Genome-wide association study of recalcitrant atopic dermatitis in Korean children

Kyung Won Kim, MD, PhD,[a,b] Rachel A. Myers, PhD,[b] Ji Hyun Lee, PhD,[c] Catherine Igartua, BS,[b] Kyung Eun Lee, PhD,[a] Yoon Hee Kim, MD,[a] Eun-Jin Kim, PhD,[d] Dankyu Yoon, PhD,[d] Joo-Shil Lee, PhD,[d] Tomomitsu Hirota, DDS, PhD,[e] Mayumi Tamari, MD, PhD,[e] Atsushi Takahashi, PhD,[f] Michiaki Kubo, MD, PhD,[g] Je-Min Choi, PhD,[h] Kyu-Earn Kim, MD, PhD,[a] Dan L. Nicolae, PhD,[b,i] Carole Ober, PhD,[b] and Myung Hyun Sohn, MD, PhD[a] *Seoul and Osong, Korea, Chicago, Ill, and Yokohama and Tokyo, Japan*

Background: Atopic dermatitis (AD) is a heterogeneous chronic inflammatory skin disease. Most AD during infancy resolves during childhood, but moderate-to-severe AD with allergic sensitization is more likely to persist into adulthood and more often occurs with other allergic diseases.

Objective: We sought to find susceptibility loci by performing the first genome-wide association study (GWAS) of AD in Korean children with recalcitrant AD, which was defined as moderate-to-severe AD with allergic sensitization.

Methods: Our study included 246 children with recalcitrant AD and 551 adult control subjects with a negative history of both allergic disease and allergic sensitization. DNA from these subjects was genotyped; sets of common single nucleotide polymorphisms (SNPs) were imputed and used in the GWAS after quality control checks.

Results: SNPs at a region on 13q21.31 were associated with recalcitrant AD at a genome-wide threshold of significance ($P < 2.0 \times 10^{-8}$). These associated SNPs are more than 1 Mb from the closest gene, protocadherin (*PCDH*)9. SNPs at 4 additional loci had P values of less than 1×10^{-6}, including SNPs at or near the neuroblastoma amplified sequence (*NBAS*; 2p24.3), thymus-expressed molecule involved in selection (*THEMIS*; 6q22.33), *GATA3* (10p14), and S-phase cyclin A-associated protein in the ER (*SCAPER*; 15q24.3) genes. Further analysis of total serum IgE levels suggested 13q21.31 might be primarily an IgE locus, and analyses of published data demonstrated that SNPs at the 15q24.3 region are expression quantitative trait loci for 2 nearby genes, *ISL2* and proline-serine-threonine phosphatase interacting protein 1 (*PSTPIP1*), in immune cells.

Conclusion: Our GWAS of recalcitrant AD identified new susceptibility regions containing genes involved in epithelial cell function and immune dysregulation, 2 key features of AD, and potentially extend our understanding of their role in pathogenesis. (J Allergy Clin Immunol 2015;136:678-84.)

Key words: *Genome-wide association study, atopic dermatitis, allergic sensitization, IgE, severity, children*

Atopic dermatitis (AD) is a complex chronic inflammatory skin disease that commonly presents during childhood, when it is strongly associated with allergic sensitization.[1] Although AD has a varied disease course, children with moderate-to-severe AD with allergic sensitization are more likely to have disease persisting to adulthood and more concomitant allergic diseases, such as asthma or allergic rhinitis, which result in significant health care costs.[2] Available treatment options for the prevention and treatment of this subtype of recalcitrant AD are still insufficient,[3] reflecting our poor understanding of disease pathogenesis.

AD is highly heritable, with heritability estimates of 72% in European twin pairs[4,5] and genetic studies supporting a significant role for aberrant gene expression in patients with AD.[6] In particular, low-frequency and rare loss-of-function variants in the filaggrin gene *(FLG)* are major predisposing factors for persistent AD, as well as for skin infections with AD and multiple allergic diseases.[7-9] Filaggrin deficiency results in skin barrier dysfunction, resulting in accelerated water loss, skin alkalinization, and colonization by microbial pathogens.[10] Based on these findings, epithelial barrier dysfunction (in particular filaggrin) has been placed at the center of AD pathogenesis.

Three large genome-wide association studies (GWASs),[11-13] a meta-analysis of GWASs from 16 population-based European cohorts,[14] and targeted studies using the immunochip[15] have identified several candidate AD genes in addition to *FLG*. However, none reached genome-wide significance in the discovery samples, and moreover, there were no shared loci among the top associations in studies of European and Han Chinese AD populations.[11,12] These combined results suggest genetic heterogeneity in AD between continental populations, particularly when broad case definitions are used. AD is characterized by genetic and

From [a]the Department of Pediatrics, Severance Hospital, Institute of Allergy, Brain Korea 21 PLUS project for Medical Science, Yonsei University College of Medicine, Seoul; the Departments of [b]Human Genetics and [i]Medicine and Statistics, University of Chicago; [c]the Department of Oral Biology, Yonsei University College of Dentistry, Seoul; [d]the Division of Allergy and Chronic Respiratory Diseases, Center for Biomedical Sciences, Korea National Institute of Health, Osong; [e]the Laboratory for Respiratory and Allergic Diseases and [g]the Laboratory for Genotyping Development, Center for Integrative Medical Sciences, RIKEN, Yokohama; [f]the Laboratory for Statistical Analysis, Center for Integrative Medical Sciences, RIKEN, Tokyo; [h]the Department of Life Science, Research Institute for Natural Sciences, Hanyang University, Seoul.

Supported by the grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI) funded by the Ministry of Health & Welfare, Republic of Korea (grant no. HI11C1404, HI14C0234, and A092076), a National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIP; no. 2007-0056092), the Korea Research Foundation Grant funded by the Korean Government (KRF-2010-0025171), and National Institutes of Health grants U19 AI095230 and R01 HL085197 (to C.O.). This study included biospecimens and data from the Korean Genome Analysis Project (4845-301), the Korean Genome and Epidemiology Study (4851-302), and the Korea Biobank Project (4851-307, KBP-2014-033), which were supported by the Korea Center for Disease Control and Prevention, Republic of Korea.

Disclosure of potential conflict of interest: The authors declare that they have no relevant conflicts of interest.

Received for publication January 15, 2015; revised March 5, 2015; accepted for publication March 13, 2015.

Available online April 30, 2015.

Corresponding author: Myung Hyun Sohn, MD, PhD, Department of Pediatrics, Institute of Allergy, Yonsei University College of Medicine, 50 Yonsei-ro, Seodaemun-gu, Seoul 120-752, Republic of Korea. E-mail: mhsohn@yuhs.ac.

0091-6749/$36.00

http://dx.doi.org/10.1016/j.jaci.2015.03.030

Abbreviations used
AD: Atopic dermatitis
eQTL: Expression quantitative trait locus
FLG: Filaggrin gene
GWAS: Genome-wide association study
LD: Linkage disequilibrium
NBAS: Neuroblastoma amplified sequence
PCDH: Protocadherin
PSTPIP1: Proline-serine-threonine phosphatase interacting protein 1
SCAPER: S-phase cyclin A-associated protein in the ER
SNP: Single nucleotide polymorphism
THEMIS: Thymus-expressed molecule involved in selection

phenotypic heterogeneity, and this is consistent with the finding that susceptibility loci discovered to date, including *FLG*, account for only 14.4% of the heritability of AD in Europeans.[15] This also suggests that studies in additional populations and with narrower clinical definitions are needed to fully characterize the genetic architecture of AD.

Here we conducted the first GWAS of AD in Korean children and focused on the distinct phenotype of recalcitrant AD, which is defined as moderate-to-severe AD with allergic sensitization. Moreover, we included as control subjects nonallergic adults without a history of allergic diseases. We identified a novel region of chromosome 13q21.31 as likely to contain genes controlling AD risk, which was genome-wide significant, with additional loci, including neuroblastoma amplified sequence *(NBAS)*, thymus-expressed molecule involved in selection *(THEMIS)*, *GATA3*, and S-phase cyclin A-associated protein in the ER (*SCAPER*), as suggested genes for AD.

METHODS

Sample compositions

Our study included 246 Korean children with both moderate-to-severe AD and allergic sensitization and 551 Korean adult control subjects without a history of allergic diseases or evidence of allergic sensitization. In addition, because IgE levels vary with age, we performed association studies of selected single nucleotide polymorphisms (SNPs) with total serum IgE levels in the 246 case children and 108 healthy Korean children without allergic disease or evidence of allergic sensitization where measured levels of total serum IgE as controls were available (see Table E1 in this article's Online Repository at www.jacionline.org). Case and control children were recruited from Severance Children's Hospital, Seoul, Korea, and adult control subjects were from the Ansung population-based cohort (n = 5108), which was established as part of the KoGES by the Korea Center for Diseases Control and Prevention.[16]

Clinical evaluations

AD was diagnosed by pediatric allergists based on the revised Hanifin and Rajka criteria.[17] We first determined severity by using SCORAD indexes[18] of 572 children and then recruited 275 case children with moderate-to-severe AD (SCORAD score ≥30; mean ± SD, 59.9 ± 14.4) for our studies. Allergic sensitization was defined by specific IgE levels of greater than 0.7 kU_A/L to at least 1 of the following food or airborne allergens: egg white, milk, peanut, soybean, wheat, *Dermatophagoides pteronyssinus*, *Dermatophagoides farinae*, *Alternaria* species, or *Blattella germanica*. Of the 275 children with moderate-to-severe AD, 246 had specific IgE to at least 1 of the allergens. We selected 551 adult control subjects with a negative history of both allergic diseases and allergic sensitization among 1214 adults. Negative histories of allergic diseases, including asthma and AD, were based on a self-administered questionnaires; lack of sensitization was based on a negative skin prick test response to 12 common allergens (*D pteronyssinus*, *D farinae*, 2 tree pollen mixtures, grass pollen mixture, ragweed, mugwort, cockroach, *Alternaria* species, *Aspergillus* species, cat dander, and dog dander). Additionally, 108 control children were recruited during routine hospital visits and included in our study if they had a negative history of allergic diseases based on interviews with their parents, had negative serum specific IgE levels to 6 common allergens (egg white, milk, *D pteronyssinus*, *D farinae*, *Alternaria* species, or *Blattella germanica*), and had total serum IgE levels of less than 100 kU/L. All cases and control subjects were unrelated, and either they or their parents provided written informed consent for participation in the study according to the hospital's institutional review board.

Genotyping, imputation, and quality control in the GWAS

Blood samples were collected from each participant, and the derived genomic DNA was genotyped with the Affymetrix Axiom array in the children with AD and control children and the Affymetrix 5.0 chip (Affymetrix, Santa Clara, Calif) in the adult control subjects (see Table E1). We excluded samples with call rates for autosomal SNPs of less than 95% and excluded SNPs with minor allele frequencies of less than 5% or Hardy-Weinberg P values of less than 10^{-4}. Quality control was performed with PLINK 1.07.[19] After quality control exclusions, 402,919 SNPs remained in the children, and 287,622 SNPs remained in the adults. The common sets of SNPs were then used for imputation by using minimac[20] and the 1000 Genomes Asian reference panel.[21] The resulting genotype data for 14,598,181 SNPs were subjected to further quality control checks and selected for high imputation accuracy ($r^2 > .9$) and minor allele frequency of greater than 5%. As a final quality filter, SNPs were excluded if their allele frequencies differed ($P \geq .001$) between the adult and child control samples. In all, 2,501,352 autosomal SNPs were used for the GWAS analysis.

Association of the most significant SNPs with total serum IgE levels

We tested 20 of the 53 SNPs with P values of less than 10^{-6} in the GWAS of AD after pruning for linkage disequilibrium (LD; $r^2 > 0.8$ in the Asian 1000 genomes data) for association with total serum IgE levels in 108 control children (see Table E1). These studies were performed in nonallergic control children to determine whether these variants were also associated with IgE levels independent of AD (see Fig E1 in this article's Online Repository at www.jacionline.org).

Replication studies

To examine the association of SNPs with P values of less than 10^{-6} in our GWAS, we obtained P values of those SNPs from a GWAS of AD in Japanese subjects (see Table E1). That study included 1472 cases with physician-diagnosed AD and 7966 control subjects, including 6042 subjects with one of 5 non-AD diseases (cerebral aneurysm, esophageal cancer, endometrial cancer, chronic obstructive pulmonary disease, and glaucoma) and 1929 healthy volunteers without a history of asthma or AD.[13]

Gene expression and expression quantitative trait loci analysis

To determine whether the SNPs associated with AD in our GWAS were expression quantitative trait loci (eQTLs), we used the eQTL browser (GTEx, http://www.gtexportal.org/)[22] and published reports from eQTL studies of different cell types, including skin,[15,23-25] B cells and monocytes,[26,27] and $CD14^+$ monocytes stimulated with IFN-γ or LPS.[28]

Statistical analysis

We performed logistic regression analysis for binary phenotypes (AD) and linear regression analysis for continuous phenotypes (total serum IgE) by using R software for an additive model. The statistical significance of the association with each SNP was assessed by using a 1-*df* Cochran-Armitage trend test. Regional association plots were generated with LocusZoom.[29]