Exploration of association between EPHX1 and chronic obstructive pulmonary disease on the basis of combined data mining

L. An¹*, H. Xia²,³*, P. Zhou²,³ and L. Hua²,³

¹Beijing Institute of Respiratory Medicine, Beijing Chao-Yang Hospital, Capital Medical University, Beijing, China
²School of Biomedical Engineering, Capital Medical University, Beijing, China
³Beijing Key Laboratory of Fundamental Research on Biomechanics in Clinical Application, Capital Medical University, Beijing, China

*These authors contributed equally to this study.

Corresponding author: L. Hua
E-mail: hualin7750@139.com

Received March 23, 2016
Accepted April 11, 2016
Published April 26, 2016
DOI http://dx.doi.org/10.4238/gmr.15028639

ABSTRACT. Chronic obstructive pulmonary disease (COPD) is an important respiratory disease with high mortality. Although smoking is the major environmental risk factor for the development of COPD, only 10% of heavy smokers develop symptomatic disease, suggesting association between genetic susceptibilities and environmental influences. In recent years, as one of the most widely studied genes including tests for associations between a genetic variant and COPD, epoxide hydrolase 1 (EPHX1) was found to be involved in the metabolism of tobacco smoke, an important risk factor of COPD. However, genetic associations with COPD identified in studies on EPHX1 are controversial. To address this issue, except for performing the meta-analysis, which specially added our current study on two polymorphisms (T337C and A416G) of EPHX1, we performed
combined data mining based on functional prediction algorithms of nonsynonymous single-nucleotide polymorphisms and gene-based variable threshold testing. Genetic variations in EPHX1 did not affect COPD in Caucasian and Eastern Asian population, which is supported by recent evidence. We found no association between EPHX1 and COPD; however, a minor effect of EPHX1 on COPD risk was not completely excluded; further replication studies with large samples are needed to confirm our findings.

**Key words:** Chronic obstructive pulmonary disease; Meta-analysis; Epoxide hydrolase 1; Nonsynonymous single-nucleotide polymorphisms; Variable threshold

**INTRODUCTION**

Chronic obstructive pulmonary disease (COPD) is a progressive disease, defined as airflow limitation that is not fully reversible (Sotiriou and Makris, 2013). Although cigarette smoking is the major environmental risk factor for COPD, previous studies have revealed that smokers show considerable variation in developing symptomatic COPD (Sethi and Rochester, 2000). Thus, COPD is possibly caused due to links between genetic susceptibilities and environmental factors. Epoxide hydrolase 1 (EPHX1) is an enzyme involved in the detoxification of smoking-induced reactive substances (Chappell et al., 2008). Genetic variants of this enzyme have been widely studied for their role in the genetic susceptibility to COPD. EPHX1 is strongly expressed in the lung, but down-regulated in COPD (Tomaki et al., 2007; Lee et al., 2011). However, genetic associations with COPD identified in EPHX1 studies have been inconsistent. Some studies suggested an association between EPHX1 and COPD, including previous and recent meta-analysis results (Brøgger et al., 2006; Li et al., 2013). However, other studies found no association between EPHX1 and COPD. For example, Chappell et al. (2008) found that several single-nucleotide polymorphisms (SNPs) in EPHX1 were not associated with the presence or severity of COPD. A subsequent study by Siedlinski et al. (2009) confirmed that there was no effect of EPHX1 polymorphisms on the level or forced expiratory volume in 1 s in the general population. These conflicting results make it difficult to conduct further analyses (Nakamura, 2011). Notably, none of the investigated SNPs involved in EPHX1 were found to be associated with COPD in previous genome-wide association studies (GWAS) (Pillai et al., 2009; Todd et al., 2011; Cho et al., 2012). The stringent criteria for selecting SNPs may have resulted in these observations. Although GWAS is limited in finding novel biomarkers, these studies have greatly advanced pulmonary medicine. Therefore, these GWAS results regarding the roles of EPHX1 in COPD require further investigation using larger data sets and deeper data mining.

Nonsynonymous SNPs (nsSNPs) can alter the amino acid sequences of proteins and further affect the function of proteins, accounting for the susceptibility to complex disease (Stenson et al., 2008). Thus, to determine whether EPHX1 is a candidate gene for COPD, except for performing the meta-analysis which specially added our current study on two polymorphisms (T337C and A416G) of EPHX1, further data mining based on functional prediction algorithms of nsSNPs and gene-based variable threshold (VT) tests (Price et al., 2010) was conducted in this study. The VT method allows for the incorporation of
Determine possible association between EPHX1 and COPD

computational predictions of the functional effects of nsSNPs into the association test, and therefore is more powerful than traditional association test methods because it combines both functional and non-functional alleles (Adzhubei et al., 2010). Given the large differences in genetic backgrounds of Asians and Caucasians, we performed the VT tests using three datasets: two datasets included Caucasian text data and Eastern Asian text data extracted from the meta-analysis, respectively, and included our Chinese Han data.

MATERIAL AND METHODS

Datasets

Two SNPs of \textit{EPHX1} (T337C (rs1051740) and A416G (rs2234922)) are frequently chosen as COPD biomarkers because they tag the entire coding region and are functional SNPs; we therefore also investigated these two SNPs in the current study. The three datasets used in this paper are described below.

Caucasian text data and Eastern Asian text data extracted from the meta-analysis

We obtained literature data using Gene Prospector (http://www.hugenavigator.net/HuGENavigator/home.do), an online gateway for searching human genes and their relationships with disease. We performed a systematic literature search of the, Medline, and Embase databases using the following key terms: “EPHX1”, “microsomal epoxide hydrolase”, “polymorphism”, “COPD”, “chronic obstructive pulmonary disease”, “genetic variant”, “EPHX1 His139Arg”, “EPHX1 Tyr113His”, and their various combinations. The language was restricted to English and the study populations were limited to Caucasian and Eastern Asian (Japanese, Korean, and Chinese) populations. The selected studies were published after January 2000 and up to May 2013.

Studies eligible for meta-analysis met the following inclusion criteria: i) an unrelated case-control design for the association studies, ii) cases were COPD with the diagnosis based on clinical criteria, iii) sufficient data were provided to calculate odd ratio (OR) values for two SNPs (\textit{EPHX1} T337C and \textit{EPHX1} A416G).

For study selection and quality assessment, we excluded studies that contained overlapping data. The quality of included observational studies was assessed by two authors and judged according to the Newcastle-Ottawa scale. The following information was extracted if it met our criteria: first author, year of publication, ethnicity, sample size of the cohorts, \textit{EPHX1} gene T337C and A416G allele, and genotype distributions among case and control groups.

Our current study cohort: Chinese Han population

A total of 310 unrelated COPD patients aged 40-75 years and 203 normal subjects with the same age range were recruited from the respiratory outpatient clinics and the medical examination center at 12 hospitals in Beijing from October of 2007 to March 2009. Written informed consent was obtained from all participating subjects, and the study protocol was approved by the research ethics boards of all participating hospitals. The entry criteria were the same as those in our previous study (An et al., 2012; Hua et al., 2012).
A conventional phenol-chloroform method was used to isolate genomic DNA from whole blood leukocytes. Illumina VeraCode technology on a BeadXpress platform (Illumina, San Diego, CA, USA) was used to perform SNP genotyping. Forty-four tagging SNPs (minor allele frequency >0.05) were genotyped to identify common variants of the four genes (EPHX1, GSTP1, SERPINE2, and TGFB1) under pair-wise mode with an r² threshold of 0.8 (An et al., 2012; Hua et al., 2012). In this analysis, we focused on two SNPs (EPHX1 T337C and EPHX1 A416G) to explore whether the genetic variations in EPHX1 are associated with COPD in Chinese Han population.

Data mining method

Meta-analysis

In this study, we conducted a meta-analysis including different case-control studies; the samples in these studies were not necessarily typed using the same genotyping product or imputed to the same reference panel. We therefore assumed that the filter studies had been subjected to quality control, while those including poorly genotyped samples or imputed SNPs were excluded. The Cochrane Q test was used to test for heterogeneity and I² was used to measure inconsistency. The I² results range from 0 to 100% (I² = 0-25%, no heterogeneity; I² = 25-50%, moderate heterogeneity; I² = 50-75%, large heterogeneity; I² = 75-100%, extreme heterogeneity) (Zhang et al., 2011). The minor allele (C of T337C and G of A416G) of EPHX1 may increase the susceptibility to COPD; we therefore investigated the association between these two genetic variants and COPD risk under an allelic model (C vs T and G vs A) for Caucasian and Eastern Asian populations, respectively. The pooled OR was calculated using a fixed-effect model or a random-effect model according to the heterogeneity among studies. The statistical significance of the pooled ORs was determined by the Z test, and P < 0.05 indicated a statistically significant difference. We applied RevMan 5.1 software (http://ims.cochrane.org/revman) and R software (http://www.r-project.org) for the meta-analysis. For the meta-analysis of the Eastern Asian population, the genotype data from our current study were included.

VT tests based on functional prediction algorithms of nsSNPs

To further explore the association between EPHX1 and COPD, we performed combined data mining based on functional prediction algorithms of nsSNPs and gene-based VT tests. The following are the detailed description of the methods.

1) Prediction of the potential functional effect scores of two SNPs of EPHX1

Here, we applied three programs (VarionWatch (Hua et al., 2012), SIFT (Kumar et al., 2009), and PolyPhen-2 (Adzhubei et al., 2010)) to predict the functional effect scores of EPHX1 T337C and EPHX1 A416G. VarionWatch incorporates 6 databases and can use different criteria to predict the function of the variant. This prediction is divided into 5 risk levels, including very low, low, medium, high, and very high. We quantified these 5 risk levels as 0.2, 0.4, 0.6, 0.8, and 1, respectively. SIFT prediction is based on sequence homology and the physical properties of amino acids. An nsSNP can be predicted by SIFT as ‘damaging’ (0.00-0.05) or ‘tolerated’ (0.05-1.00). PolyPhen-2 predicts functional effects of nsSNPs using straightforward physical and comparative considerations. According to the naïve Bayes
Determine possible association between EPHX1 and COPD

posterior probability calculated by PolyPhen-2, an nsSNP can be classified into ‘benign’ [0-0.2], ‘possibly damaging’ (0.2-0.85), or ‘probably damaging’ [0.85-1] groups. These three programs are described in detail in our previous study (Hua et al., 2012).

2) VT tests based on three functional effect scores

In the VT test (Price et al., 2010), rare alleles are grouped together and a threshold quantified by a z-score is used to maximize the difference between distributions of disease status for individuals with and without rare alleles. For each gene, given some unknown threshold T, variants with a minor allele frequency below T are substantially more likely to be functional than those with minor allele frequency above T. We assumed that gene G includes m SNPs for a given threshold T, with z-score defined as:

\[
z(T) = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} S_i \xi_i^T C_{ij}(\pi_j - \bar{\pi})}{\sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} (S_i \xi_i^T C_{ij})^2}}
\]  
(Equation 1)

Where,

\[
\xi_i^T = \begin{cases} 
1, & \text{MAF } T_i \text{ < } T \\
0, & \text{other} 
\end{cases}
\]  
(Equation 2)

\(C_{ij}\) is the reference allele count of SNP i in subject j and \(\pi_j\) is the phenotype of subject j, which equals 0 for controls or 1 for cases. \(\bar{\pi}\) is the mean value of \(\pi_j\) across subjects j. \(S_i\) is the functional effect score of SNP i; a larger value indicates a higher probability of a damaging effect. In this study, we incorporated three functional effect scores of nsSNPs into the gene-based VT test. An opposite trend exists between SIFT and PolyPhen-2; to maintain the positive correlation between these tests, we used “1-SIFT” to express the functional scores of nsSNPs. \(S_i\) was expressed as follows:

1) \(S_i = \text{VarioWatch} \)
2) \(S_i = 1 - \text{SIFT} \)
3) \(S_i = \text{PolyPhen} - 2 \)

For factors missing functional effect scores that could not be calculated using the SIFT and PolyPhen-2 programs, we imputed these values with the corresponding median scores: 0.9 for 1-SIFT and 0.2 for PolyPhen-2. Furthermore, the maximum z-score was defined as VT-z = max Z(T), where VT-z indicates the risk function score of gene G. In order to control type error and estimate the statistical significance of VT-z, we permuted the sample labels 1000 times and obtained an exact disease association evaluated by the VT-p value for each gene. The VT tests were performed using three datasets: two datasets included Caucasian text data and Eastern Asian text data extracted from the meta-analysis, while the third dataset included our Chinese Han population.

RESULTS

Meta-analysis

Literature search

As a result of literature search, 19 studies (include our current study and another
18 studies) investigating the relationship of \( EPHX1 \) T337C and A416G with COPD met our selection criteria. These studies included 9 Caucasian populations (Rodriguez et al., 2002; Korytina et al., 2003; Hersh et al., 2005; Park et al., 2005; Brøgger et al., 2006; Chappell et al., 2008; Židzik et al., 2008; Penyige et al., 2010; Lee et al., 2011) and 10 Eastern Asian populations (Takeyabu et al., 2000; Yim et al., 2000; Yoshikawa et al., 2000; Zhang et al., 2002; Budhi et al., 2003; Park et al., 2003; Cheng et al., 2004; Xiao et al., 2004; Fu et al., 2007; Hua et al., 2012) (Table 1). The total number of samples in the Caucasian and Eastern Asian populations were 47,048 (6471 cases vs 40,577 controls) and 2586 (1228 cases vs 1358 controls), respectively.

**Meta-analysis of two SNPs of EPHX1 for Caucasian population and Eastern Asian population**

For the Caucasian population, the meta-analysis under the allelic model revealed no association between \( EPHX1 \) T337C and COPD risk (pooled OR = 1.12, 95%CI = 1.00-1.25, \( P = 0.05 \)). A similar result was also observed for \( EPHX1 \) A416G (pooled OR = 1.01, 95%CI = 0.96-1.06, \( P = 0.65 \)) (Figure 1). For the Eastern Asian population, the meta-analysis showed no association between \( EPHX1 \) T337C and COPD risk (pooled OR = 1.04, 95%CI = 0.86-1.25, \( P = 0.69 \)). Similarly, no significant association was observed for \( EPHX1 \) A416G (pooled OR = 0.92, 95%CI = 0.79-1.08, \( P = 0.31 \)) (Figure 2). In summary, our meta-analysis revealed no statistically significant evidence for an association between two SNPs of \( EPHX1 \) and COPD risk under the allelic model for the Caucasian population and Eastern Asian population.

**Table 1. Characteristics of the individual studies of \( EPHX1 \) T337C and A416G polymorphisms included in this meta-analysis.**

<table>
<thead>
<tr>
<th>First author</th>
<th>Year</th>
<th>Country</th>
<th>Case</th>
<th>Control</th>
<th>First author</th>
<th>Year</th>
<th>Country</th>
<th>Case</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rodriguez F</td>
<td>2002</td>
<td>Spain</td>
<td>79</td>
<td>146</td>
<td>Takeyabu K</td>
<td>2000</td>
<td>Japan</td>
<td>79</td>
<td>58</td>
</tr>
<tr>
<td>Korytina GF</td>
<td>2003</td>
<td>Russia</td>
<td>88</td>
<td>162</td>
<td>Yim JJ</td>
<td>2000</td>
<td>Korea</td>
<td>83</td>
<td>76</td>
</tr>
<tr>
<td>Hersh CP</td>
<td>2005</td>
<td>USA</td>
<td>304</td>
<td>441</td>
<td>Yoshikawa M</td>
<td>2000</td>
<td>Japan</td>
<td>40</td>
<td>140</td>
</tr>
<tr>
<td>Park FY</td>
<td>2005</td>
<td>USA</td>
<td>131</td>
<td>262</td>
<td>Zhang RH</td>
<td>2002</td>
<td>China</td>
<td>55</td>
<td>52</td>
</tr>
<tr>
<td>Brøgger J</td>
<td>2006</td>
<td>Norway</td>
<td>239</td>
<td>241</td>
<td>Budhi A</td>
<td>2003</td>
<td>Japan</td>
<td>63</td>
<td>172</td>
</tr>
<tr>
<td>Chappell S</td>
<td>2008</td>
<td>European countries</td>
<td>1,017</td>
<td>912</td>
<td>Park SS</td>
<td>2003</td>
<td>Korea</td>
<td>58</td>
<td>79</td>
</tr>
<tr>
<td>Penyige A</td>
<td>2010</td>
<td>Hungary</td>
<td>269</td>
<td>289</td>
<td>Xiao D</td>
<td>2004</td>
<td>China</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Lee J</td>
<td>2011</td>
<td>Denmark</td>
<td>4,127</td>
<td>37,964</td>
<td>Fu WP</td>
<td>2006</td>
<td>China</td>
<td>256</td>
<td>266</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Our current study</td>
<td>2011</td>
<td>China</td>
<td>310</td>
<td>203</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>6,471</td>
<td>40,577</td>
<td>Total</td>
<td></td>
<td></td>
<td>1,228</td>
<td>1,358</td>
</tr>
</tbody>
</table>

**Publication bias**

We used funnel plots (Begg and Mazumdar, 1994) to identify potential publication bias (Figure 3), and the funnel plot asymmetry tests (Egger et al., 1997) showed no significant evidence of publication bias for \( EPHX1 \) T337C (\( P = 0.233 \)) and for \( EPHX1 \) A416G (\( P = 0.142 \)).

**VT tests based on three functional effect scores**

**VT tests based on three functional effect scores using Caucasian text data and Eastern Asian text data extracted from the meta-analysis**

Because we were unable to retrieve the original individual genotype data from the text data, which means that the two SNPs of \( EPHX1 \) could not be subjected to the VT test.
Determine possible association between EPHX1 and COPD

### Figure 1
Forest plot describing the association between two SNPs of EPHX1 and COPD risk under the allelic model for Caucasian population. **A.** EPHX1 T337C (rs1051740). **B.** EPHX1 A416G (rs2234922).

### Figure 2
Forest plot describing the association between two SNPs of EPHX1 and COPD risk under the allelic model for Eastern Asian population. **A.** EPHX1 T337C (rs1051740). **B.** EPHX1 A416G (rs2234922).
simultaneously, we performed the VT tests based on *EPHX1* T337C and A416G respectively for the Caucasian and Eastern Asian populations. For *EPHX1*, the risk function score z-values based on three functional effect scores were computed and the statistical significance of z-values were determined by permuting the sample labels 1000 times. All three functional effect scores showed that *EPHX1* T337C is a high-risk SNP, whereas *EPHX1* A416G was found to be a low-risk SNP (Figure 4). For Caucasian text data, the VarioWatch-based VT test showed a minor association between *EPHX1* T337C and COPD risk (VT-P = 0.048), while SIFT-based and Polyphen-2-based VT tests did not show any significant association between *EPHX1* and COPD (VT-P > 0.05) (Table 2). For Eastern Asian text data, the VT tests based on three functional effect scores consistently showed that *EPHX1* was not associated with COPD (VT-P > 0.05) (Table 3).

![Funnel plots for publication bias tests of two SNPs of *EPHX1*. A. *EPHX1* T337C (C vs T). B. *EPHX1* A416G (G vs A). Each point represents an individual study for the indicated association.](image1)

![Functional effect scores of two SNPs based on three functional prediction algorithms (1-SIFT, PolyPhen-2, and VarioWatch). A. *EPHX1* T337C (rs1051740). B. *EPHX1* A416G (rs2234922).](image2)
Determine possible association between EPHX1 and COPD

Table 2. Results of VT tests based on three functional effect scores for EPHX1 using Caucasian text data extracted from the meta-analysis.

<table>
<thead>
<tr>
<th>Functional prediction programs</th>
<th>EPHX1 T337C (rs1051740)</th>
<th>EPHX1 A416G (rs2234922)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Functional effect score</td>
<td>VT-(z) VT-P</td>
</tr>
<tr>
<td>VarioWatch</td>
<td>0.800</td>
<td>1.020 0.048*</td>
</tr>
<tr>
<td>SIFT</td>
<td>0.990</td>
<td>0.418 0.147</td>
</tr>
<tr>
<td>PolyPhen-2</td>
<td>0.998</td>
<td>0.524 0.231</td>
</tr>
</tbody>
</table>

VT = variable threshold; *VT-P < 0.05.

Table 3. Results of VT tests based on three functional effect scores for EPHX1 using Eastern Asian text data extracted from the meta-analysis.

<table>
<thead>
<tr>
<th>Functional prediction programs</th>
<th>EPHX1 T337C (rs1051740)</th>
<th>EPHX1 A416G (rs2234922)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Functional effect score</td>
<td>VT-(z) VT-P</td>
</tr>
<tr>
<td>VarioWatch</td>
<td>0.800</td>
<td>0.665 0.109</td>
</tr>
<tr>
<td>SIFT</td>
<td>0.990</td>
<td>0.525 0.310</td>
</tr>
<tr>
<td>PolyPhen-2</td>
<td>0.998</td>
<td>0.691 0.566</td>
</tr>
</tbody>
</table>

VT = variable threshold.

VT tests based on three functional effect scores using our true Chinese Han data

Among the 13 SNPs involved in EPHX1 genotyped from our Chinese Han population, five SNPs (rs1051740, rs2234922, rs1009668, rs1051741, and rs2292568) were nsSNPs whose functional effect scores were calculated using the different programs. The VarioWatch-based VT test was used to determine the functional effect scores of five nsSNPs of EPHX1, while the SIFT-based and PolyPhen-2 based VT tests utilized the functional effect scores of three nsSNPs because of the missing scores, which could not be calculated by the corresponding functional prediction programs. As a result, the VT tests based on three functional effect scores were completely consistent (VT-P > 0.05), suggesting that there is no significant association between EPHX1 and COPD in the Chinese Han population (Table 4).

Table 4. Results of VT tests based on three functional effect scores for EPHX1 using our true Chinese Han data.

<table>
<thead>
<tr>
<th>nsSNP name</th>
<th>P value (allelic association test)</th>
<th>VT-(z) VT-P</th>
<th>Functional effect score</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1051740</td>
<td>0.2056</td>
<td>0.8</td>
<td>0.990 0.998</td>
</tr>
<tr>
<td>rs2234922</td>
<td>0.8669</td>
<td>0.4</td>
<td>0.310 0.000</td>
</tr>
<tr>
<td>rs1051741</td>
<td>0.1006</td>
<td>0.2</td>
<td>NA 0.391</td>
</tr>
<tr>
<td>rs1009668</td>
<td>0.0105</td>
<td>0.3</td>
<td>NA 0.384</td>
</tr>
<tr>
<td>rs2292568</td>
<td>0.6254</td>
<td>0.4</td>
<td>NA 0.413</td>
</tr>
<tr>
<td>VT-P</td>
<td>0.084</td>
<td>0.982</td>
<td>0.117</td>
</tr>
</tbody>
</table>

NA = not available due to the missing scores which could not be calculated by the corresponding functional prediction program for this nsSNP.

DISCUSSION

Numerous previous studies examining the genetic susceptibility to COPD have focused on EPHX1 because it encodes a xenobiotic metabolizing enzyme that detoxifies the reactive epoxides produced by smoking to form water and soluble dihydrodiol compounds (Li et al., 2013). However, genetic associations with COPD identified in these studies have
been inconsistent; therefore, it is difficult to interpret these findings. Although meta-analyses can improve statistical power and avoid error sources by summarizing related studies, they are limited by their potential bias. Here, we conducted combined data mining based on a meta-analysis and a gene-based VT test using the functional effect scores of nsSNPs predicted by three algorithms to determine the potential association between \textit{EPHX1} and COPD in Caucasian and Eastern Asian populations. Our combined data analysis results suggest that genetic variations in \textit{EPHX1} are not related to COPD in Caucasian and Eastern Asian populations.

There were some limitations to this meta-analysis. First, this analysis was limited to population-based case-control studies in order to decrease the complexity. The inclusion of family-based studies may strengthen the analysis. Second, discrepancies in the observed results may be caused by a series of factors, including differences in study methods, genotyping errors, heterogeneity of enrolled cases, outcome definition variability, and limited statistical power. Third, we use the allelic model to perform the meta-analysis, limiting the study to the genetic main effect. In fact, COPD susceptibility is more likely caused by gene-gene and gene-smoking interactions. Fourth, a small proportion of studies can increase the instability of meta-analysis. When we divided the selected studies into Caucasian and Eastern Asian populations, the number of studies included in each subgroup was decreased, decreasing the statistical test power of the meta-analysis. In addition, our analysis did not include GWAS results because of a lack of data. Including GWAS results will help overcome publication bias. To overcome the limitations and potential bias, further well-designed studies with larger sample sizes should be performed in the future. As a response to the limited sample size, incorporation of functional effect scores of nsSNPs into the gene-based VT test will improve statistical power. It has been proposed that statistical power can be increased by 15-50\% when combining SNP-based analysis with gene-disease association studies (Thomas et al., 2003). In fact, if different mutations in a gene affect disease risk, the entire gene should be focused on as a risk gene rather than focusing on individual risk variants. In the present study, we conducted combined data mining based on functional prediction algorithms of nsSNPs and gene-based VT test using three datasets. The results revealed that genetic variations in \textit{EPHX1} were unrelated to COPD. This result is consistent with the performed meta-analysis. Furthermore, among the 13 genotyped SNPs involved in \textit{EPHX1} from our Chinese Han population, single SNP association analysis showed that only one SNP (7.7\%) was significant after Bonferroni correction. This result also supports a lack of an association between \textit{EPHX1} with COPD. However, because the data used in the current meta-analysis were not extracted from the original data sets but rather taken directly from published articles, the VT tests based on these meta-data may show bias for the association. Therefore, we cannot completely exclude the effect of \textit{EPHX1} on COPD risk, which must be validated using more data in the future.

In conclusion, we combined data mining based on a meta-analysis and a gene-based VT test using three functional effect scores of nsSNPs. The results revealed no association between genetic variations in \textit{EPHX1} and COPD risk in Caucasian and Eastern Asian populations. However, a minor effect of \textit{EPHX1} on COPD risk cannot be completely excluded and should be validated using more data obtained from further studies in the future.

\textbf{Conflicts of interest}

The authors declare no conflict of interest.
ACKNOWLEDGMENTS

Research supported by Beijing Natural Science Foundation (Grant #7142015) and National Natural Science Foundation of China (Grant #31100905). Research supported by the Excellent Talent Cultivation Project of Beijing (#2012D005018000002) and the Foundation-Clinical Cooperation Project of Capital Medical University (#14JL43).

REFERENCES


Korytina GF, Ianbaeva DG and Viktorova TV (2003). Role of polymorphic variants of cytochrome P450 genes (CYP1A1, CYP2E1) and microsomal epoxide hydrolase (mEPHX) in pathogenesis of cystic fibrosis and chronic respiratory tract diseases. *Mol. Biol.* 37: 784-792.


Yim JJ, Park GY, Lee CT, Kim YW, et al. (2000). Genetic susceptibility to chronic obstructive pulmonary disease in Koreans: combined analysis of polymorphic genotypes for microsomal epoxide hydrolase and glutathione S-transferase M1 and T1. *Thorax* 55: 121-125. [http://dx.doi.org/10.1136/thorax.55.2.121](http://dx.doi.org/10.1136/thorax.55.2.121)


