Replication of genome-wide association studies on asthma and allergic diseases in Korean adult population

Dankyu Yoon1,2,*, Hyo-Jeong Ban3,8, Young Jin Kim4, Eun-Jin Kim4, Hyung-Cheol Kim4, Bok-Ghee Han5, Jung-Won Park1,6, Soo-Jong Hong1,7, Sang-Heon Cho1,8, Kiejung Park3,* & Joo-Shil Lee1,2,*

1Research Center for Standardization of Allergic Disease, 2Allergy TF, Department of Immunology and Pathology, Korea National Institute of Health, 3Division of Bio-Medical Informatics, Center for Genome Science, Korea National Institute of Health, 4Division of Structural and Functional Genomics, Center for Genome Science, Korea National Institute of Health, 5Department of Internal Medicine and Institute of Allergy, Yonsei University College of Medicine, Seoul 120-752, 6Childhood Asthma Atopy Center, Department of Pediatrics, Asan Medical Center, University of Ulsan College of Medicine, Seoul 138-736, 7Department of Internal Medicine, Seoul National University College of Medicine, Seoul 110-799, Korea

Allergic diseases such as asthma, allergic rhinitis, and atopic dermatitis are heterogeneous diseases characterized by multiple symptoms and phenotypes. Recent advancements in genetic study enabled us to identify disease associated genetic factors. Numerous genome-wide association studies (GWAS) have revealed multiple associated loci for allergic diseases. However, the majority of previous studies have been conducted in populations of European ancestry. Moreover, the associations of single nucleotide polymorphisms (SNPs) with allergic diseases have not been studied amongst the large-scale general Korean population. Herein, we performed the replication study to validate the previous variants, known to be associated with allergic diseases, in the Korean population. In this study, we categorized three allergic related phenotypes, one allergy and two asthma related phenotypes, based on self-reports of physician diagnosis and their symptoms from 8,842 samples. As a result, we found nominally significant associations of 6 SNPs with at least one allergic related phenotype in the Korean population. [BMB reports 2012; 45(5): 305-310]

INTRODUCTION

Allergic diseases are complex ailments caused by crosstalk between multiple genes and environmental factors (1). According to a recent report from the World Allergy Organization (WAO), the prevalence of allergic diseases is rising dramatically on a global scale (2). Allergic diseases brought the significant burden of global morbidity and mortality in the world and are regarded as one of the major contributors to the considerable drain on health budgets of developed and emerging economies (2). Given the reports from the Korean National Health and Nutrition Examination Survey (3), the prevalence of allergic diseases is increasing in Korea, and the prevalence of asthma, allergic rhinitis and atopic dermatitis in adult population are 7.6%, 11.9% and 2.95%, respectively. Therefore, it is valuable to identify genes or loci associated with allergic diseases for understanding underlying complex mechanisms of diseases.

Numerous previous genetic studies have been conducted using linkage designs and candidate gene association studies to elucidate etiology of allergic diseases (4). Although these cases provide strong evidence for the involvement of genetic factors in allergic diseases, identifying these susceptible genes has been a challenge. A number of genome-wide association studies (GWAS) have been conducted for various allergic disease/phenotypes, including asthma (5-12), atopy (13), atopic dermatitis (14), serum IgE levels (15), eosinophilic esophagitis (16), and plasma eosinophil count (17). However, most of these GWASs have been focused primarily on samples of European origin. Currently, more than two dozen GWAS for asthma and allergic disease have been performed (18). However, few GWAS have been conducted in populations of racial and ethnic minorities. Moreover, several prior studies reported significant discrepancies in allele frequencies of the variants and differences in genetic architecture between European and Asian (19, 20). In this context, we conducted the replication study on allergic disease amongst an Asian population.

RESULTS

Description of KoGES for KARE project in its relation allergic phenotypes
The Korean Association REsource (KARE) was established as a part of the Korean Genome and Epidemiology study (KoGES).
KARE, 8,842 subjects with 352,228 SNPs were analyzed. Average ages of male and female were 51.78 ± 8.78 (SD) and 52.61 ± 9.01 (SD), respectively. Female participants were 52.7% (N=4,659) of all subjects, 5.5% and 2.2% of these had allergy and asthma diagnostic history, respectively. The characteristics of the study are shown in Table 1.

Comparison analysis between previous GWAS results and KARE
To gain insight into the genetic influence of previously allergic diseases associated loci in an Asian population, we conducted a replication study for those known candidate loci in the Korean population. From GWAS catalog, 46 significant associations of 12 independent GWA studies on allergic diseases were retrieved. Supplementary Table 1 shows the summarized information of all 46 known associations. For comparison analysis, we retrieved the corresponding SNPs comprising directly genotyped, imputed genotype and proxy SNPs in LD (r2 > 0.8). As a result, 32 SNPs were available in KARE (Supplementary Table 1). As expected, difference in allele frequencies across populations was observed. For example, the SNPs including rs1342326, rs17525472, rs3184504 and rs4815617 are monomorphic in Asian populations. Due to this disparity in allele frequencies, those loci were not able to directly compare to those of SNPs in our dataset. We defined SNPs with P values less than 0.05 as a replicated result. Given these criteria, we observed 6 replicated SNPs in total. We summarized the comparison analysis (Supplementary Table 2) and the replicated results (Table 2). We successfully replicated three allergy phenotype associated SNPs comprising rs12619285 near IKZF2; IKAROS family zinc finger 2), rs1295686 (at IL13; Interleukin 13) and rs2073643 (at IL13; Interleukin 13) and rs2073643 (at SLCL22A5; solute carrier family 22 (organic cation/carnitine transporter), member 5). In asthma phenotypes, rs13106227 (at SHROOM3; shroom family member 3) for both asthma I and asthma II, rs2416257 (at WDR36; WD repeat domain 36 and near TSLP; Thymic stromal lymphopoietin) for asthma II and rs3806932 (near WDR36) for asthma I were replicated in our study.

In order to scrutinize the reason for replication failure of 26 SNPs, the statistical power was calculated. The power analysis revealed that 26 non-replicated SNPs in our study would be detected due to the lack of statistical power (Supplementary Table 2). All 6 replicated loci have relatively high statistical power (0.36-0.86) for the disease association while the power of the remaining variants was found to be relatively low (< 0.47).

Since the statistical power is the main reason for non-replication, we performed an alternative association analysis. Gene-based association analysis is a well-known approach for testing the association between multiple markers within gene region and a trait. It is known to be more powerful than the classical single marker association test under the condition of genetic architecture (21). We performed a gene-based association test for the candidate gene regions near previously reported 46 variants. We observed significant results (P-value < 0.05) at the loci near replicated variants from single SNP analysis in one or more allergic disease related phenotypes (Supplementary Table 3). Unfortunately, however, we did not observe any newly replicated locus that is not listed in the single SNP analysis.

DISCUSSION
We performed the replication study on previously known allergic diseases associated loci in the Korean population. Previously, Kim et al. (8) reported GWA results of toluene-induced asthma. However, the investigation was mainly conducted in a population with specific material induced disease. Thus, to the best of our knowledge, we present the first replication study focused on the general Korean population.

From 46 previously reported allergic disease related variants, 32 SNPs were available for the single SNP association test in the current study excluding 4 monomorphic SNPs and 9 SNPs with low imputation quality. Among 32 SNPs, 6 SNPs were successfully replicated in this study (P-value < 0.05). The result from gene-based association test also confirmed the associations of 4 loci that were identified by single marker analysis. Four of six SNPs were found to be located at chromosome 5 including rs1295686 at IL13, rs2073643 at SLCL22A5, rs2416257 at WDR36 and rs3806932 near WDR36. Both IL13 and SLCL22A5 have been reported to be associated with asthma (6). WDR36 was previously associated with eosinophilic esophagitis (16), plasma eosinophil count, as well as with atopic asthma (17). Interestingly, Gudbjartsson et al. (17) confirmed this signal in both Europeans and Asians. SHROOM3 and IKZF2 were reported to be associated with eosinophilic esophagitis (16) and plasma eosinophil count (17), respectively. Here we excluded 9 SNPs due to low im-

Table 1. Characteristics of the study population

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Total</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size (N, %)</td>
<td>8,842</td>
<td>4,183 (47.31)</td>
<td>4,659 (52.69)</td>
</tr>
<tr>
<td>Age (year) (mean ± SE)</td>
<td>52.22 ± 8.91</td>
<td>31.78 ± 8.78</td>
<td>52.61 ± 9.01</td>
</tr>
<tr>
<td>BMI</td>
<td>24.60 ± 3.12</td>
<td>24.25 ± 2.92</td>
<td>24.90 ± 3.25</td>
</tr>
<tr>
<td>FEV1 (%)</td>
<td>111.7 ± 17.7</td>
<td>106.3 ± 16.25</td>
<td>116.6 ± 17.58</td>
</tr>
<tr>
<td>FEV1/FVC (%)</td>
<td>79.81 ± 7.76</td>
<td>77.91 ± 8.39</td>
<td>81.66 ± 6.65</td>
</tr>
<tr>
<td>Allergic disease phenotype</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allergy diagnosis (N)</td>
<td>8,348</td>
<td>4,008</td>
<td>4,340</td>
</tr>
<tr>
<td>Control</td>
<td>490</td>
<td>173</td>
<td>317</td>
</tr>
<tr>
<td>Asthma I (N)</td>
<td>8,645</td>
<td>4,111</td>
<td>4,534</td>
</tr>
<tr>
<td>Control</td>
<td>193</td>
<td>70</td>
<td>123</td>
</tr>
<tr>
<td>Asthma II (N, %)</td>
<td>7,450</td>
<td>3,632</td>
<td>3,818</td>
</tr>
<tr>
<td>Control</td>
<td>44</td>
<td>14</td>
<td>30</td>
</tr>
</tbody>
</table>

BMI: Body Mass Index (kg/m²), FEV1(%) : % predicted value of forced expiratory volume in 1 second, FEV1/FVC (%) : % predicted value of FEV1, and forced vital capacity (FVC). Asthma I: asthma diagnosis, Asthma II: asthma diagnosis and self-reported symptom.
pulation quality prior to comparison analysis based on single SNP analysis. Despite the unavailability for directly matched or proxy makers of these 9 SNPs, we were able to perform the gene-based association test on 7 of 9 excluded loci in the single SNP analysis while 2 gene desert loci were excluded from the analysis. However, these 7 loci were not replicated in the gene-based association analysis (P-value > 0.05). The remainder of the unreplicated 26 SNPs was due to lack of statistical power. In the pool of the failed SNPs, all SNPs were below 50% of statistical power. These results together indicate that further analysis including more sampling and genotyping is required for exploring the possible association of non-replicated loci and SNPs with low imputation quality.

As mentioned above, some SNPs (such as rs1342326, rs3184504, rs4815617, and rs17525472) were not polymorphic in Asian populations. All 4 loci showed distinct LD patterns and some evidence of genetic differentiation across populations (Supplementary Fig. 1 and Supplementary Fig. 2). Given iHS (Integrated Haplotype Score) values within the loci, little evidence of population difference was shown in terms of positive selection. However, Fst score indicates that genetic diversity between populations exists around the loci. For instance, the replication quality prior to comparison analysis based on single SNP analysis.

Table 2. Significant results of replication analysis on allergic disease phenotype in KARE

<table>
<thead>
<tr>
<th>SNP</th>
<th>chr</th>
<th>Position</th>
<th>Gene</th>
<th>Allele</th>
<th>Allergy OR</th>
<th>SE</th>
<th>P value</th>
<th>Asthma OR</th>
<th>SE</th>
<th>P value</th>
<th>Asthma&amp; self reported symptoms OR</th>
<th>SE</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs12619285</td>
<td>2q34</td>
<td>213532290</td>
<td>IKZF2</td>
<td>A</td>
<td>0.865</td>
<td>0.07</td>
<td>0.038</td>
<td>0.907</td>
<td>0.109</td>
<td>0.371</td>
<td>1.028</td>
<td>0.22</td>
<td>0.899</td>
</tr>
<tr>
<td>rs13106227</td>
<td>4q21.1</td>
<td>77637705</td>
<td>SHROOM3</td>
<td>G</td>
<td>0.901</td>
<td>0.067</td>
<td>0.123</td>
<td>0.733</td>
<td>0.107</td>
<td>0.004</td>
<td>0.462</td>
<td>0.241</td>
<td>0.001</td>
</tr>
<tr>
<td>rs1295686</td>
<td>5q22.1</td>
<td>132023742</td>
<td>IL13</td>
<td>T</td>
<td>1.237</td>
<td>0.069</td>
<td>0.002</td>
<td>1.042</td>
<td>0.11</td>
<td>0.707</td>
<td>1.19</td>
<td>0.222</td>
<td>0.433</td>
</tr>
<tr>
<td>rs2073643</td>
<td>5q31.1</td>
<td>131751187</td>
<td>SLC22A5</td>
<td>C</td>
<td>0.84</td>
<td>0.082</td>
<td>0.033</td>
<td>0.785</td>
<td>0.129</td>
<td>0.059</td>
<td>0.763</td>
<td>0.271</td>
<td>0.319</td>
</tr>
<tr>
<td>rs2416257</td>
<td>5p22.1</td>
<td>110463389</td>
<td>WDR36, TSLP</td>
<td>T</td>
<td>1.007</td>
<td>0.151</td>
<td>0.962</td>
<td>1.079</td>
<td>0.227</td>
<td>0.738</td>
<td>2.723</td>
<td>0.327</td>
<td>0.002</td>
</tr>
<tr>
<td>rs3806932</td>
<td>5q22.1</td>
<td>110433574</td>
<td>WDR36</td>
<td>G</td>
<td>0.962</td>
<td>0.07</td>
<td>0.586</td>
<td>1.271</td>
<td>0.107</td>
<td>0.025</td>
<td>1.342</td>
<td>0.22</td>
<td>0.182</td>
</tr>
</tbody>
</table>

Replication of GWAS on asthma and allergic diseases in Korean

Dankyu Yoon, et al.

Definition of Allergic related phenotype

Allergic disease related phenotypes are defined based on information from self-report of physician diagnosis and their symptoms. On the basis of the survey questionnaire, three allergic disease phenotypes were defined: one of them for allergy and two of them for asthma related phenotypes. Allergy diagnosis phenotype was defined based on the history of physician diagnosis on allergic rhinitis and atopic dermatitis. Asthma phenotypes were defined according to the history of physician diagnosis on asthma and self-reported symptoms. Asthma related

http://bmbreports.org
phenotypes were classified in detail by the presence of self-reported symptoms. Asthma I was defined based on the history of physician diagnosis on asthma. Asthma II was defined by both the history of physician diagnosis as well as awakening with chest tightness, cough, wheeze, or night time awakening due to shortness of breath during the last 12 months. The control samples were retrieved from 8,842 samples based on no history of asthma, wheezing and allergy.

Comparison between previous GWAS result and KARE

Twelve GWAS on allergic diseases have been reported and 46 statistically significant results from each study (P-value < 1 × 10^{-7}) were retrieved from GWAS catalog (24). Due to the limited amount of directly genotyped data, we used imputed genotype data for the comparison analysis. Those SNPs that were not available in our directly genotyped or imputed data were compared to proxy markers (SNPs in linkage disequilibrium, r^2 > 0.8 in CHB/JPT HapMap Phase II samples) within our dataset. Among 46 previously reported SNPs, 13 SNPs were excluded for further analysis due to monomorphism in Asians (4 SNPs) and low imputation quality (9 SNPs, information score < 0.5).

Statistical analysis

Statistical analysis was performed using PLINK (25) and R software. We performed logistic regression analysis for the binary phenotypes, adjusting age, sex, and recruiting geographic area under additive models. Imputation was carried out by using the IMPUTE program on the basis of NCBI build 36 and dbSNP build 126, we initially used 90 individuals from JPT and CHB founders in HapMap as a reference panel (release 22) (26). Before the association analysis on the imputed genotypes, we removed imputed SNP markers with low genotype information content (info < 0.5), posterior probability score < 0.90, low call rate < 0.90, MAF < 0.01 and HWE (P-value < 1 × 10^{-7}). Among the markers, we selected 32 previously known variants for further analysis. To calculate the statistical power, we used the R statistics software and its package “GeneticsDesign”. The power analysis was performed under the additive model and with specific parameters such as odds ratio as estimated in this study, a disease prevalence of 0.07 (3), marker allele frequency and low imputation quality (9 SNPs, information score < 0.5).

Gene-based associations

A gene-based approach tests for the association between a phenotype and all variants within a gene region. We used set-based tests implemented in PLINK software for the gene-based association test. Gene information near 46 previously allergic diseases associated SNPs was retrieved from UCSC genome browser (hg18). In the association test, we selected options for analysis as r-square = 0.5, 10,000 permutations, and no threshold for p-value of the variants within gene region. After excluding the gene regions with no available SNPs, we performed the set-based test on 34 genes near previously reported variants from asthma and allergic GWAS.

Assessment of linkage disequilibrium (LD) and population structure

To view the LD patterns of genomic regions across populations, we used Haploview software (27) to draw LD plots based on HapMap phase II CEU and CHB/JPT populations. In addition to the graphical view of difference in genetic structure, we also assessed the differential levels of genetic structure using iHS (Integrated Haplotype Score) and Fst score. Haplotter is used for retrieving pre-calculated iHS and Fst score (28). iHS is a statistic representing the degree of recent positive selection at a locus. Fst is a measure of the magnitude of population differentiation between two populations.

Acknowledgements

This study was supported by a grant from the Korea Healthcare Technology R&D Project, Ministry of Health & Welfare, Republic of Korea (A092076). This work was also supported by grants from the Korea Centers for Disease Control and Prevention (The Korean Genome Analysis Project (4845-301), KoGES (4851-302), and Korea Biobank Project (4851-307)) and an intramural grant from the Korea National Institute of Health (2001-N73002-00, 2009-N72001-00), the Republic of Korea. We thank Nam H. Cho (Department of Preventive Medicine, Ajou University, South Korea) and Chol Shin (Department of Internal Medicine, Korea University Ansan Hospital, South Korea) for their great efforts in generating and providing this valuable data.

REFERENCES


ciated with eight hematological parameters in the HaemGen 

nome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. Nat. Genet. 41, 527-534.


