Quantitative assessment of the diagnostic role of FHIT promoter methylation in non-small cell lung cancer

By Xiaofeng Chen

WORD COUNT 2904
TIME SUBMITTED 11-NOV-2016 12:48AM
PAPER ID 25825442

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 lacking of symptoms in its early stages, however, 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 events [5]. It has been observed and confirmed that DNA methylation change is wide-spread 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 low requirements for sample quality [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 cancer 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]. FHIT has been recently seen as a genome caretaker which is of great importance for genome stability. Multiple studies have found the reduction of FHIT expression in precancerous lesions, indicating its potential suppressing role in carcinogenesis [14-19]. The FHIT -/- mice were more prone to develop carcinogen-induced tumors as well as the spontaneous tumors than wild type mice [20, 21]. And FHIT viral gene therapy was found to be able to prevent and reverse carcinogen-induced tumors in a gastric cancer mouse model [22]. Moreover, recent studies have found that FHIT can also function as the tumor suppressor by inhibiting EMT [23, 24]. 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 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 [25]. Several studies have showed the improved robustness of combining data from papers and databases [26, 27]. 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. More detailed information about the inclusion or exclusion criteria was shown in Figure 1. Finally, 12 studies [28-39] with data on the relationship between FHIT gene promoter methylation and NSCLC were pooled for analysis (Figure 2 and Table 1). 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 2).

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 (Ad/Sc) 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 3A) 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 3B). Interestingly, difference was found between Asian (3.50, 95% CI: 1.50 to 8.14, P = 0.005) and Caucasian population (2.55, 95% CI: 0.86 to 7.57, P = 0.09) subgroup (Figure 3C), and the differential methylation in Caucasian population is not significant, indicating that diagnostic ability of FHIT methylation might be limited in Caucasian population. Both tissue and plasma groups showed significant association between FHIT methylation and NSCLC (OR = 3.68 and 3.89, respectively) (Figure 3D) which suggested that FHIT methylation can be taken as a potential biomarker for NSCLC diagnosis using either tissue or plasma samples. FHIT has been reported to be related with smoking history but not with cancer, thus we conducted the subgroups of the percentage of smoking samples. And we found no significant difference between the smoker%<68% and smoker% >=68% subgroups (Figure S7). In addition, 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 (I2 = 78.8%, Q2 = 61.05, P < 0.0001) (Figure 2), 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 was 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 and Figure 4).

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 GSE39279 dataset, a total of 244 normal lung tissues were included, and both of the dataset used the Illumina HumanMethylation450 Bead Chip 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 and Figure S5). 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 S6).

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 4).

Discussion

The FHIT gene loss was observed in 64% of non-small-cell lung cancer 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.

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. We should be noticed that all

the independent datasets from TCGA and GEO were based on Caucasian population. The result about Caucasian population from datasets is consistent with the result from meta-analysis, so the relationship between FHIT methylation and NSCLC in Caucasian population is robust. Besides, we need more micro-assay and RNA-Seq data based on Asian population to distinguish whether the diagnostic role of FHIT is specific in the Asians.

In our meta-analysis, we found high rate of heterogeneity between the studies ($p < 0.0001$). Thus we did further research to explore the influential confounding factors. In our subgroup analysis, we found that ages, stages as well as the aims are the sources of heterogeneity (Table 2). However, significant odds ratios between FHIT promoter methylation and NSCLC were still retained in most of the subgroups, which is in accordance with the overall meta-analysis results (Table S2). 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 Caucasian 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. Wu et al have previously conducted the meta-analysis to test the association between FHIT methylation and NSCLC, and they found differential methylation of FHIT promoter in both Caucasian and Asian populations, which was different with our findings [40]. In addition, the much more significant difference of FHIT promoter methylation between NSCLC and normal controls was observed both in our meta-analysis and in Wu's research. The above consistencies and inconsistencies between the two studies implied the need to test the association between FHIT methylation and NSCLC with larger sample sizes and more advanced technology.

There are several limitations in our review and meta-analysis. Firstly, there is a relatively strong heterogeneity between the included studies, which may decrease the statistical power of the meta-analysis results. Secondly, the publication bias may still present, though we have conducted the trim and fill analysis and drew the same conclusions as the meta-analysis. Thirdly, we have searched the papers only written in English, while many other related papers written in other languages might be ignored. As a result, the English bias may also exist. Due to the previous limitations in our work,

we strongly recommend to use more advanced methylation detection methods, like WGBS (whole genome bisulfite sequencing) and RRBS (restricted region bisulfite sequencing), to explore the association between FHIT promoter methylation and NSCLC with larger sample sizes.

Conclusion

The diagnostic role of FHIT gene in the lung cancer is relatively limited in the Caucasian population but may be useful in the Asians. However, more datasets and studies with large sample sizes are needed for further confirmation.

Quantitative assessment of the diagnostic role of FHIT promoter methylation in non-small cell lung cancer

ORIGINALITY REPORT

34%

SIMILARITY INDEX

PRIMARY SOURCES

1	fe-for-worst-offenders- Internet	746 words — 25%
2	www.clinicalepigeneticsjournal.com Internet	77 words — 3%
3	Jeffrey Aufman. "Carcinogenesis of Lung Cancer", Mechanisms of Oncogenesis, 2010 Crossref	27 words — 1%
4	en.wikipedia.org Internet	20 words — 1%
5	Clinical and Biological Basis of Lung Cancer Prevention, 1998. Crossref	20 words — 1%
6	www.iaeng.org Internet	19 words — 1%
7	www.biomedcentral.com Internet	18 words — 1%
8	Yang, Jingyu, Heng Niu, Yingze Huang, and Kunxian Yang. "A Systematic Analysis of the Relationship of CDH13 Promoter Methylation and Breast Cancer Risk and Prognosis", PLoS ONE, 2016. Crossref	15 words — 1%
9	Methods in Molecular Biology, 2015. Crossref	15 words — 1%

10	Zhong, Yun-Hua, Hao Peng, Hong-Zhong Cheng, and Ping Wang. "Quantitative Assessment of the Diagnostic Role of CDH13 Promoter Methylation in Lung Cancer", Asian Pacific Journal of Cancer Prevention, 2015. Crossref	13 words — < 1%
11	cancerres.aacrjournals.org Internet	11 words — < 1%
12	Sun, J., J. Liu, X. Pan, D. Quimby, N. Zanesi, T. Druck, G. P. Pfeifer, C. M. Croce, L. Y. Fong, and K. Huebner. "Effect of zinc supplementation on N-nitrosomethylbenzylamine-induced forestomach tumor development and progression in tumor suppressor-deficient mouse strains", Carcinogenesis, 2011. Crossref	9 words — < 1%
13	Kapral, M.. "Transforming growth factor @b isoforms (TGF-@b1, TGF-@b2, TGF-@b3) messenger RNA expression in laryngeal cancer", American Journal of Otolaryngology--Head and Neck Medicine and Surgery, 200807 Crossref	7 words — < 1%
14	Molecular Testing in Cancer, 2014. Crossref	6 words — < 1%

EXCLUDE QUOTES OFF
EXCLUDE MATCHES OFF
EXCLUDE BIBLIOGRAPHY OFF