Clinical Implications of Single Nucleotide Polymorphisms in Diagnosis of Asthma and its Subtypes

Jong-Sook Park¹, Ji-Hye Son², Choon-Sik Park¹, and Hun Soo Chang²

¹Division of Allergy and Respiratory Medicine, Department of Internal Medicine, Soonchunhyang University Bucheon Hospital, Bucheon; ²Department of Interdisciplinary Program in Biomedical Science, Graduate School, Soonchunhyang University, Bucheon, Korea.

For the past three decades, a large number of genetic studies have been performed to examine genetic variants associated with asthma and its subtypes in hopes of gaining better understanding of the mechanisms underlying disease pathology and to identify genetic biomarkers predictive of disease outcomes. Various methods have been used to achieve these objectives, including genome-wide linkage analysis, candidate gene polymorphism analysis, and genome-wide association studies (GWAS); however, the degree to which genetic variants contribute to asthma pathogenesis has proven to be much less significant than originally expected. Subsequent application of GWAS to well-defined phenotypes, such as occupational asthma and non-steroidal anti-inflammatory drug-exacerbated respiratory diseases, has overcome some of these limitations, although with only partial success. Recently, a combinatorial analysis of single nucleotide polymorphisms (SNPs) identified by GWAS has been used to develop sets of genetic markers able to more accurately stratify asthma subtypes. In this review, we discuss the implications of the identified SNPs in diagnosis of asthma and its subtypes and the progress being made in combinatorial analysis of genetic variants.

Key Words: Asthma, aspirin, non-steroidal anti-inflammatory agents, biomarkers, single nucleotide polymorphism, genetic techniques

INTRODUCTION

Asthma and its associated traits exhibit a strong familial connection, indicating that a genetic component likely contributes to the disease pathology. Twin studies are commonly used to determine whether complex multifactorial diseases, such as asthma, have a measurable genetic component. In a study of 7000 monozygotic twins performed over 100 years ago, the concordance rate for asthma between monozygotic twin pairs was 19%, nearly four times higher than that of dizygotic twins (4.8%).¹ The overall heritability of asthma has since been estimated at anywhere from 15–60%, and genetic variants, such as single nucleotide polymorphisms (SNPs), are presumed to play an important role in the etiology of the disease. To identify SNPs in multifactorial complex diseases, several approaches, including genome-wide linkage studies, biologically plausible candidate gene association studies, and genome-wide association studies (GWAS) have been performed, revealing more than 100 loci on autosomal and sex chromosomes to be linked to asthma. Linkage analysis has identified SNPs on several genes, including ADAM33,² DPP10,³ PHF11,⁴ GPR,⁵ and PTGDR,⁶ as strong genetic variants for asthma.

Over the past decade, several GWAS have been performed, examining 300K–500K SNPs and producing a nearly complete survey of all common genetic variability.⁷ In 2010, the GABRIEL consortium used GWAS to confirm the association of asthma with many of the previously identified SNPs associated with antigen presentation (HLA-DR/DQ), inflammation (ORMDL3-GSDMB: Gasdermin B), and T(H)1/T(H)2 immune responses (IL33, IL-1RL1-IL18R1, RAD50-IL13, and TSLP-WDR36).⁸ However, the odds ratios (ORs) for these markers ranged from 0.5–2.0, much lower than originally expected. Furthermore, combination of these seven SNPs for the classification of childhood asthma pro-
duced a sensitivity of only 35% and a specificity of 75%, indicating these markers as poor predictors of asthma overall.

Despite the poor predictive power suggested by the GABRIEL consortium study,8 analysis of genetic variants associated with asthma continues to flourish. As of December 2017, 3437 publications were identified in the PubMed database using the searching terms “asthma” and “polymorphism.” Interestingly, the ORs for most of the associated SNPs ranged between 0.5 and 2.0, indicating a much smaller genetic component in the development of asthma than initially expected. Possible explanations for this poor predictive power included a lack of rare allele variants in the GWAS database, imprecise disease phenotypes, and incomplete consideration of environmental factors.9 Furthermore, a more extensive combinatorial analysis using the identified SNPs has not been thoroughly studied to date. The present review focuses on the genetic impact of rare allele variants on asthma and the clinical usefulness of combinatorial analysis using identified SNPs on nonsteroidal anti-inflammatory drug-exacerbated respiratory diseases (NERD).

Genetic Effect of Rare Variants in Asthma

Until recently, much of the speculation regarding the missing heritability not detected using standard GWAS has focused on the possible contribution of rare variants [minor allele frequency (MAF) <0.5%], as only common variants (MAF>5%) were available in previous GWAS datasets. For more information on these rare variants, the 1000 Genomes Project sequenced 1000 individual genomes, revealing more than 15 million new SNPs, 1 million short insertions and deletions, and 20000 structural variants.10,11 More recently, Igartua, et al.11 investigated associations among rare (<1%) and low-frequency (1–5%) variants using a Human Exome Bead Chip array containing 246139 SNPs in ~5000 asthmatic patients and 5000 controls, representing a broad cross section of European, African American/African Caribbean, and Latino subjects. This analysis revealed a number of rare variants for asthma to be associated with ethnicity, including SNPs in GRASP (general receptor for phospohoinositides 1-associated scaffold protein), MTHFR (methylene tetrahydrofolate reductase), and GSDMB in association with asthma in Latino, African, and all ethnicities, respectively. These data suggest that associations with rare and low-frequency variants are specific for each ethnicity and not likely to explain a significant proportion of the genetic variants for asthma.

Improvement of Genetic Effects According to Stratification of Phenotypes

Because asthma is a heterogeneous disease, use of well-defined sub-phenotypes may improve the genetic power of SNPs. Although asthma has been traditionally divided into two major pathways based on the mechanism of pathogenesis,12 further subgrouping is possible based on the factors that trigger asthma symptoms,13 including allergic asthma, exercise-induced asthma, NERD, occupational asthma, menstruation-associated asthma, and others. Furthermore, inflammatory patterns in sputum and peripheral blood analysis can be used to divide asthma into four groups: eosinophilic, neutrophilic, mixed, and paucigranulocytic types. Clinically and physiologically, asthmatics are also stratified into well controlled, exacerbation-prone, and remodeled asthma.

Recently, biological classification has been introduced on the basis of molecular pathogenesis. Candidate endotypes include several biological mediators and functional immune subsets, including Th1, Th2, Th17, innate Th2, epithelial, and smooth muscle dysfunctions, as well as others (Table 1). Identification of genetic associations would therefore have to be studied using cluster analysis of well-defined sub-phenotypes, with patients stratified based on clinical manifestations and endotypes.14,15 Furthermore, quantitative loads of environmental factors should also be considered, as asthma is known to be triggered or exacerbated by a wide range of compounds, including air pollutants,16 occupational materials,17,18 nutritional status,19,20 and food allergens.21-24 Of these factors, occupation is one of the most easily accessible triggers. In a

---

**Table 1. Classification of Asthma Phenotypes according to Etiologic Agent, Clinical and Inflammatory Patterns, and Molecular Mechanisms**

1. **Phenotypes related to disease etiology**
   - Aspirin or non-steroidal anti-inflammatory drugs
   - Environmental allergens
   - Occupational allergens or irritants
   - Estrogens
   - Exercise
   - Viral infection
   - Obesity

2. **Clinical or physiological phenotypes**
   - Severity-defined: from mild to severe
   - Exacerbation-prone: brittle vs. non-brittle
   - Defined by chronic airflow restriction: remodeled
   - Treatment-resistant: resistance to steroids
   - Defined by age at onset: childhood, adulthood, old age

3. **Inflammatory phenotypes**
   - Eosinophilic
   - Neutrophilic
   - Pauci-granulocytic

4. **Molecular phenotypes**
   - Th1 dominant
   - Th2 dominant
   - Th17 dominant
   - Innate Th1,2 dominant
   - Epithelial dysfunction
   - Smooth muscle dysfunction

---

https://doi.org/10.3349/ymj.2019.60.1.1
GWAS using the well-defined Korean toluene diisocyanate (TDI)-induced asthmatic cohort, \(^{25}\) SNPs in \(CTNN43\) (catenin alpha 3, alpha-T catenin) were significantly associated with the risk of TDI-induced asthma (OR=5.84 for rs10762058). The attributable fraction increased up to 24%, indicating that the missing heritability of asthma has been solved in part by the consideration of environmental variables. Very recently, homozygosity of the minor allele SNPs rs10762058 and rs7088181 was shown to increase the risk for di-isocyanate asthma in Caucasian workers, with an OR=9.05 [95% confidence interval (CI): 1.69–48.54], \(^{26}\) a level high enough to be considered useful as a genetic marker. Another example is that of drug-induced reactions in susceptible persons, such as those with NERD. Initial genetic association studies of NERD patients used a set of biologically plausible genes associated with the over- or under-production of arachidonic acid metabolites. Genes such as \(LT CAS, ALOX5, NAT2, CysLTR1,\) and \(CysLTR2,\) all members of the cysteinyl leukotriene pathway, exhibited several NERD-associated SNPs, with ORs ranging from 1.88 to 9.78. \(^{18,27}\) Genes of prostaglandin and thromboxane pathways discovered using GWAS \(^{28}\) also showed good ORs, although not at levels high enough for use as a stand-alone diagnostic biomarker.

**IMPROVEMENT IN DIAGNOSTIC VALUES BY COMBINATORIAL ANALYSIS OF MULTIPLE SNPS OBTAINED FROM GWAS**

The method most often used for measuring predictive factors is the area under a receiver operator characteristic (ROC) curve (Fig. 1).\(^{29}\) The area under this curve (AUC) is frequently used to discriminate between those with and without disease. AUC values range from 0.5 (providing no discrimination between the two conditions) to 1.0 (perfect discrimination), with ideal classification exhibiting almost square curves, indicative of high true-positive and low false-positive rates. Disease risk prediction is frequently described using a combination of sensitivity, specificity, and accuracy. In order to obtain a sensitivity and specificity of 80% or more, which is typically required for clinical utility, the absolute OR should be greater than 10 (Fig. 1). AUC values predictive of a hypothetical condition carry modest (1.5), sizeable (10), and large (50) odds ratios (ORs), showing false-positive fractions at 80% sensitivity (dotted line; fractions are >75, >25, and <10%, respectively). The graph demonstrates that very large ORs are needed to provide acceptably low false-positive fractions. Reprinted from Jakobsdottir, et al. PLoS Genet 2009;5:e1000337, with permission of Jakobsdottir, et al.\(^{30}\)

![Image](https://doi.org/10.3349/ymj.2019.60.1.1)

**Fig. 1.** Area under receiver operator characteristic (ROC) curve values predictive of a hypothetical condition carry modest (1.5), sizeable (10), and large (50) odds ratios (ORs), showing false-positive fractions at 80% sensitivity (dotted line; fractions are >75, >25, and <10%, respectively). The graph demonstrates that very large ORs are needed to provide acceptably low false-positive fractions.

in and around the 11 identified genes were genotyped on 163 NERD and 429 ATA as the second-stage mapping.\(^{31}\) Among the SNPs, rs7572857 G>A (Gly74Ser) of \(CEP 68\) was most significantly associated with the risk of NERD (\(p=6.06 \times 10^{-5}\), 95% CI=1.64–4.21, OR=2.63). Because the ratio of ATA and NERD subjects varies among studies, the number of NERD and ATA patients should be normalized according to the prevalence of NERD in a given population. When we applied the incidence of positive response in oral aspirin challenge (16.03%) in Korean adult asthmatics,\(^{32}\) the diagnostic value of rs7572857 G>A on \(CEP68\) was not high enough to justify clinical application due to the low accuracy (53%), sensitivity (18%), and specificity (88%) of this marker. Taken together, these data highlight the need for better methods capable of increasing the discriminating power of individual SNPs, as well as the evaluation of combinatorial effects of SNPs on disease resolution.\(^{33}\)

Using the 10 SNPs with the lowest \(p\) values from the first GWAS of NERD patients,\(^{31}\) relative risk (RR) was calculated for each individual subject using a multiple logistic regression analysis of the 1023 possible combinations (2\(^{10}-1\)) of SNPs.\(^{34}\) The ROC curve analysis demonstrated that a combinatorial set of eight SNPs was the best model for maximizing the AUC (0.9) for distinguishing NERD from ATA (Fig. 2). This model exhibited 82% accuracy, 78% sensitivity, and 88% specificity with an OR of 20.74 (\(p=3.24 \times 10^{-19}\)).\(^{34}\) Each of the eight SNPs, including Parkin RBR E3 Ubiquitin Protein Ligase, DDB1 and CUL4 Associated Factor 4, Vaccinia Related Kinase 2, Carbo-

---

\(^{25}\) Jong-Sook Park, et al.
hydrate Sulfotransferase 11, Long Intergenic Non-Protein Coding RNA 314, PDZ Domain Containing 2, Small G Protein Signaling Modulator 1, and Centromere Protein F exhibited ORs ranging from 0.32 to 2.91 (Table 2). These data suggest that the diagnostic accuracy of the SNPs as genetic markers is robustly increased by combinatorial analysis, compared to that of single SNPs alone.

In the second GWAS using the 660K chip, as well as a replication study on 141 subjects with NERD and 996 ATA subjects, rs1042151 (Met105Val) in exon 2 of HLA-DPB1 showed the highest susceptibility to NERD ($p=5.11\times10^{-7}$; OR=2.40). The rs1042151 acts as a potential cis regulator of the expression of HLA-DPB1 with an expression quantitative trait loci (eQTL) score of 36.83, as calculated using the eQTL browser (http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/). Further analysis of SNP function (https://snpinfo.niehs.nih.gov/snpinfo/snpfunc.html) indicated that the rs1042151 SNP is located in an exonic splicing enhancer region; however, despite the location of this SNP, its true effect remains questionable, as rs1042151 alone exhibits very low accuracy (54.7%) and sensitivity (17.7%) for predicting NERD, despite strong specificity (91.7%). Furthermore, because this GWAS was limited to common variants (MAF>5%) with minimal coverage of exonic SNPs, a second NERD-associated SNP study was extended to exonic SNPs using an Exome BeadChip assay containing 240K SNP. In this analysis, another exonic SNP (exm537513

![Fig. 2. ROC curves used to devise the best combinatorial model using the 10 most statistically significant SNPs. Values were taken from the first genome-wide association study, which used 100K Bead Chips, to examine 80 NERD and 100 ATA subjects. RR was calculated for each individual subject using a multiple logistic regression analysis examining all 1023 combinations ($2^{10}-1$) of the 10 SNPs described in reference 34. A model consisting of eight SNPs shows the highest area under the ROC curve of 0.9 with an accuracy of 82.01%. The sensitivity and specificity are 78% and 88%, respectively, with an odds ratio of 20.74 ($p=3.24E^{-19}$). The eight SNPs are listed in Table 2. Reprinted from Shin, et al. DNA Cell Biol 2012;31:1604-9, with permission of Mary Ann Liebert, Inc.]

### Table 2. List of the Eight SNPs in Fig. 2 and their RRs for Diagnosis of Nonsteroidal Anti-Inflammatory Drug-Exacerbated Respiratory Diseases

| SNP       | Gene              | Position | Alleles   | $p$ value | OR     | RR    | CC  | CM  | MM  |
|-----------|-------------------|----------|-----------|-----------|--------|-------|-----|-----|-----|
| rs6911768 | PARK2             | Intron   | A>C       | 6.75E-05  | 2.901  | 0.43  | 1.248|     |     |
| rs3213729 | WDR21             | 5’ UTR   | C>G       | 0.000411  | 2.724  | 0.507 | 1.382| 3.764|
| rs6714952 | VRK2              | Intron   | C>T       | 0.000723  | 2.486  | 0.526 | 1.307| 3.248|
| rs746035  | CHST11            | Intron   | C>T       | 0.00532   | 2.292  | 0.492 | 1.127| 2.583|
| rs4501026 | LINCO0314         | Intergenic| A>C      | 4.76E-05  | 0.4198 | 1.44  | 0.604| 0.254|
| rs4867084 | PDZK3             | Intron   | G>A       | 0.00017   | 0.4177 | 1.269 | 0.53 | 0.221|
| rs139719  | SGSM1             | Intron   | C>T       | 3.18E-05  | 0.4109 | 1.416 | 0.582| 0.239|
| rs6659655 | CENPF             | Intergenic| A>G      | 3.49E-05  | 0.3181 | 1.159 | 0.369| 0.117|

SNP, single nucleotide polymorphism; RR, relative risk; OR, odds ratio; PARD2, Parkin RBR E3 Ubiquitin Protein Ligase; WDR21, DDB1 and CUL4 Associated Factor 4; VRK2, Vaccinia Related Kinase 2; CHST11, Carbohydrate Sulfotransferase 11; LINCO0314, Long Intergenic Non-Protein Coding RNA 314; PDZK3, PDZ Domain Containing 2; SGSM1, Small G Protein Signaling Modulator 1; CENPF, Centromere Protein F. Reprinted from Shin, et al. DNA Cell Biol 2012;31:1604-9, with permission of Mary Ann Liebert, Inc.
**Table 3. List of the Top 30 SNPs Associated with NERD**

| n  | Name | Position | Genes in EXON | Mutation(s) | MAF (NERD, n=165) | OR (95% CI) | p value  |
|----|------|----------|---------------|-------------|-------------------|-------------|---------|
| 1  | rs1042136 | 6 | HLA-DPB1,*** | A>C | 0.167 | 3.28 (2.13–5.04) | 3.40E-08 |
| 2  | rs1042151 | 6 | HLA-DPB1,*** | A>G | 0.194 | 2.50 (1.71–3.67) | 3.18E-06 |
| 3  | rs1042157 | 6 | HLA-DPB1,*** | G>C | 0.205 | 2.37 (1.68–3.35) | 4.62E-06 |
| 4  | rs1042162 | 6 | HLA-DPB1,*** | T>C | 0.197 | 2.36 (1.70–3.33) | 6.00E-06 |
| 5  | rs1042165 | 6 | HLA-DPB1,*** | C>G | 0.185 | 2.28 (1.70–3.34) | 6.26E-06 |
| 6  | rs1042169 | 6 | HLA-DPB1,*** | G>A | 0.216 | 2.24 (1.70–3.34) | 6.68E-06 |
| 7  | rs1042171 | 6 | HLA-DPB1,*** | G>A | 0.205 | 2.23 (1.70–3.34) | 6.75E-06 |
| 8  | rs1042173 | 6 | HLA-DPB1,*** | C>G | 0.185 | 2.21 (1.70–3.34) | 6.89E-06 |
| 9  | rs1042174 | 6 | HLA-DPB1,*** | T>C | 0.197 | 2.20 (1.70–3.34) | 6.96E-06 |
| 10 | rs1042175 | 6 | HLA-DPB1,*** | C>G | 0.185 | 2.19 (1.70–3.34) | 6.99E-06 |
| 11 | rs1042176 | 6 | HLA-DPB1,*** | G>A | 0.205 | 2.18 (1.70–3.34) | 7.02E-06 |
| 12 | rs1042177 | 6 | HLA-DPB1,*** | G>A | 0.205 | 2.17 (1.70–3.34) | 7.05E-06 |
| 13 | rs1042178 | 6 | HLA-DPB1,*** | C>G | 0.185 | 2.16 (1.70–3.34) | 7.08E-06 |
| 14 | rs1042179 | 6 | HLA-DPB1,*** | T>C | 0.197 | 2.15 (1.70–3.34) | 7.11E-06 |
| 15 | rs1042180 | 6 | HLA-DPB1,*** | C>G | 0.185 | 2.14 (1.70–3.34) | 7.14E-06 |
| 16 | rs1042181 | 6 | HLA-DPB1,*** | G>A | 0.205 | 2.13 (1.70–3.34) | 7.17E-06 |
| 17 | rs1042182 | 6 | HLA-DPB1,*** | C>G | 0.185 | 2.12 (1.70–3.34) | 7.20E-06 |
| 18 | rs1042183 | 6 | HLA-DPB1,*** | T>C | 0.197 | 2.11 (1.70–3.34) | 7.23E-06 |
| 19 | rs1042184 | 6 | HLA-DPB1,*** | C>G | 0.185 | 2.10 (1.70–3.34) | 7.26E-06 |
| 20 | rs1042185 | 6 | HLA-DPB1,*** | G>A | 0.205 | 2.09 (1.70–3.34) | 7.29E-06 |
| 21 | rs1042186 | 6 | HLA-DPB1,*** | C>G | 0.185 | 2.08 (1.70–3.34) | 7.32E-06 |
| 22 | rs1042187 | 6 | HLA-DPB1,*** | T>C | 0.197 | 2.07 (1.70–3.34) | 7.35E-06 |
| 23 | rs1042188 | 6 | HLA-DPB1,*** | C>G | 0.185 | 2.06 (1.70–3.34) | 7.38E-06 |
| 24 | rs1042189 | 6 | HLA-DPB1,*** | T>C | 0.197 | 2.05 (1.70–3.34) | 7.41E-06 |
| 25 | rs1042190 | 6 | HLA-DPB1,*** | C>G | 0.185 | 2.04 (1.70–3.34) | 7.44E-06 |
| 26 | rs1042191 | 6 | HLA-DPB1,*** | T>C | 0.197 | 2.03 (1.70–3.34) | 7.47E-06 |
| 27 | rs1042192 | 6 | HLA-DPB1,*** | C>G | 0.185 | 2.02 (1.70–3.34) | 7.50E-06 |
| 28 | rs1042193 | 6 | HLA-DPB1,*** | T>C | 0.197 | 2.01 (1.70–3.34) | 7.53E-06 |
| 29 | rs1042194 | 6 | HLA-DPB1,*** | C>G | 0.185 | 2.00 (1.70–3.34) | 7.56E-06 |
| 30 | rs1042195 | 6 | HLA-DPB1,*** | T>C | 0.197 | 1.99 (1.70–3.34) | 7.59E-06 |

NERD, nonsteroidal antiinflammatory drug-exacerbated respiratory diseases; MAF, minor allele frequency; ATA, aspirin-tolerant asthma; SNP, single nucleotide polymorphism; AUC, area under the curve; OR, odds ratio.
in HLA-DPB1, rs1042136) showed the lowest p value ($p=3.4\times10^{-4}$, OR: 3.28) in association with the risk of NERD, although with low sensitivity (16.7%) and accuracy (55.5%) as before, in spite of relatively high specificity (94.3%). These data suggest that SNPs on HLA-DPB1 should contribute to the development of NERD, but are not clinically useful as genetic markers of NERD.

In the exon-wide association study, 16 of the top 20 SNPs were located on chromosome 6 (Table 3). Among them, six SNPs were exonic and located on HLA-DPB1 [exm537513 (rs1042136, 194L), exm537522 (re1042151, M105V), and exm537523 (rs1042153, M105I)], HLA-DPA1 [exm-rs3097671 (silent) and exm537485 (rs1126504, L327V)], and HLA-DPB2 [exm-rs3129294 (silent)]. Furthermore, a strong linkage disequilibrium (LD) was noted between the six exonic SNPs in the HLA genes (Fig. 3). AUCs for the combinations of these SNPs were exonic and located on HLA-DPB1 [exm537513, exm537522, exm-rs3097671, and exm537485] showed the best AUC of 0.75 ($p=7.94\times10^{-21}$), with OR=7.291, 40.8% sensitivity, 88.0% specificity, and 75% accuracy for the discrimination of NERD from ATA. To date, 2161 SNPs in HLA-DPB1 and 2602 SNPs in HLA-DPA1 have been identified (http://www.ncbi.nlm.nih.gov/snp/?term=HLA-DPB1 and http://www.ncbi.nlm.nih.gov/snp/?term=HLA-DPA1), of which 239 and 193 exonic SNPs are present in HLA-DPB1 and HLA-DPA1, respectively. This indicates that the number of SNPs within the 660K GWAS and 240K exon chips used in our and other studies is still not sufficient to discover all of the SNPs associated with NERD. Thus, further discovery of SNPs in HLA-DPB1 and HLA-DPA1 will be necessary for the development of diagnostic genetic markers for NERD.

Our third combinatorial analysis of SNPs was performed with the 660W genome-wide association data examining 142 NERD and 996 ATA.24 Fourteen genes [HLA-DPB1, MEIS1, NAB1, NRNP2, CLDN1, DCCD2, TMEM196, AK8 (C9orf98), TSC1, ZFYVE1-RBM25, PCSK6, ILVBL, GP6, and RDH13] were selected on the basis of OR and p values (Table 4). In this study, summed genetic risk scores were calculated as a combinatorial analysis and used in place of OR.25 To exclude the effect of LD, one SNP on each of the 14 genes was chosen, and the genetic risk score of each of the subjects was calculated by assigning ORs to the risk alleles in comparison to reference alleles. This was followed by summing the ORs from multiple variants in different genetic loci into a single RR value using the following formula:

$$\text{Summed risk score} = \exp \left( \sum_{i=1}^{n} \ln(\text{odds ratio} \times \text{risk allele}) \right)$$

The summed risk scores of the 14 SNPs ranged from 0.774 to 1.142, with an average of 0.973. These summed risk scores were significantly higher in NERD patients than in ATA con-

### Table 4. List of the 14 Top SNPs Predictive of NERD from the 660K GWAS

| SNPID | Chr. | Gene (nearest gene) | Location | Odds ratio (95% CI) | Results of GWAS |
|-------|------|---------------------|----------|---------------------|-----------------|
| rs12619205  | 2     | MEIS1               | Intron   | 1.72 (1.30–2.28)    | 1.14×10^{-4}    |
| rs10931468   | 2     | NAB1                | Intron   | 1.45 (1.09–1.93)    | 9.94×10^{-3}    |
| rs849530    | 2     | NRNP2               | Intron   | 1.8 (1.35–2.41)     | 6.91×10^{-2}    |
| rs10513846   | 3     | CLDN1               | Intron   | 0.68 (0.51–0.91)    | 8.15×10^{-3}    |
| rs2281389    | 3     | HLA-DPB1            | 3'       | 2.4 (1.61–3.58)     | 5.6×10^{-4}     |
| rs3789224    | 6     | DCCD2               | Intron   | 1.98 (1.41–2.77)    | 8.55×10^{-3}    |
| rs9886152    | 7     | TMEM196             | Intron   | 0.54 (0.39–0.76)    | 3.09×10^{-4}    |
| rs2771994    | 9     | AK8 (C9orf98)       | Intron   | 1.87 (1.39–2.52)    | 2.48×10^{-3}    |
| rs1050700    | 9     | TSC1                | 3'UTR    | 1.82 (1.35–2.45)    | 5.82×10^{-4}    |
| rs12432987   | 14    | ZFYVE1-RBM25        | Intron   | 1.88 (1.38–2.57)    | 4.36×10^{-3}    |
| rs3825915    | 15    | PCSK6               | Intron   | 0.51 (0.37–0.70)    | 2.5×10^{-4}     |
| rs2260299    | 19    | ILVBL               | Intron   | 0.51 (0.37–0.72)    | 7.61×10^{-5}    |
| rs11669150   | 19    | GP5                 | Intron   | 0.68 (0.47–0.98)    | 3.74×10^{-2}    |
| rs1671215    | 19    | RDH13               | 3'       | 0.49 (0.32–0.74)    | 8.34×10^{-4}    |

SNP, single nucleotide polymorphism; NERD, nonsteroidal antiinflammatory drug-exacerbated respiratory diseases; GWAS, genome-wide association studies; CI, confidence interval.

Using the 14 SNPs listed in the table, combinatorial analysis was performed, and the results are summarized in Table 5. Reprinted from Chang, et al. Pharmacogenomics J 2015;15:316-21, with permission of Springer Nature.26
trols (1.033±0.005 vs. 0.963±0.002, \(p=8.58\times10^{-37}\)). The AUC of the NERD score was 0.821. The best cutoff value with a score of 1.01328 was determined using the ROC curve and the Youden’s Index. A higher proportion of NERD subjects showed summed risk scores above the cutoff value relative to ATA subjects [64.66% vs. 15.04%, \(p=1.38\times10^{-15}\), OR=10.34 (6.69–15.96)]. The sensitivity and specificity of the scores were 64.7% and 85.0%, respectively, with an accuracy of 82.0% (42.1% positive and 93.4% negative predictive values) (Table 5).

**TABLE 5. Diagnostic Value of the Summed Risk Scores Using the Top 14 SNPs to Predict NERD**

| RR score | ATA | NERD | Total | Sensitivity | Specificity | Accuracy |
|----------|-----|------|-------|-------------|-------------|----------|
| ≤1.01328 | 582 (72.7) | 41 (5.1) | 623 (77.8) | 64.70% | 85.00% | 82.00% |
| >1.01328 | 103 (12.9) | 75 (9.4) | 178 (22.2) | 64.70% | 85.00% | 82.00% |
| Total | 685 (85.5) | 116 (14.5) | 801 (100) | 64.70% | 85.00% | 82.00% |

SNP, single nucleotide polymorphism; NERD, nonsteroidal antiinflammatory drug-exacerbated respiratory diseases; ATA, aspirin-tolerant asthma; \(\chi^2\), positive predictive value; \(\chi^2\), negative predictive value; LR+, positive likelihood ratio; LR-, negative likelihood ratio; OR, odds ratio; RR, relative risk.

To calculate the genetic risk score of each of the subjects, ORs were assigned to the risk alleles in contrast to the value of 1 assigned to reference alleles. ORs from multiple variants in different genetic loci were summed into a single RR value. The summed risk scores of the 14 SNPs in the study subjects ranged from 1.01328 to 1.142. The summed risk scores were significantly higher in the NERD group than in the ATA group (1.033±0.005 vs. 0.963±0.002, \(p=8.58\times10^{-37}\)). A higher proportion of NERD subjects showed summed risk scores above the cutoff value [64.66% vs. 15.04%, \(p=1.38\times10^{-15}\), OR=10.34 (6.69–15.96)]. The sensitivity and specificity of the scores were 64.7% and 85.0%, respectively, with 42.1% positive and 93.4% negative predictive values. The overall accuracy of the test was 82.0%. Reprinted from Chang, et al. Pharmacogenomics J 2015;15:316-21, with permission of Springer Nature.37

**Table 6. Diagnostic Accuracy of SNPs Identified in GWAS and Exon-Wide Association Studies for NERD**

| Genetic makers | OR | Accuracy (%) | Sensitivity (%) | Specificity (%) | Reference |
|----------------|----|--------------|----------------|----------------|----------|
| 1) 7572857 G>A on CEP68 | 2.63 | 53 | 18 | 88 | 32 |
| 2) rs1042151 (Met105Val) in exon 2 of HLA-DPB1 obtained from a 660K GWAS | 2.40 | 54.7 | 17.7 | 91.7 | 29 |
| 3) rs1042136 (exm537513) in HLA-DPB1 obtained from a 240K exome chip | 3.28 | 55.5 | 16.7 | 94.3 | 39 |
| 4) A combination model of the eight SNPs obtained from a 100K GWAS \* | 20.74 | 82.01 | 78 | 88 | 34 |
| 5) Four exonic SNPs on HLA-DPB1 obtained from a 240K exome chip | 3.67 | 68.4 | 29.8 | 90.3 | 36 |
| 6) Seven exonic SNPs obtained from a 240K exome chip | 7.29 | 75 | 40.8 | 88 | 36 |
| 7) A summed risk score of 14 SNPs obtained from 660K GWAS | 10.34 | 82 | 64.7 | 85 | 37 |

SNP, single nucleotide polymorphism; GWAS, genome-wide association studies; NERD, nonsteroidal antiinflammatory drug-exacerbated respiratory diseases; OR, odds ratio.

\*A combination set of eight SNPs listed in Table 2, \*A combination of four exonic SNPs (exm537513, exm537522, exm-rs3097671, and exm-rs3129294) listed in Table 3, \*A combination model of seven SNPs (exm537513, exm83523, exm1884673, exm538564, exm2264237, exm396794, and exm791954) listed in Table 3, \*A combination of eight SNPs obtained from a 100K GWAS and the summed risk score of 14 SNPs obtained from 660K GWAS showed higher accuracy (82.0% and 82.01%), compared with that observed for the 240K exome chip (68.4% and 75.0%). The improved accuracy of GWAS in these studies was derived from the higher sensitivity of the method (64.7% and 78.0%), compared to that seen using exome-wide methods (29.8% and 40.8%), despite equal specificities of ~90%.

**COMPARISON OF THE GENETIC MARKERS IN DIAGNOSIS OF NERD**

The three most significant SNPs identified in the GWAS and exome chip studies exhibited ORs of only 2.4–3.28, with accuracies of 53.0–55.5% (Table 6). Although the specificity of these markers was very high, nearing 90% in many cases, the sensitivity of these markers was consistently low, typically at less than 20%. In contrast, both the OR and accuracy were elevated using combinatorial analysis of the top 10 SNPs obtained from the GWAS. A combinatorial analysis of several SNPs from GWAS and exome-wide association studies provided a more comprehensive genetic set for the diagnosis of NERD, compared to that of single SNPs alone. In the data presented here, the combination of eight SNPs obtained from a 100K GWAS and the summed risk score of 14 SNPs obtained from 660K GWAS showed higher accuracy (82.0% and 82.01%), compared with that observed for the 240K exome chip (68.4% and 75.0%). The improved accuracy of GWAS in these studies was derived from the higher sensitivity of the method (64.7% and 78.0%), compared to that seen using exome-wide methods (29.8% and 40.8%), despite equal specificities of ~90%.

**PERSPECTIVES FOR IMPROVED DIAGNOSTIC VALUES USING COMBINATORIAL ANALYSIS OF MULTIPLE SNPS**

GWAS have been used to identify numerous loci associated with multifactorial complex diseases and quantitative traits. However, most GWAS have not studied the gene-gene interactions (epistasis) that could be important in complex trait genetics and in combinatorial analysis of the SNPs. Existing statistical techniques, such as logistic regression and multifactor dimensionality reduction, can be used to identify the combinatorial effects of multiple SNPs in disease susceptibility; however, methods for the analysis of these techniques remains computationally intensive. To overcome some of these
computational bottlenecks, a series of methods have been developed, including the use of computers equipped with particular graphical processing units, such as the BiForce Toolbox (University of Turku, Turku, Finland). The BiForce Toolbox incorporates additional tests of interactions involving SNPs with significant marginal effects and identifies interesting interaction signals and pathways.\textsuperscript{38} Pathway-based approaches may also narrow the search space and enhance power, for example by seeking pathways\textsuperscript{39} or by identifying common pathways enriched in epistatic genes with modest interaction signals detected from multiple GWAS populations.\textsuperscript{40} In either case, rapid screening of pairwise interactions in individual GWAS populations appears to be critical in order to provide the information necessary to design and perform pathway-based analyses.

Another limitation is the relatively small sizes used in these studies, thereby limiting the ability to replicate significant epistasis signals.\textsuperscript{41} A recently proposed statistical method, called Limitless Arity Multiple-testing Procedure (LAMP),\textsuperscript{42} provides the possibility of detecting statistically significant higher-order interactions. LAMP is a multiple testing procedure for identifying statistically significant combinatorial effects by introducing a theoretical upper bound for a family-wise error rate tighter than a Bonferroni correction. Its application to GWAS may uncover synergistic effects of SNPs associated with diseases, while filtering out uninformative combinations. Another program, LAMPLINK (National Institute of Advanced Industrial Science and Technology, Tokyo, Japan), also has the ability to quickly identify combinatorial effects of SNPs.

Outside of computational methodologies, an important factor complicating the diagnosis of NERD is the low overall penetration of genetic traits. If the genomically susceptible subjects do not take aspirin or NSAID in their lifetime, they will never develop NERD or aspirin-induced urticaria. Thus, a history of aspirin or NSAID use should be available at the time of diagnosis, although the exact dosage and duration of drug exposure remains difficult to estimate for both patients and physicians. In the future, better combinatorial data will be possible as more information related to the measurable exposome becomes available and incorporated into big data-type analyses, including factors such as lifestyle, prescription drug use, and exposure to air pollution.

In the effort to identify SNP markers, GWAS has provided useful information for asthma. Understandably, however, identified SNP markers alone show poor discriminating power for diagnosis, mainly due to biological complexity and phenotypic heterogeneity of asthma. Research to identify rare variants in asthma not only seems to be unsuccessful at discovering genetic markers, but also fails to provide meaningful information on the missing heritability of the disease. In contrast, studies using sub-phenotypes and combination/integration of SNPs seem to be more successful in achieving clinically useful diagnostics. In the near future, it is expected that statistical methods providing improved sensitivity and specificity could be achieved by the combination and integration of common genetic variants with exposome data for individual patients.

**ACKNOWLEDGEMENTS**

This study was supported by the Ministry of Health, Welfare and Family Affairs, Republic of Korea (HI13C0319) and a research grant from Soonchunhyang University to JS Park.

**ORCID iDs**

Jong-Sook Park https://orcid.org/0000-0003-2977-0255
Ji-Hye Son https://orcid.org/0000-0002-5380-1245
Choon-Sik Park https://orcid.org/0000-0003-2977-0255
Hun Soo Chang https://orcid.org/0000-0002-4025-4182

**REFERENCES**

1. Edfors-Lubs ML. Allergy in 7000 twin pairs. Acta Allergol 1971;26:249-85.
2. Van Eerdewegh P, Little RD, Dupuis J, Del Mastro RG, Falls K, Simon J, et al. Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness. Nature 2002;418:426-30.
3. Allen M, Heinzmanna A, Noguchia E, Abecasis G, Broxholmec J, Ponting cp, et al. Positional cloning of a novel gene influencing asthma from chromosome 2q14. Nat Genet 2003;35:258-63.
4. Zhang Y, Leavesc NI, Anderson GG, Broxholme J, Holt R, et al. Positional cloning of a quantitative trait locus on chromosome 13q14 that influences immunoglobulin E levels and asthma. Nat Genet 2003;34:181-6.
5. Laitinnen T, Polvia A, Rydman P, Vendelin J, Pullkinenv N, Salmikan-gas P, et al. Characterization of a common susceptibility locus for asthma-related traits. Science 2004;304:300-4.
6. Oguma T, Palmer LJ, Birben E, Sonna LA, Asanok L, Lilly CM. Role of prostanoid DP receptor variants in susceptibility to asthma. N Engl J Med 2004;351:1752-63.
7. Hardy J, Singleton A. Genomewide association studies and human disease. N Engl J Med 2009;360:1759-68.
8. Moffat MF, Gut IG, Denemais F, Strachan DP, Bouzigon E, Heath S, et al. A large-scale, consortium-based genomewide association study of asthma. N Engl J Med 2010;363:2121-21.
9. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunt er DJ, et al. Finding the missing heritability of complex diseases. Nature 2009;461:747-53.
10. 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, et al. A map of human genome variation from population-scale sequencing. Nature 2010;467:1061-73.
11. Igartua C, Myers RA, Mathias RA, Pino-Yanes M, Eng C, Graves PE, et al. Ethnic-specific associations of rare and low-frequency DNA sequence variants with asthma. Nat Commun 2015;6:5965.
12. Green RH, Brightling CE, Bradding P. The reclassification of asthma based on subphenotypes. Curr Opin Allergy Clin Immunol 2007;7:43-50.
13. Lee SH, Park JS, Park CS. The search for genetic variants and epigenetics related to asthma. Allergy Asthma Immunol Res 2011;3:236-44.
14. Moore WC, Meyers DA, Wenzel SE, Teague WG, Li H, Li X, et al.
Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program. Am J Respir Crit Care Med 2010;181:315-23.

15. Jang AS, Kwon HS, Cho YS, Bae YJ, Kim TB, Park JS, et al. Identification of subtypes of refractory asthma in Korean patients by cluster analysis. Lung 2013;191:87-93.

16. Lee JH, Lee HS, Park MR, Lee SW, Kim EH, Cho JB, et al. Relationship between indoor air pollutant levels and residential environment in children with atopic dermatitis. Allergy Asthma Immunol Res 2014;6:517-24.

17. Jeong I, Kim I, Park HJ, Roh J, Park JW, Lee JH. Allergic diseases and multiple chemical sensitivity in Korean adults. Allergy Asthma Immunol Res 2014;6:409-14.

18. Hur GY, Ye YM, Koh DH, Kim SH, Park HS. IL-4 receptor α polymorphisms may be a susceptible factor for work-related respiratory symptoms in bakery workers. Allergy Asthma Immunol Res 2013;5:371-6.

19. Kim SH, Sutherland ER, Gelfand EW. Is there a link between obesity and asthma? Allergy Asthma Immunol Res 2014;6:189-95.

20. Lee SY, Kang MJ, Kwon JW, Park KS, Hong SJ. Breastfeeding might have protective effects on atopy in children with the CD14: -159T CT/CC genotype. Allergy Asthma Immunol Res 2013;5:239-41.

21. Huiyan W, Yuhe G, Juan W, Junyan Z, Shan W, Xiaojun Z, et al. The importance of allergen avoidance in high risk infants and sensitized patients: a meta-analysis study. Allergy Asthma Immunol Res 2014;6:525-34.

22. Peden D, Reed CE. Environmental and occupational allergies. J Allergy Clin Immunol 2010;125(2 Suppl 2):S150-60.

23. von Mutius E. Gene-environment interactions in asthma. J Allergy Clin Immunol 2009;123:3-11.

24. Park HJ, Lim HS, Park KH, Lee JH, Park JW, Hong CS. Changes in allergen sensitization over the last 30 years in Korean respiratory allergic patients: a single-center. Allergy Asthma Immunol Res 2014;6:434-43.

25. Kim SH, Cho BY, Park CS, Shin ES, Cho EY, Yang EM, et al. Alpha-T-catenin (CTNNA3) gene was identified as a risk variant for toluene diisocyanate-induced asthma by genome-wide association analysis. Clin Exp Allergy 2009;39:203-12.

26. Bernstein DI, Kashon M, Lummus ZL, Johnson VJ, Fluharty K, Gautrin D, et al. CTNNA3 (α-catenin) gene variants are associated with diisocyanate asthma: a replication study in a Caucasian worker population. Toxicol Sci 2013;131:242-6.

27. Park SM, Park JS, Park HS, Park CS. Unraveling the genetic basis of aspirin hypersensitivity in asthma beyond arachidionate pathways. Allergy Asthma Immunol Res 2013;5:258-76.

28. Park BL, Kim TH, Kim JH, Bae JS, Pasaje CF, Cheong HS, et al. Genome-wide association study of aspirin-exacerbated respiratory disease in a Korean population. Hum Genet 2013;132:313-21.

29. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010;21:128-38.

30. Jakobsdottir J, Gorin MB, Conley YP, Ferrell RE, Weeks DE. Interpretation of genetic association studies: markers with replicated highly significant odds ratios may be poor classifiers. PLoS Genet 2009;5:e1000337.

31. Kim JH, Park BL, Cheong HS, Bae JS, Park JS, Jang AS, et al. Genome-wide and follow-up studies identify CEP68 gene variants associated with risk of aspirin-intolerant asthma. PLoS One 2010;5:e13818.

32. Chang HS, Park JS, Jang AS, Park SW, Uh ST, Kim YH, et al. Diagnostic value of clinical parameters in the prediction of aspirin-exacerbated respiratory disease in asthma. Allergy Asthma Immunol Res 2013;5:256-64.

33. Wei WH, Hemani G, Haley CS. Detecting epistasis in human complex traits. Nat Rev Genet 2014;15:722-33.

34. Shin SW, Park J, Kim YJ, Uh ST, Choi BW, Kim MK, et al. A highly sensitive and specific genetic marker to diagnose aspirin-exacerbated respiratory disease using a genome-wide association study. DNA Cell Biol 2012;31:1604-9.

35. McCarthy MI, Hirschhorn JN. Genome-wide association studies: potential next steps on a genetic journey. Hum Mol Genet 2008;17:R156-65.

36. Shin SW, Park BL, Chang H, Park JS, Bae DJ, Song HJ, et al. Exonic variants associated with development of aspirin exacerbated respiratory diseases. PLoS One 2014;9:e111887.

37. Chang HS, Shin SW, Lee TH, Bae DJ, Park JS, Kim YH, et al. Development of a genetic marker set to diagnose aspirin-exacerbated respiratory disease in a genome-wide association study. Pharmacogenomics J 2015;15:316-21.

38. Gauderman WJ. Sample size requirements for association studies of gene-gene interaction. Am J Epidemiol 2002;155:478-84.

39. Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: Genetic interactions create phantom heritability. Proc Natl Acad Sci U S A 2012;109:1193-8.

40. Wei WH, Hemani G, Gyneselai A, Vitart V, Navarro P, Hayward C, et al. Genome-wide analysis of epistasis in body mass index using multiple human populations. Eur J Hum Genet 2012;20:857-62.

41. Gauderman WJ. Sample size requirements for association studies of gene-gene interaction. Am J Epidemiol 2002;155:478-84.

42. Terada A, Okada-Hatakeyama M, Tsuda K, Sese J. Statistical significance of combinatorial regulations. Proc Natl Acad Sci U S A 2013;110:12996-3001.