Cancer Genome Atlas project (TCGA) as well as the Gene Expression Omnibus (GEO) database had collected hundreds of NSCLC samples with whole genome DNA methylation datasets and comprehensive clinical information, providing an additional resource for validation and without publication bias[14]. Several studies have showed the improved robustness of combining data from papers and databases[15, 16]. Therefore, in our work, we innovatively integrated these microarray data and published articles to evaluate and validate the diagnostic ability of FHIT methylation test in NSCLC.

Results

Study characteristics

The electronic search strategy identified 948 potentially relevant articles (Medline, 229; Web of science, 549; 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) [17-28]. The selection of the criteria was described in method section. All these articles were written in English. In total, 1090 lung cancer tissues/plasma and 1029 normal counterpart tissues/plasma were collected. The age of the subjects in the 12 studies ranged from 28 to 86, with mean or median age ranging from 53 to 68. As for the study aim, 4 articles were especially aiming at diagnosis, while the others were for prognosis or pathogenesis. 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% (Table 1). 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 (Methylight). Three kinds of methylation detection primers or probes were found to be utilized for most of the 12 studies (Table S1).

Meta-analysis

The odd ratio (OR) 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 plasma), age, counterpart categories (autogenous or heterogeneous), proportion of stage I, proportion of stage I and II, proportion of male, aim of the study (for diagnosis or non-diagnosis), ratio of adenocarcinoma to squamous (Ad2Sc) and other possible confounding factors (Table S2). 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 (Figure 2A) and between the ORs of higher (29.58, 95% CI: 6.82 to 128.37) and lower (2.67, 95% CI: 1.32 to 5.40) proportion of stage I and II subgroup (Figure 2B). Interestingly, difference was found between Asian (3.50, 95% CI: 1.50 to 8.14, $P = 0.005$) and white population (2.55, 95% CI: 0.86 to 7.57, $P = 0.09$) subgroup (Figure 2C), and the differential methylation in white population is not significant, indicating that diagnostic ability of *FHIT* methylation might be limited in white population. Both tissue and plasma groups showed significant association between *FHIT* methylation and NSCLC (OR = 3.68 and 3.89, respectively) (Figure 2D) which suggested that *FHIT* methylation can be taken as a potential biomarker for NSCLC diagnosis using either tissue or plasma samples. No significant difference was found between subgroups of MSP and qMSP (OR = 3.22 and 4.31, respectively), which suggested both of the methods were equivalent in methylation detection. In addition, proportion of male, counterpart categories, ratio of adenocarcinoma to squamous, the aim of the studies, the primer set and other factors were not the sources of heterogeneity (Table S2).

Analysis revealed that heterogeneity existed among 13 studies ($I^2 = 78.8\%$, $Q^2 = 61.05$, $P < 0.0001$) (Figure 1), whereas age, aim and stage were significant heterogeneity resources. The trend in ORs was inversely correlated with age ($\beta = -3.92$, $P = 0.05$), and age counted for 40.03% of total variance. The result is consistent with the subgroup analysis, in which the OR of the elder group (OR = 3.30) was smaller than the younger group (OR = 51.4). The aim and stage were also two important heterogeneity sources ($P = 0.028$ and 0.006), explaining about 51.44% and 17.07% of overall heterogeneity respectively. As shown in Table 1, the subgroup of aiming at diagnosis or non-diagnosis didn't concordant with the subgroup of different stages, indicating aim was an independent heterogeneity source. Other factors such as sample type, proportion of males, detection methods, and other factors failed to explain the heterogeneity counting for type I error at level of 0.05 (Table 2).

Bias analysis and robust estimation of pooled OR

A funnel plot of methylation status of lung cancer tissues versus normal tissues based on linear regression test showed significant publication bias (Egger test, $z=2.7571$, $P=0.01865$) and 7 studies exceeded the 95% confidence intervals (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 was 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 demonstrated a significantly positive association between FHIT methylation and NSCLC (Figure S2).

In sensitivity analysis aimed 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 (Figure S3).

A cumulative meta-analysis at the time of the published literature was also conducted, and we found the OR was tending to be stable (Figure S4). The stable result indicated that our meta-analysis might be more credible when more incoming researches added.

Similarly, the influence on meta-regression was determined by omitting one study each time to explore heterogeneity source. The ORs of omitting each study were similar and indicating the meta-analysis result was credible.

Validation with independent TCGA and GEO lung cancer datasets

In order to make independent validation of the above meta-analysis results, we searched and collected several datasets from TCGA (The Cancer Genome Atlas) and GEO (Gene Expression Omnibus). For datasets from TCGA, we downloaded lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) methylation data. Eight CpG sites located in the same CpG islands as the three sets of primers (Table 3) were obtained after data filtering. In LUAD dataset, though five out of the eight CpG sites had a p -value <0.05 both in Wilcoxon rank sum test and logistic regression, the absolute mean difference were less than 0.1 for all (Table 3). As a result, none of the eight CpG sites could be considered as differentially methylated between lung adenocarcinoma tissues and adjacent normal tissues. Concordantly, in the LUSC dataset, 3 out of 8 CpG sites showed a p -value <0.05 after multiple correction but the absolute mean difference of the 3 CpG sites were less than 0.1, which was the same as in the LUAD dataset and couldn't be regarded as significant methylated as well (Table 3).

Because of the conflicting results came from the meta-analysis and TCGA dataset, we obtained other datasets from the GEO (Gene Expression Omnibus) website. The first dataset was the combination of GSE39279 and GSE52401. In GSE39279 dataset, 322 lung adenocarcinoma and 122 lung squamous cell tissues were included. While in GSE52401 dataset, a total of 244 normal lung tissues were included, and both of the dataset used the Illumina HumanMethylation450 BeadChip for methylation measurement. The two datasets were combined and a total of 444 tumor tissues and 244 normal tissues were included in the subsequent analysis. We performed the same analysis as in TCGA dataset and the result was almost the same. Due to the large number of samples, all the p-values of the eight CpG sites were less than 0.05 after multiple corrections (Table S3). However, the absolute mean difference of the eight CpG sites were also less than 0.1 and still couldn't be considered as significant methylated CpG sites.

Moreover, we downloaded GSE56044 with 124 NSCLC tissues and 12 adjacent normal tissues for further validation. GSE56044 didn't have clinical information on the subtypes of NSCLC and thus we just utilized NSCLC tissues for subsequent comparison. And the result was unsurprisingly the same as the two datasets mentioned before, showing no significant methylation state of the eight CpG sites (Figure S5).

Gene Expression data with TCGA RNA-Seq dataset

DNA methylation played a key factor in regulating gene expression. It may be informative to see if the gene expression of *FHIT* was changed due to the very different results obtained from microarray data and the meta-analysis. We downloaded level 3 RNA-Seq data of lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) from TCGA project. Per million mapped reads (RPKM) was chosen as the measurement for gene expression. Interestingly, however, after calculating the fold change and p-value with multiple correction, no significantly differential expression was shown both in LUAD ($P = 0.58$, Fold change: 1.30) and LUSC ($P = 5.7E-7$, Fold change: 1.86) when compared with the adjacent normal tissues. Furthermore, the expression level of *FHIT* is relatively low in LUAD (mean RPKM: 37.04) and its adjacent normal tissues (mean RPKM: 28.49) as well as in LUSC (mean RPKM: 17.29) and its adjacent normal tissues (mean RPKM: 32.18), which implied that the role of *FHIT* gene played in NSCLC carcinogenesis need to be further confirmed (Figure 3).

Discussion

The *FHIT* gene loss was observed in 64% of non-small-cell lung carcinoma patients and is reported to be significantly associate with squamous cell carcinoma and poor tumor grade. 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 diagnostic ability using *FHIT* promoter methylation level as a biomarker in NSCLC, and a significant association was identified between *FHIT* methylation and NSCLC (OR = 3.43). 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.

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 plasma), proportion of males, and detection methods could not explain the heterogeneity. Subgroup analysis showed that *FHIT* methylation is significant relevant to NSCLC in Asians (OR = 3.50, 95% CI: 1.50 to 8.14) but not in white population (OR = 2.55, 95% CI: 0.86 to 7.57), indicating that aberrant methylation of *FHIT* can be a diagnostic biomarker for NSCLC in Asian population.

In the validation stage, all the results from three independent datasets showed no significance of differential methylation between NSCLC and normal tissues on account of the small mean methylation difference. It was found that in the dataset from TCGA dataset, none of the eight CpG sites which shared the same CpG island with the primers in the meta-analysis is significantly different methylated. And the result is further confirmed by other two datasets from the GEO database. Furthermore, we downloaded the RNA-Seq data from TCGA project and still no significant differential expression of *FHIT* gene was found both in LUAD and LUSC when compared with adjacent normal tissues. Besides, the expression level of the *FHIT* gene is relatively low in comparison with other functional genes in cancer.

It is easy to explain because all the independent datasets from TCGA and GEO were based on white population. The result about white population from datasets is consistent with the result from meta-analysis, so the relationship between *FHIT* methylation and NSCLC in white population is robust. Obviously we need more microassay and RNA-Seq data based on Asian people to distinguish whether the diagnostic role of *FHIT* is specific in the Asians and may lead a new way to elucidate the etiology of lung cancer.

Conclusion

The diagnostic role of *FHIT* gene in the lung cancer is relatively limited in the white population but may be useful in the Asians. However we need more solid data to confirm that.

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 October 2015. 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 contained FHIT gene promoter methylation data from tissue, blood or plasma, 3) the studies had to be case-control studies which included tissue-tissue, blood-blood or plasma-plasma in case and controls respectively, 4) OR can be calculated or extracted from the text. 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 and II 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 (Table 1).

Meta-analysis

Data were analyzed and visualized mainly using R Software (R version 3.1.0) including meta, metefor and mada packages[29][30]. 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 I^2 statistic with values over 50% and Chi-squared test with $P \leq 0.1$ indicating strong heterogeneity between the studies[31]. Tau-

squared (τ^2) was used to determine how much heterogeneity was explained by subgroup differences. The data was pooled using the DerSimonian and Laird random effects model ($I^2 > 50\%$, $P \leq 0.1$) or fixed effects model ($I^2 < 50\%$) according to heterogeneity statistic I^2 [32]. A two-sided $P \leq 0.05$ was set as the threshold of being significant without special annotation. With a lack of heterogeneity among included studies, the pooled odds ratio estimates were calculated using the fixed-effects model [33]. Otherwise, the random-effects model was used [34]. Random effects meta-regression was employed to determine how much of the heterogeneity (between-study variance) was 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 and II samples, mean or median age (> 59 or ≤ 59), single or multiple target detection, sample types (plasma or tissue), methylation detection methods (MSP, qMSP), study designs (diagnosis or non-diagnosis) and primer sets. Sensitivity analyses were performed to assess the contribution of single study to the final result with the abandonment of one article each time. Publication bias was analyzed by funnel plot with mixed-effects version of the Egger test [35]. If bias was suspected, the conventional meta-trim method was used to re-estimate the effect size.

TCGA and GEO datasets extraction and analysis

TCGA DNA methylation datasets which included 23 lung adenocarcinoma and 40 lung squamous cell carcinoma tissues as well as 63 paired adjacent tissues, were collected from TCGA project[<http://cancergenome.nih.gov/>] using Illumina HumanMethylation 450K Beadchip. And The GEO datasets including GSE39279, GSE52401 and GSE56044 were downloaded from Gene Expression Omnibus[<http://www.ncbi.nlm.nih.gov/geo/>], including a sum of 568 NSCLC tissues and 256 adjacent or normal lung tissues [36-38]. All of the above datasets are using Illumina HumanMethylation450 BeadChip for methylation measurement. The estimation of methylation for each CG probe was calculated between methylated (M) and unmethylated (U) alleles. Specifically:

$$\text{beta} = \frac{\max(M,0)}{\max(M,0)+\max(U,0)}$$

M and U represent the mean signal intensities for about 30 replicates on the array. The methylation signals of the CpG sites in the datasets previously mentioned were all defined according to the beta value. CpG site would be immediately omitted if there was one missing samples or more in the dataset. CpG sites of FHIT gene in TCGA dataset and GEO dataset were not completely the same due to the quality control previously mentioned.

P-value was calculated with Wilcoxon rank sum test. To correct for multiple testing, Benjamini and Hochberg procedure was conducted. For identification of differentially methylated CpG sites,

adjusted P-value ≤ 0.05 and absolute mean difference ≥ 0.1 was set as the criteria. Besides, logistic regression was also conducted to calculate the OR and p-value for every CpG site with Benjamini and Hochberg multiple comparison correction followed. Data was analyzed and visualized mainly with R software (R 3.1.0)[39][40].

RNA-Seq data extraction and analysis

RNA-Seq data was downloaded from TCGA Data Portal, which was available for 114 lung adenocarcinoma and 104 lung squamous cell carcinoma and 218 paired adjacent normal lung tissues. Level 3 RNA-Seq data was obtained and per million mapped reads (RPKM) was chosen as the measurement for gene expression. We assessed the significance of the differential gene expression by comparing the tumor tissues with paired adjacent normal tissues using Wilcoxon rank sum test and a Benjamini and Hochberg false discovery rate (FDR) correction[40]. For identification of differentially expression genes, adjusted p-value ≤ 0.05 and fold change ≥ 2.0 was set as the criteria. All the data analysis was conducted with open-source R software (version 3.1.0).

Acknowledgements

The computations involved in this study were supported by Fudan University High-End Computing Center.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

XG and WP, XC, JW contributed to the conception, design and final approval of the submitted version. XG, SG, ZL, AW, YT contributed to the meta-analysis and interpretation of data, XG, SG, WP, LT, SC contributed to TCGA NSCLC data analysis. All authors read and approved the final manuscript.

Grant Support

This study was partially supported by the grants from the 111 Project (B13016).

Reference

1. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2015*. CA: A Cancer Journal for Clinicians, 2015. **65**(1): p. 5-29.
2. Hankey, B.F., L.A. Ries, and B.K. Edwards, *The surveillance, epidemiology, and end results program: A national resource*. Cancer Epidemiology Biomarkers & Prevention, 1999. **8**(12): p. 1117-1121.
3. van Rens, M.T.M., 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. 374-379.
4. He, Y.H., 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-U110.
5. Flores, K., et al., *Genome-wide association between DNA methylation and alternative splicing in an invertebrate*. BMC Genomics, 2012. **13**.
6. Das, P.M., *DNA Methylation and Cancer*. Journal of Clinical Oncology, 2004. **22**(22): p. 4632-4642.
7. Barnes, L.D., et al., *Fhit, a putative tumor suppressor in humans, is a dinucleoside 5',5'''-P-1,P-3-triphosphate hydrolase*. Biochemistry, 1996. **35**(36): p. 11529-11535.
8. Nishimura, A., *The timing of cell division: Ap4A as a signal*. Trends in Biochemical Sciences, 1998. **23**(5): p. 157-159.
9. Huebner, K., et al., *The role of the FHIT/FRA3B locus in cancer*. Annual Review of Genetics, 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*. Applied Immunohistochemistry & Molecular Morphology, 2014. **22**(4): p. 275-283.
11. Wang, H.L., et al., *Abnormal FHIT protein expression may be correlated with poor prognosis in gastric cancer: a meta-analysis*. Tumor Biology, 2014. **35**(7): p. 6815-6821.
12. Guo, X.Q., et al., *DNA methylation and loss of protein expression in esophageal squamous cell carcinogenesis of high-risk area*. Journal of Experimental & Clinical Cancer Research, 2007. **26**(4): p. 587-594.
13. Kapitanovic, S., et al., *Reduced FHIT expression is associated with tumor progression in*

sporadic colon adenocarcinoma. Experimental and Molecular Pathology, 2014. **96**(1): p. 92-97.

14. Edgar, R., M. Domrachev, and A.E. Lash, *Gene Expression Omnibus: NCBI gene expression and hybridization array data repository*. Nucleic Acids Research, 2002. **30**(1): p. 207-210.
15. Geng, X., W.L. Pu, and S.C. Guo, *Aberrant methylation of CDH13 can be a diagnostic biomarker for lung adenocarcinoma*. submitted, 2015.
16. Guo, S., et al., *Quantitative assessment of the diagnostic role of APC promoter methylation in non-small cell lung cancer*. Clin Epigenetics, 2014. **6**(1): p. 5.
17. de Fraipont, F., et al., *Promoter methylation of genes in bronchial lavages: A marker for early diagnosis of primary and relapsing non-small cell lung cancer?* Lung Cancer, 2005. **50**(2): p. 199-209.
18. Fischer, J.R., et al., *Prognostic significance of RASSF1A promoter methylation on survival of non-small cell lung cancer patients treated with gemcitabine*. Lung Cancer, 2007. **56**(1): p. 115-123.
19. Haroun, R.A.H., et al., *Assessment of the Prognostic Value of Methylation Status and Expression Levels of FHIT, GSTP1 and p16 in Non-Small Cell Lung Cancer in Egyptian Patients*. Asian Pacific Journal of Cancer Prevention, 2014. **15**(10): p. 4281-4287.
20. Hsu, H.S., et al., *Characterization of a multiple epigenetic marker panel for lung cancer detection and risk assessment in plasma*. Cancer, 2007. **110**(9): p. 2019-26.
21. Kim, D.S., et al., *Aberrant DNA methylation profiles of non-small cell lung cancers in a Korean population*. Lung Cancer, 2007. **58**(1): p. 1-6.
22. Kim, H., et al., *Tumor-specific methylation in bronchial lavage for the early detection of non-small-cell lung cancer*. Journal of Clinical Oncology, 2004. **22**(12): p. 2363-2370.
23. Li, W., et al., *Association of 5'-CpG island hypermethylation of the FHIT gene with lung cancer in Southern-central chinese population*. Cancer Biology & Therapy, 2010. **10**(10): p. 997-1000.
24. Li, W., J. Deng, and J.X. Tang, *Combined Effects Methylation of FHIT, RASSF1A and RAR beta Genes on Non-Small Cell Lung Cancer in the Chinese Population*. Asian Pacific Journal of Cancer Prevention, 2014. **15**(13): p. 5233-5237.
25. Verri, C., et al., *Fragile Histidine Triad Gene Inactivation in Lung Cancer The European Early Lung Cancer Project*. American Journal of Respiratory and Critical Care Medicine, 2009. **179**(5): p. 396-401.

26. Yanagawa, N., et al., *Promoter hypermethylation of RASSF1A and RUNX3 genes as an independent prognostic prediction marker in surgically resected non-small cell lung cancers*. Lung Cancer, 2007. **58**(1): p. 131-138.
27. Zhang, Y.W., et al., *Methylation of multiple genes as a candidate biomarker in non-small cell lung cancer*. Cancer Letters, 2011. **303**(1): p. 21-28.
28. Zochbauer-Muller, S., et al., *5 ' CpG island methylation of the FHIT gene is correlated with loss of gene expression in lung and breast cancer*. Cancer Research, 2001. **61**(9): p. 3581-3585.
29. Schwarzer, G., *meta: General Package for Meta-Analysis. R package version 4.2-0*. 2015, Vienna, Austria: R Foundation.
30. Team R C. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2012[J]. 2014
31. Guo, S.C., et al., *Quantitative assessment of the diagnostic role of APC promoter methylation in non-small cell lung cancer*. Clinical Epigenetics, 2014. **6**.
32. DerSimonian, R. and N. Laird, *Meta-analysis in clinical trials*. Control Clin Trials, 1986. **7**(3): p. 177-88.
33. Mantel, N. and W. Haenszel, *Statistical aspects of the analysis of data from retrospective studies of disease*. J Natl Cancer Inst, 1959. **22**(4): p. 719-48.
34. Huizenga, H.M., I. Visser, and C.V. Dolan, *Testing overall and moderator effects in random effects meta-regression*. Br J Math Stat Psychol, 2011. **64**(Pt 1): p. 1-19.
35. Egger, M., et al., *Bias in meta-analysis detected by a simple, graphical test*. BMJ, 1997. **315**(7109): p. 629-34.
36. Sandoval, J., et al., *A Prognostic DNA Methylation Signature for Stage I Non-Small-Cell Lung Cancer*. Journal Of Clinical Oncology, 2013. **31**(32): p. 4140-+.
37. Shi, J.X., et al., *Characterizing the genetic basis of methylome diversity in histologically normal human lung tissue*. Nature Communications, 2014. **5**.
38. Karlsson, A., et al., *Genome-wide DNA Methylation Analysis of Lung Carcinoma Reveals One Neuroendocrine and Four Adenocarcinoma Epitypes Associated with Patient Outcome*. Clinical Cancer Research, 2014. **20**(23): p. 6127-6140.
39. Team R C. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2012[J]. 2014
40. Benjamini, Y. and Y. Hochberg, *Controlling the False Discovery Rate - a Practical and*

Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Methodological, 1995. **57**(1): p. 289-300.

Table 1. Characteristics of eligible studies considered in the report

Author	Sample Type	Age ^a	Stage I%	Stage (I+II)%	Gender Ratio	Patients (M/T)	Control (M/T)	Methods	Aim	Multipletarget	Control design	Ad2Sc	primer set
Haroun et al	Tissue	53	0.18	0.57	0.71	15/28	1/28	qMSP	Non-Diagnosis	Multi	hom	1.777778	1
Li et al (2014, China)	Serum	53.15	0.27	0.39	NA	19/56	0/56	MSP	Non-Diagnosis	Multi	heter	0.586207	1
Li et al (2010, China)	Serum	55.03	NA	NA	0.71	42/123	0/105	MSP	Non-Diagnosis	Single	heter	0.602941	1
Zhang et al	Tissue	59	0.32	0.74	0.74	1/40	1/40	MSP	Diagnosis	Multi	hom	0.833333	2
Fischer et al	Serum	60.9	0	0	0.65	43/92	0/7	MSP	Non-Diagnosis	Multi	heter	1.714286	1
Zochbauer et al	Tissue	61	0.57	0.77	0.71	40/107	9/104	MSP	Non-Diagnosis	Single	hom	1.046512	1
Kim.D et al	Tissue	63	0.57	0.75	0.81	34/99	17/99	MSP	Non-Diagnosis	Multi	hom	0.622951	3
Verri et al	Tissue	63.9	0.65	NA	0.84	84/229	68/208	MSP	Non-Diagnosis	Single	hom	1.109091	1
Yanagawa et al	Tissue	68.1	0.67	0.74	0.71	34/101	7/101	MSP	Non-Diagnosis	Multi	hom	1.589744	1
Hsu et al ^b	Tissue	69	NA	0.65	0.71	22/57	9/63	qMSP	Diagnosis	Multi	hom	0.759259	2
Fraipont et al	Serum	NA	NA	NA	NA	6/16	18/56	MSP	Diagnosis	Multi	heter	NA	1
Hsu et al ^b	Serum	NA	NA	0.65	0.71	18/57	7/35	qMSP	Diagnosis	Multi	heter	0.759259	2
Kim.H et al	Serum	NA	0.59	1	0.67	19/85	36/127	MSP	Diagnosis	Multi	heter	0.72093	1

^amean or median age from articles; ^bwith two records since there are Tissue and serum data simultaneously in this article. M and T means methylation positive and total, respectively.

**Table 2. Meta-regression analysis for the main potential interference factors with
random-effects model**

Subgroup	Coefficient(95% CI)	P-value	τ^2	QE	QE. P-value
sampletype	0.18(-1.14, 1.49)	0.79377	0.9	52.84	1.92E-07
age	-0.15(-0.3, 0)	0.052152	0.94	40.04	3.16E-06
stage1	-3.92(-7.89, 0.05)	0.052921	0.76	35.74	8.13E-06
stage2	-3.98(-6.83, -1.14)	0.00609	0.32	17.07	0.029374
genderatio	-5.38(-18.19, 7.44)	0.411213	1.03	46.81	4.26E-07
methods	0.33(-1.14, 1.81)	0.656529	0.85	53.51	1.45E-07
aim	1.45(0.15, 2.74)	0.028296	0.88	51.44	3.44E-07
Multipltarget	0.36(-1.23, 1.95)	0.655878	1.08	55.52	6.22E-08
controldesign	0.18(-1.14, 1.49)	0.79377	0.9	52.84	1.92E-07
Ad2Sc	1.04(-0.74, 2.82)	0.251308	0.92	51.41	1.47E-07

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

Table 3. Differential FHIT methylation, odds ratio between adenocarcinoma, squamous cell carcinoma and their counterparts from TCGA dataset

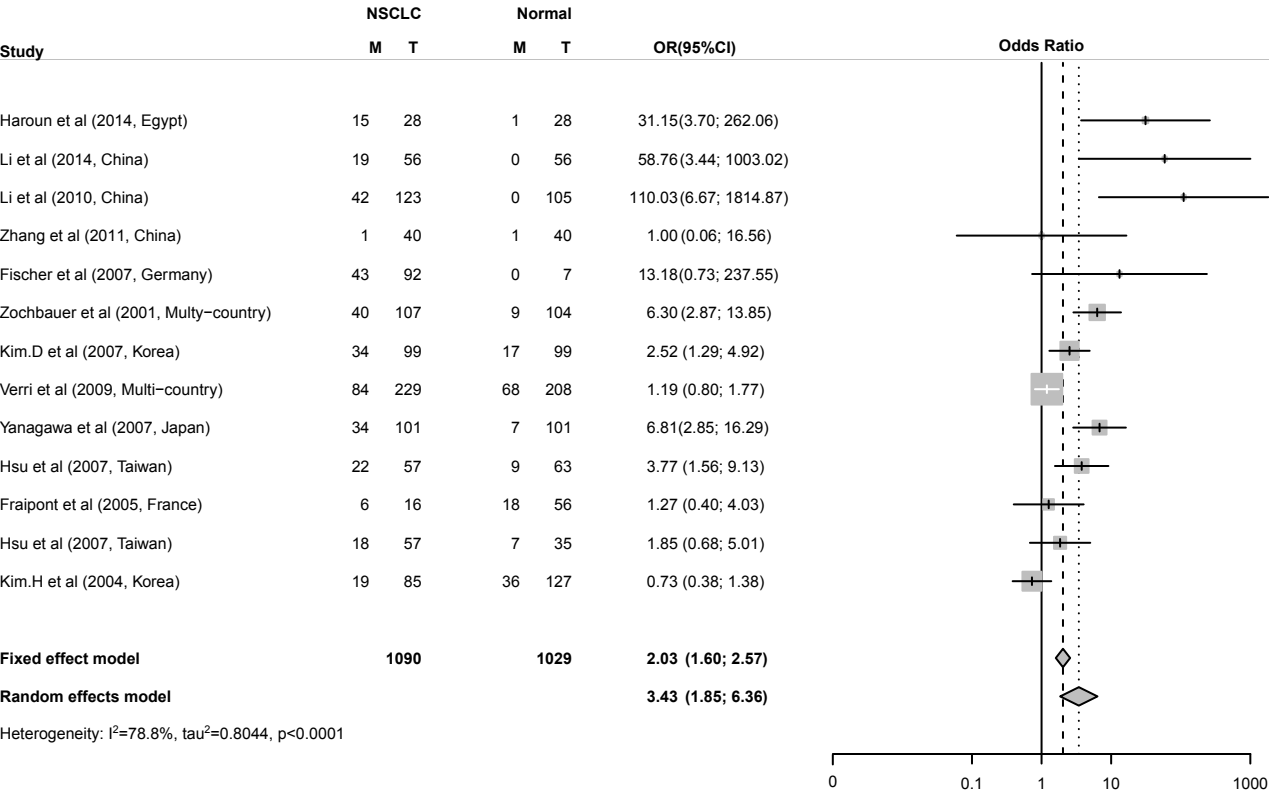
Type	CpG site	Position	McaM	McoM	$\Delta\beta$	P-value ^a	P-value ^b	OR ^b	95%CI ^b
LUAD	cg22215728	61236652	0.12	0.16	0.04	0.00081	0.0097	4.33	1.93-12.55
	cg15931943	61236909	0.10	0.12	0.02	0.00718	0.18873	1.80	0.91-4.62
	cg02854288	61236911	0.11	0.13	0.02	0.00168	0.04887	2.81	1.29-7.97
	cg19049316	61237063	0.03	0.03	0.00	0.22515	0.14417	1.73	0.93-3.58
	cg26322434	61237156	0.03	0.04	0.01	0.29934	0.17031	1.63	0.89-3.28
	cg24796403	61237172	0.04	0.04	0.00	0.66265	0.40238	1.35	0.73-2.89
	cg16986494	61237206	0.04	0.05	0.01	0.01932	0.04887	3.12	1.33-9.44
	cg12030002	61237226	0.04	0.05	0.01	0.2052	0.24634	1.49	0.82-2.9
LUSC	cg22215728	61236652	0.10	0.14	0.04	1.00E-05	0.00066	3.92	2.07-8.44
	cg15931943	61236909	0.09	0.10	0.01	1.00E-05	0.01163	2.51	1.35-5.13
	cg02854288	61236911	0.08	0.10	0.02	1.00E-05	0.00485	2.76	1.5-5.58
	cg19049316	61237063	0.03	0.03	0.00	0.52339	0.31038	0.66	0.26-1.16
	cg26322434	61237156	0.03	0.02	0.01	0.51658	0.29719	0.77	0.45-1.21
	cg24796403	61237172	0.04	0.03	0.01	0.52339	0.29486	0.31	0.03-1.13
	cg16986494	61237206	0.04	0.03	0.01	0.02458	0.54451	0.84	0.26-1.37
	cg12030002	61237226	0.04	0.03	0.01	0.11177	0.12536	0.10	0.005-0.99

McaM and McoM represent the mean of case methylation (Beta) and mean of control methylation (Beta). Methylation levels are calculated with formula: Beta = (M/M + U).

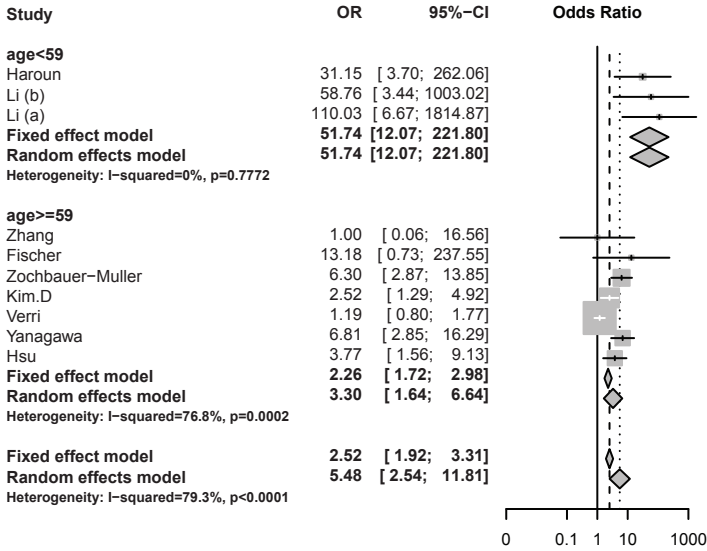
Position represents the chromosome position of each CpG site according to GRCh37/hg19.

P-values^a are calculated from Wilcoxon rank sum test after false discovery rate(FDR adjustment).

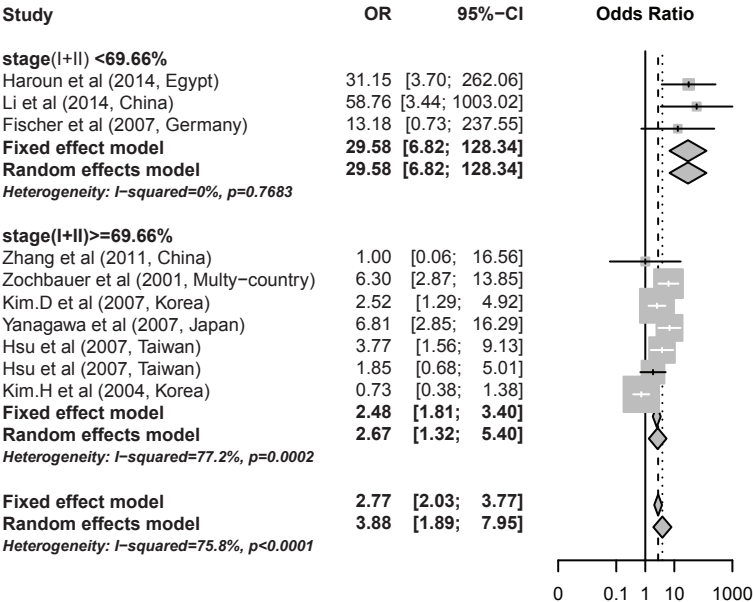
P-value^b and OR^b and 95%CI^b are from logistic regression analysis with P-value^b were also after false discovery rate (FDR adjustment).



A



B



C

