# A genome-wide association study for allergen component sensitizations identifies allergen component–specific and allergen protein group–specific associations

Wataru Morii, PhD,[a] Koki Kasai, BS,[a] Takako Nakamura, MSc,[a] Daisuke Hayashi, MD,[a] Monami Hara, MD,[a] Tatsuhiko Naito, MD, PhD,[b,c,d] Kyuto Sonehara, MD, PhD,[b,e] Tatsuki Fukuie, MD, PhD,[f] Mayako Saito-Abe, MD, PhD,[f] Limin Yang, MD, PhD,[f] Kiwako Yamamoto-Hanada, MD, PhD,[f] Masami Narita, MD, PhD,[f,g] Kazushi Maruo, PhD,[h] Yukinori Okada, MD, PhD,[b,d,e,i,j] Emiko Noguchi, MD, PhD,[a] and Yukihiro Ohya, MD, PhD[f] *Tsukuba, Suita, Tokyo, and Kanagawa, Japan*

**Background: Allergic diseases are some of the most common diseases worldwide. Genome-wide association studies (GWASs) have been conducted to elucidate the genetic factors of allergic diseases. However, no GWASs for allergen component sensitization have been performed.**

**Objective: We sought to detect genetic variants associated with differences in immune responsiveness against allergen components.**

**Methods: The participants of the present study were recruited from the Tokyo Children's Health, Illness, and Development study, and allergen component–specific IgE level at age 9 years was measured by means of allergen microarray immunoassays. We performed GWASs for allergen component sensitization against each allergen (single allergen component sensitization, number of allergen components analyzed, n = 31), as well as against allergen protein families (allergen protein group sensitization, number of protein groups analyzed, n = 16).**

**Results: We performed GWAS on 564 participants of the Tokyo Children's Health, Illness, and Development study and found associations between Amb a 1 sensitization and the immunoglobulin heavy-chain variable gene on chromosome 14 and between Phl p 1 sensitization and the *HLA* class II region on chromosome 6 ($P < 5.0 \times 10^{-8}$). A GWAS-significant association was also observed between the *HLA* class II region and profilin sensitization ($P < 5.0 \times 10^{-8}$).**

**Conclusions: Our data provide the first demonstration of genetic risk for allergen component sensitization and show that this genetic risk is related to immune response genes including immunoglobulin heavy-chain variable gene and *HLA*. (J Allergy Clin Immunol Global 2023;2:100086.)**

***Key words:*** *Genome-wide association study, immunogenetics, allergen components, HLA, IGHV*

Allergic diseases including allergic rhinitis, allergic asthma, and atopic dermatitis are common diseases worldwide.[1] Allergic diseases have been increasing in prevalence in both developed and developing countries in the last decade,[1] and the high prevalence of allergic diseases has become a social problem in many countries. A nationwide survey of Japanese primary school children (aged 6-8 years) found that 10.2% had wheeze; 18.7%, rhinoconjunctivitis; and 14.6%, eczema.[2]

Prevention of allergic disease development has been proposed to be categorized into primary, secondary, and tertiary prevention.[3] In food allergy, primary prevention is defined as the prevention of IgE sensitization itself; secondary prevention, as the prevention of the onset of allergic symptoms in IgE-sensitized individuals; and tertiary prevention, as seeking to reduce the expression of end-organ allergic disease in children with established food allergy.[3] The status of allergen sensitization is therefore important in evaluating allergic diseases[4] as well as in preventing them. Diagnosis of allergic diseases has become remarkably more accurate through improvements in allergen peptide purification.[5] One of the most accurate allergy diagnostic methods in recent years is allergen component–specific IgE.[6] Extracts using conventional crude allergens contain various allergens, and it is now possible to measure allergen component–specific IgE using a recombinantly purified protein.[5] It has been reported that the measurement of allergen component–specific IgE improves diagnostic accuracy and is also useful for identifying cross-reactivity, such as oral allergy syndrome, and for detecting therapeutic

From [a]the Department of Medical Genetics, Faculty of Medicine, University of Tsukuba, Tsukuba, [b]the Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita, [c]the Department of Neurology, Graduate School of Medicine, The University of Tokyo, Tokyo, [d]the Laboratory for Systems Genetics, RIKEN Center for Integrative Medical Sciences, Kanagawa, [e]the Integrated Frontier Research for Medical Science Division, Institute for Open and Transdisciplinary Research Initiatives, Osaka University, Suita, [f]the Allergy Center, National Center for Child Health and Development, Tokyo, [g]the Department of Pediatrics, School of Medicine, Kyorin University, Tokyo, [h]the Department of Biostatistics, Faculty of Medicine, University of Tsukuba, Tsukuba, [i]the Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita, and [j]the Department of Genome Informatics, Graduate School of Medicine, the University of Tokyo.

This work was supported by a grant from the National Center for Child Health and Development (grant no. 26-18), by a grant from the the Japan Society for the Promotion of Science (JSPS) KAKENHI (grant no. 16H03269); and by a grant from Grant-in-Aid for JSPS Fellows (grant no. 21J11802), and by the Japan Agency for Medical Research and Development (AMED; grant no. JP20ek0410076).

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

Received for publication July 5, 2022; revised October 3, 2022; accepted for publication November 4, 2022.

Available online February 20, 2023.

Corresponding author: Emiko Noguchi, MD, PhD, Department of Medical Genetics, Faculty of Medicine, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8575, Japan. E-mail: enoguchi@md.tsukuba.ac.jp.

The CrossMark symbol notifies online readers when updates have been made to the article such as errata or minor corrections

2772-8293



https://doi.org/10.1016/j.jacig.2023.100086

*Abbreviations used*
eQTL: Expression quantitative trait locus
GWAS: Genome-wide association study
IGHV: Immunoglobulin heavy variable gene
OR: Odds ratio
PC: Principal component
SNP: Single nucleotide polymorphism
T-CHILD: Tokyo Children's Health, Illness, and Development

targets for allergen-specific immunotherapy.[7] The European Academy of Allergy and Clinical Immunology published guidelines for molecular-based diagnosis and declared that allergen-specific immunotherapy should be prescribed only when the clinical relevance of a given allergen source has been reliably demonstrated.[8] Allergen components have been proposed as predictive markers of allergen immunotherapy because they distinguish between patients with genuine reactivity to biologic sources and those with misrecognition of biologic sources including pathogenesis-related PR proteins and profilin.[8] A multiplex assay designed for multiple allergen components in human serum has been developed as a measure of allergen component–specific IgE, making it possible to evaluate an individual's sensitization profile to more than 100 allergen components.[9] Patelis et al[10] reported that IgE sensitization of allergen components may have higher clinical and prognostic values than those of extract-based measurements. We previously examined allergen component–specific IgE reactivity among children and showed high allergen sensitization rates, 57.8% at age 5 years and 74.8% at age 9 years, and that sensitization of allergen components related to allergic rhinitis was increased in children aged 9 years when compared with that in children aged 5 years.[11]

Allergic diseases and allergic sensitization are multifactorial diseases in the development of which both genetic and environmental factors play roles. Genome-wide association studies (GWASs) have been conducted to elucidate the genetic factors of allergic diseases, and many genes associated with allergic diseases have been identified.[12-16] According to the NHGRI-EBI GWAS catalog (gwas-catalog-v1.0.3; accessed October 6, 2021),[17] a curated collection of all human GWASs, 110 studies have been conducted on asthma: 11, on atopic dermatitis; 3, on eczema; 4, on food allergy; 4, on allergic rhinitis; and 2, on allergen sensitization.[15,18] Therefore, not many GWASs other than for asthma have been conducted so far, and to our knowledge, no GWAS for allergen component sensitization has been performed.

In the present study, we performed the first GWAS of the sensitization status of allergen components in a general Japanese pediatric population. The aim of the study was to detect genetic variants associated with differences in immune responsiveness against allergen components.

## METHODS

### Participants

The Tokyo Children's Health, Illness, and Development (T-CHILD) study is an ongoing single-center, prospective, and hospital-based birth cohort study by the National Center for Child Health and Development, Tokyo, Japan. Detailed information about the T-CHILD study has been described elsewhere.[11,19,20] The T-CHILD participants were recruited at the National Center for Child Health and Development during pregnancy, and a total of 1701 pregnant women and 1550 newborns participated in the T-CHILD study between 2003 and 2005. The genome study participants were recruited from among the T-CHILD study participants, and at age 4 to 5 years, 738 children participated in the genome study.[21] Informed consent was obtained from all the parents of the participants of the genome study. The flow chart of the present study is shown in Fig 1. Status of asthma, wheeze, eczema, or rhinitis was obtained from the International Study of Asthma and Allergies in Childhood (ISAAC) questionnaire.[22]

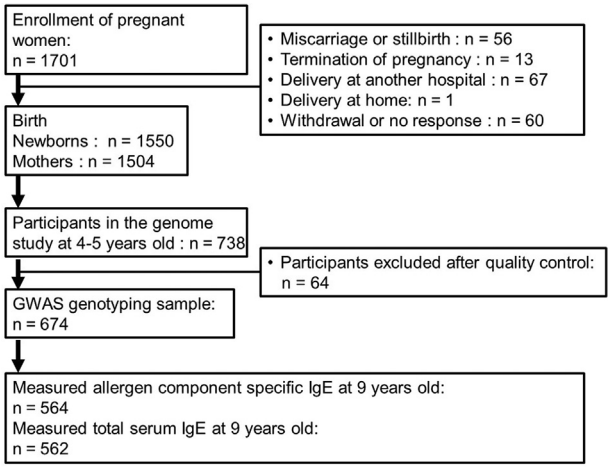


**FIG 1.** Flow chart of the present study.

The study was approved by the human genome research ethics committees of the University of Tsukuba (no. 242) and the National Center for Child Health and Development (no. 533). It was conducted in accordance with the Declaration of Helsinki.

### Allergen component IgE reactivity measurements

Detailed information about blood sampling and measurement of allergen component–specific IgE has been described previously.[11] Allergen component–specific IgE in blood was measured with ImmunoCAP ISAC (Thermo Fisher Scientific/Phadia, Uppsala, Sweden).[23] The specific IgE antibody values were within a range of 0.3 to 100 ISAC standardized units, and the allergen-specific IgE values were converted to binary values with a cutoff value of 0.3 ISAC standardized units. Total serum IgE levels were determined using ImmunoCAP (Thermo Fisher Scientific/Phadia).

Allergen components can be categorized according to their biochemical properties and protein family.[24] We defined the allergen protein group as sensitization positive if the participants were sensitized against any allergen components belonging to the same protein group (see Table E1 in this article's Online Repository at www.jaci-global.org).

We analyzed the allergen component/protein group with a greater than 3% positive sensitization rate in the study population for the GWAS analyses (n = 31 for allergen component and n = 16 for allergen protein group). The numbers of cases/controls for the allergen component/protein group for the GWAS analyses are presented in Table E2 in this article's Online Repository at www.jaci-global.org.

The correlation among allergen components and IgE reactivity was analyzed with the Spearman rank correlation coefficient ($\rho$) and visualized with the corrplot library (https://github.com/taiyun/corrplot) implemented in R software (http://www.R-project.org/).

### GWAS

Genomic DNA was extracted from saliva by use of Oragene DNA (Genotek, Ottawa, Canada) or from blood by use of a QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany) per the manufacturer's protocol. Genotyping was performed using Infinium Asian Screening Array-24 v1.0 (Illumina, San Diego, Calif) and an Illumina iScan Bead Array Reader (Illumina). The output from the Illumina iScan Bead Array Reader was processed with GenomeStudio 2.0 software (Illumina), and genotype calling

and quality control were performed with the previously reported pipeline[25] with some modifications. We selected single nucleotide polymorphisms (SNPs)/variants according to the following criteria: minor allele frequency more than 1%, variant genotyping call rate more than 95%, and Hardy-Weinberg equilibrium $P$ value more than $1.0 \times 10^{-5}$. Variants with a Gentrain score of less than 0.7 were excluded by the GenomeStudio 2.0 software. We excluded samples with a call rate of less than 98%, phenotype-genotype sex discordance, a close familial relationship estimated with PLINK[26] as PI_HAT more than 0.1875, and outliers from the Japanese ancestry estimated by use of principal-component (PC) analysis using the 1000 Genome Project Phase 3.[27] After quality controls, 674 individuals remained for further analysis (Fig 1). To perform PC analysis, we used independent markers estimated with the PLINK –indep-pairwise option, where a window of 50 SNPs was considered at a time and removed if the linkage disequilibrium was greater than 0.5. PCs were calculated by use of EIGENSOFT,[28] and the first 10 PCs were used for the analysis. The genotypes were then imputed by use of Minmac4 software[29] through the use of the Japanese reference panel (JGAS000114 reference panel; https://ddbj.nig.ac.jp/resource/jga-dataset/JGAD000220) available at the National Bioscience Database Center.[30] After imputation, we selected variants with an imputation quality of Rsq more than 0.7 and minor allele count greater than or equal to 20, and finally, we used 6,471,740 variants for the GWAS.

The GWAS for allergen component sensitization was conducted after adjustment for the first 10 PCs and sex under a logistic regression model using SAIGE version 1.0.6.[31] The odds ratios (ORs) and 95% CIs were calculated to estimate the degree of the association. The effect size estimation of variants with a $P$ value less than or equal to .05 was performed by use of Firth's bias-reduced logistic regression.

The GWAS for total serum IgE was conducted with the glm function of PLINK 2.0[32] using log-transformed values of total serum IgE, adjusting for the first 10 PCs and sex. We used alpha levels of $5.0 \times 10^{-8}$ to define GWAS-significant associations and of $1.0 \times 10^{-5}$ for GWAS-suggestive associations. The GWAS results were plotted by means of PheGWAS[33] and CMplot.[34] To assess the independence of the associations, conditional analysis was performed if there were GWAS-significant variants, and the regions around 20,000 bp were depicted by use of LocusZoom.[35] Variants were annotated with ANNOVAR.[36]

## HLA imputation and HLA association analyses

We conducted imputation of *HLA* alleles (*HLA-A*, *B*, *C*, *DRB1*, *DQA1*, *DQB1*, *DPA1*, and *DPB1* alleles/amino-acid polymorphisms) by use of DEEP*HLA software[37] when we observed significant GWAS associations on the *HLA* region. The statistical methods of the *HLA* allele and amino-acid analyses were previously described in detail.[37-39] We performed logistic regression analysis between phenotypes and *HLA* class II 4-digit alleles/amino-acid polymorphisms including *HLA-DRB1*, *DQA1*, *DQB1*, *DPA1*, and *DPB1*, in which each allele was coded as a biallelic marker (0, 1, or 2 copies of the allele). The first 10 PCs and sex were used as covariates of the logistic regression model. An omnibus $P$ value was calculated for the association analysis between an allergen component sensitization and an HLA amino-acid polymorphism.[39] The omnibus $P$ value of the variant was obtained by a log-likelihood ratio test comparing the likelihood of a null model against the likelihood of the fitted model. Firth's logistic regression analysis was performed if complete separation was observed in the logistic model.[40]

## Expression quantitative trait locus analyses

The expression quantitative trait locus (eQTL) results of 5 immune-cell subsets ($CD4^+$ T cells, $CD8^+$ T cells, B cells, natural killer cells, and monocytes) derived from 105 healthy Japanese volunteers (https://humandbs.biosciencedbc.jp/en/hum0099-v1) were extracted for the top significant variants in the GWAS-significant regions.[41] We selected the eQTL results for genes with $q$-value thresholds of 0.05. We also extracted the results of eQTL from the Genotype-Tissue Expression Portal database (https://gtexportal.org/home/, version 8) to visualize the variant effect on various tissues.[42]

**TABLE I.** The characteristics of the participants

| Characteristic | n (%) |
|---|---|
| No. of participants | 564 |
| Sex: male | 283 (50.2) |
| Asthma current* | 29 ( 5.2) |
| Asthma ever* | 116 (20.6) |
| Wheeze current* | 62 (11.0) |
| Wheeze ever* | 185 (32.9) |
| Eczema current* | 119 (21.1) |
| Eczema ever* | 141 (25.0) |
| Rhinitis current* | 324 (57.5) |
| Rhinitis ever* | 341 (60.6) |
| Geometric mean total serum IgE† (95% CI) | 135.4 IU/mL (118.3-154.9) |

*n = 563.
†n = 562.

## RESULTS

The number of participants analyzed in this study is shown in Fig 1. Of the 738 genetic study participants of the T-CHILD study, allergen component IgE data of the ImmunoCAP ISAC sIgE 112 (Thermo Fisher Scientific/Phadia) at age 9 years were available for 564 participants. The number of male participants was 283 (50.2%). The participants' baseline characteristics are presented in Table I. The geometric mean of total serum IgE (n = 562) was 135.4 IU/mL (95% CI, 118.3-154.9 IU/mL).

Cry j 1 (*Cryptomeria japonica*) showed the highest positive sensitization rate, followed by Der f 1 (*Dermatophagoides farinae*) and Der p 1 (*D pteronyssinus*) (see Fig E1 in this article's Online Repository at www.jaci-global.org); the distributions of the number of sensitization-positive components per individual are shown in Fig E2 in this article's Online Repository at www.jaci-global.org. One hundred thirty-six participants (24.1%) were not sensitized to any of the allergen components, and the maximum number of sensitization-positive components per individual was 37.

Among the allergen protein group, pectate lyase was the most common, followed by cysteine protease and the NPC2 family of house dust mites (see Fig E3 in this article's Online Repository at www.jaci-global.org); the distributions of the number of sensitization-positive allergen protein groups per individual are shown in Fig E4 in this article's Online Repository at www.jaci-global.org.

Fig E5 in this article's Online Repository at www.jaci-global.org show the correlation among IgE reactivities to allergen components calculated using the Spearman rank correlation coefficient ($\rho$). As shown in Fig E5, allergen components that belong to PR-10 or profilin and those in house dust mite allergen components showed the strong correlations.

We performed GWAS for each allergen component/allergen protein group sensitization and total serum IgE. Fig 2 shows the 3-dimensional Manhattan plot including 47 traits of allergen sensitization data and 1 trait of total serum IgE data. Manhattan plots of each allergen component and protein group sensitization, and total serum IgE, are shown in Figs E6 (allergen components), E7 (protein groups), and E8 (total serum IgE) in this article's Online Repository at www.jaci-global.org.

A variant (rs35735004) in the immunoglobulin heavy variable (*IGHV*) gene region on chromosome 14 was found to be associated with Amb a 1 sensitization that satisfied the GWAS-significant levels with the T allele as the risk-associated allele (Fig 3, *A-C*; see Table E3 in this article's Online Repository at

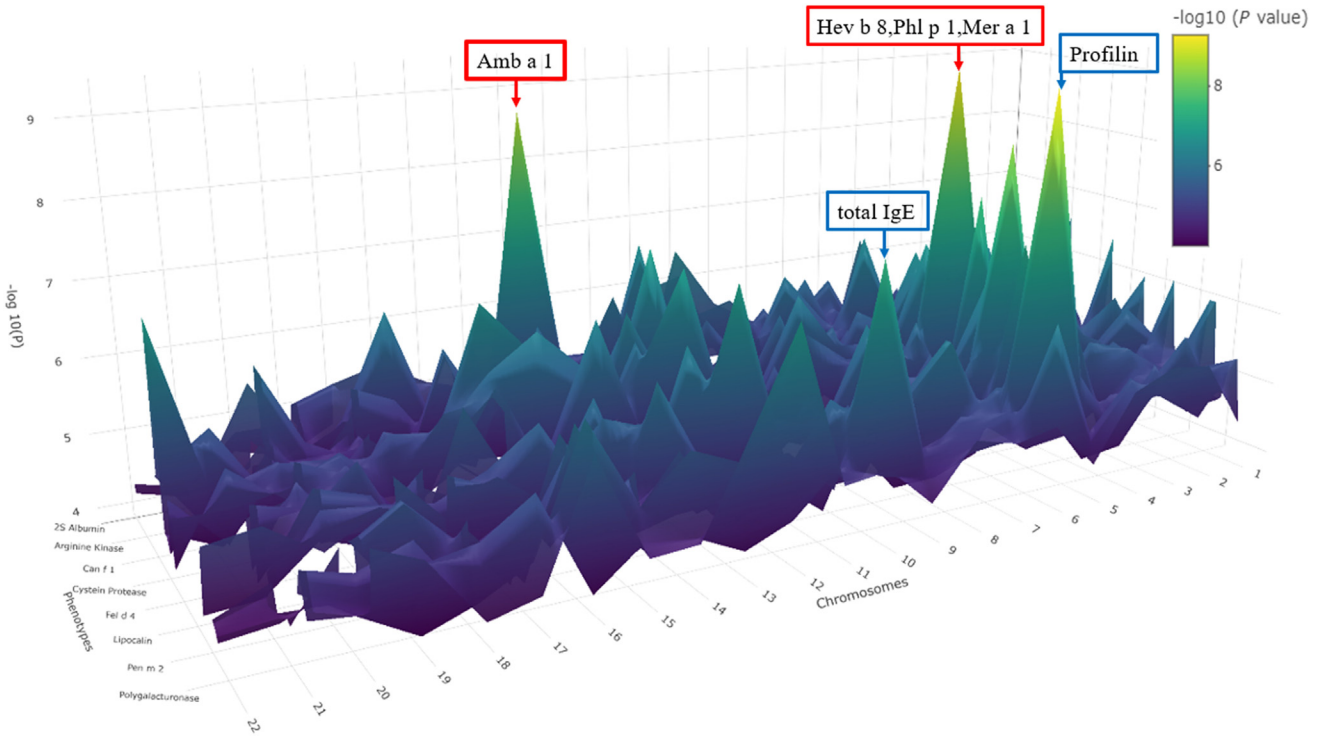


**FIG 2.** 3-Dimensional Manhattan plots of GWAS against allergen components and protein group sensitization. The x-axis (chromosome) represents the genomic locations; the y-axis, the phenotypes (ie, allergen component/allergen protein group sensitizations); and the z-axis (-log10(*P*)), the -$\log_{10}$ (*P* values).

www.jaci-global.org; $P = 1.87 \times 10^{-9}$; OR, 4.62; 95% CI, 2.83-7.55). Conditional analysis with rs35735004 revealed no additional independent associated variants on the chromosomal region (Fig 3, *D*). Then, we analyzed the effect of rs35735004 on 5 immune-cell subsets ($CD4^+$ T cells, $CD8^+$ T cells, B cells, natural killer cells, and monocytes)[41] and whole blood.[42] The risk-associated T allele of rs35735004 corresponded to decreased expression levels of *IGHV2-70* and increased expression levels of *IGHV3-64* and *IGHV3-66* (see Table E4 in this article's Online Repository at www.jaci-global.org).[41] *IGHV4-61* showed the opposite direction of expression levels in the B cells and whole blood. Tissue-specific mRNA expression data from Genotype-Tissue Expression Portal (https://gtexportal.org/home/)[42,43] revealed that the eQTL at rs35735004 with *IGHVs* was not restricted to whole blood but was also present in other tissue types (see Fig E9 in this article's Online Repository at www.jaci-global.org).

We also identified the rs72847623 located at *HLA* region on chromosome 6 and Phl p 1 sensitization, a major grass pollen allergen component (Fig 4, *A* and *B*; Table E3; $P = 3.59 \times 10^{-9}$; OR, 6.03; 95% CI, 3.29-11.04). *HLA* imputation analysis showed 3 amino-acid polymorphisms in HLA-DRβ1 and 1 amino-acid polymorphism in HLA-DQβ1 as suggestive associations (Fig 4, *C* and *D*; see Table E5 in this article's Online Repository at www.jaci-global.org; HLA-DRβ1 position 10: $P = 3.08 \times 10^{-6}$). Conditional analysis by amino-acid position 10 of HLA-DRβ1 revealed no other independent associations (see Fig E10 in this article's Online Repository at www.jaci-global.org).

Furthermore, we identified the GWAS-significant association between the *HLA* region and Hev b 8, and that between the *HLA* region and Mer a 1 (Table E3; $P = 5.27 \times 10^{-10}$, OR, 9.26, 95% CI, 4.65-18.45 for rs1289784088 of Hev b 8 and $P = 2.22 \times 10^{-8}$, OR, 7.19, 95% CI, 3.64-14.20 for rs760563972 of Mer a 1). Both Hev b 8 and Mer a 1 are profilin-associated allergens, and allergen protein group GWAS revealed an association between rs1289784088 and profilin sensitization (Fig 5, *A* and *B*; see Table E6 in this article's Online Repository at www.jaci-global.org; $P = 5.27 \times 10^{-10}$; OR, 9.26; 95% CI, 4.65-18.45). In the *HLA* imputation association analysis, *HLA-DRB1*09:01*, *HLA-DQB1*03:03*, and *HLA-DQA1*03:02* were detected as suggestive loci with profilin sensitization (Fig 5, *C*; see Table E7 in this article's Online Repository at www.jaci-global.org; *HLA-DRB1*09:01*: $P = 4.00 \times 10^{-7}$, OR, 3.82, 95% CI, 2.28-6.40). We also detected 6 amino-acid polymorphisms in *HLA-DRβ1* and 1 amino-acid polymorphism in *HLA-DQα1* to satisfy a GWAS-suggestive association level with profilin sensitization (Fig 5, *D*; see Table E8 in this article's Online Repository at www.jaci-global.org). Conditional analysis by rs1289784088, the GWAS top SNP, or by *HLA-DRB1*09:01* showed no other independent association (see Fig E11 in this article's Online Repository at www.jaci-global.org). The Manhattan plots of Bet v 2, Cor a 1.0101, Phl p 12, Mer a 1, and Hev b 8 are shown in Fig E6.

The GWAS for total serum IgE levels revealed that a variant (rs118175928 T>A) located in the intron of *LINC01515* on chromosome 10 was associated with total serum IgE levels that satisfied the GWAS-significant levels ($P = 3.61 \times 10^{-8}$; $\beta$ coefficient of A allele, −1.23; SE, 0.22; Fig E8).

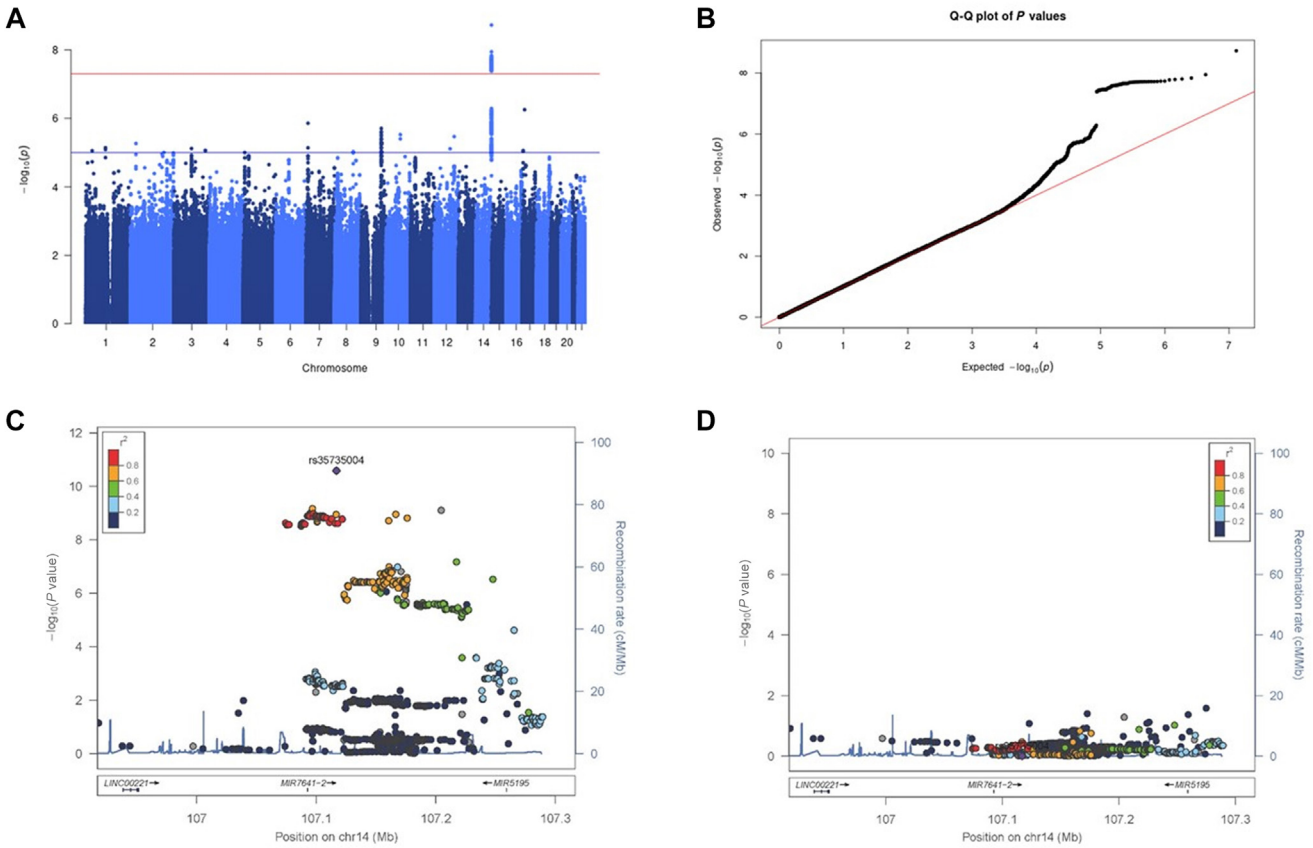


**FIG 3.** GWAS for sensitization against Amb a 1. **A,** Manhattan plot for GWAS with sensitization against Amb a 1. The x-axis indicates chromosomal positions, and the y-axis, the $-\log_{10}$ (*P* values) calculated with SAIGE. The red line indicates the genome-wide significance level ($P = 5.0 \times 10^{-8}$), and the blue line, the genome-wide suggestive level ($P = 1.0 \times 10^{-5}$). **B,** Quantile-quantile (Q-Q) plot of GWAS with sensitization against Amb a 1. The Q-Q plot indicates the expected $-\log_{10}$ (*P* values) vs the observed $-\log_{10}$ (*P* values). **C,** Regional association plots of the GWAS-significant region on chromosome 14. Variants are colored according to their linkage disequilibrium (LD) based on the 1000 Genomes Project Phase 3's EAS reference panel, and the top-associated SNP (rs35735004) is marked with a purple diamond. **D,** Regional association plots of the GWAS-significant region on chromosome 14 after conditioning on the top-associated SNP (rs35735004).

To examine the effect of the GWAS-significant variants on other phenotypes, associations between selected variants (rs35735004, rs72847623, rs1289784088, rs760563972, and rs118175928) and allergen component sensitization/protein groups/total serum IgE were evaluated. As shown in Fig E12 in this article's Online Repository at www.jaci-global.org, rs35735004 was not observed in sensitization phenotypes other than Amb a 1, and rs1289784088 and rs760563972 were found to be associated with multiple phenotypes that belong to the profilin protein group.

Gheerbrant et al[44] reported associations between HLA alleles and allergen component sensitization. Among 8 allergen components (Alt a 1, Art v 1, Bet v 1, Der p 7, Ole e 1, Phl p 2, Phl p 5, and Pla l 1) with significant associations of FDR *q* value less than or equal to 0.05 shown by Gheerbrant et al's study,[44] Alt a 1, Bet v 1, and Phl p 5 were those with greater than a 3% positive sensitization rate in the present study population. We examined the associations between the *HLA* allele and Alt a 1, Bet v 1, and Phl p 5 sensitization. As presented in Table E9 in this article's Online Repository at www.jaci-global.org, we observed the same direction of association as that of Gheerbrant et al's study, although the *P* values did not reach a significant level.

## DISCUSSION

In the present study, we performed allergen component sensitization GWAS of individuals aged 9 years, including single allergen component–specific sensitization, and allergen protein group sensitization. We found that variants located on *IGHV* are associated with sensitization against Amb a 1. We also found that variants located on the *HLA* region are associated with sensitization against Phl p 1 and profilin sensitizations.

*IGHV* is a gene that constitutes the variable portion of immunoglobulins and contributes to the acquisition of immunoglobulin diversity through genetic recombinations in B cells.[45] The variable portions of immunoglobulin heavy chains are composed of *IGHV* (100-300 types), *IGHD* (approximately 25 types), and *IGHJ* (6 types). During B-cell maturation, genes from *IGHV*, *IGHD*, and *IGHJ* are selected through genetic recombination, which results in the formation of a unique immunoglobulin heavy

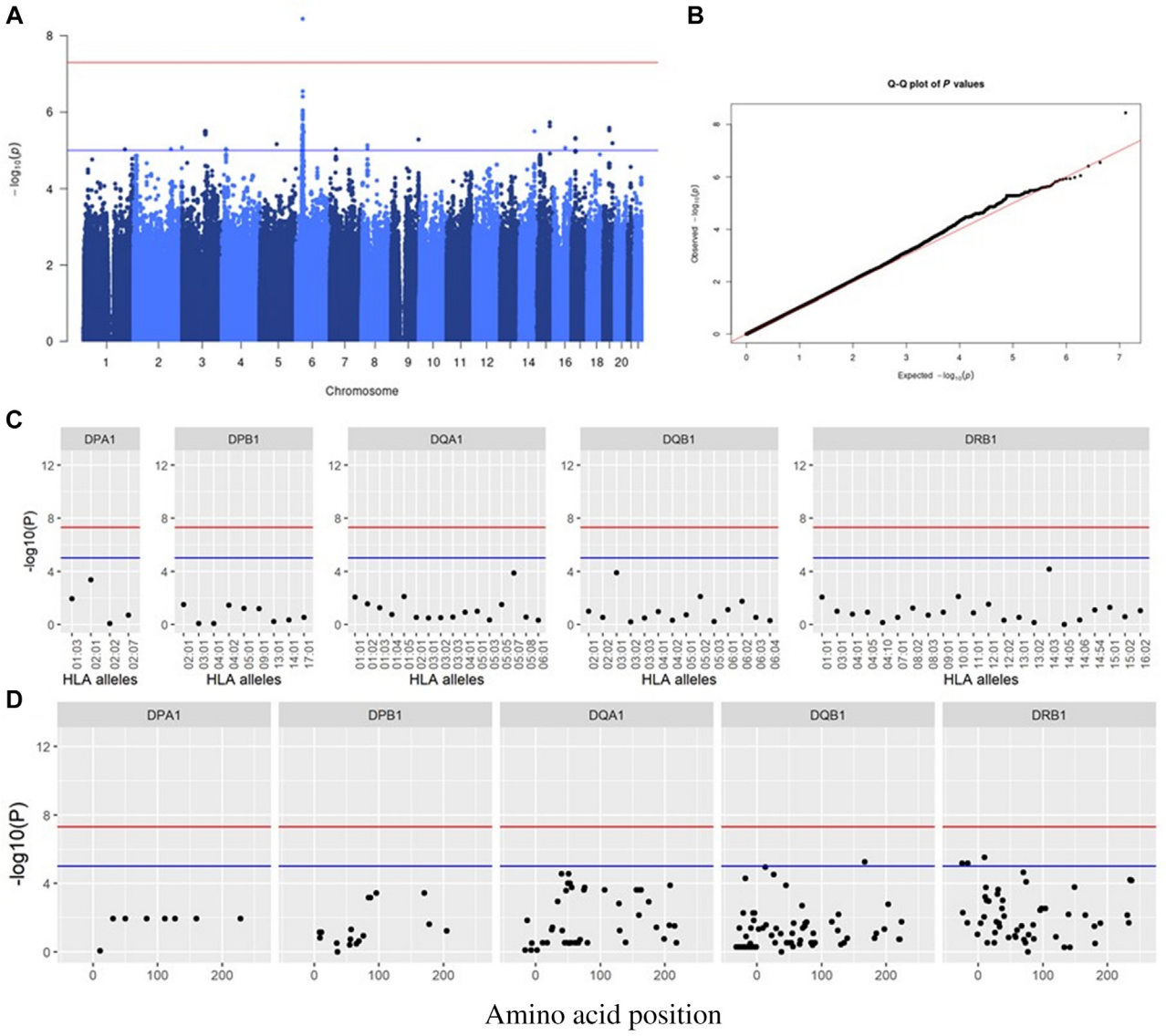


**FIG 4.** GWAS for sensitization against Phl p 1. **A,** Manhattan plot for GWAS with sensitization against Phl p 1. **B,** Association results of *HLA* class II alleles. The x-axis shows the *HLA* alleles, and the y-axis, the $-\log_{10}$ (*P* values). The red line indicates the genome-wide significance level ($P = 5.0 \times 10^{-8}$), and the blue line, the suggestive level ($P = 1.0 \times 10^{-5}$). **C,** Association results of 4-digit *HLA* class II alleles. Each circle point shows the $-\log_{10}$ (*P* values) for the 4-digit *HLA* class II alleles. The red line indicates the genome-wide significance level ($P = 5.0 \times 10^{-8}$), and the blue line, the genome-wide suggestive level ($P = 1.0 \times 10^{-5}$). **D,** Association results of *HLA* class II amino-acid polymorphisms. Each circle point shows the $-\log_{10}$ (*P* values) for each amino-acid polymorphism. The x-axis indicates the amino-acid positions. The red line indicates the genome-wide significance level ($P = 5.0 \times 10^{-8}$), and the blue line, the genome-wide suggestive level ($P = 1.0 \times 10^{-5}$).

chain in each B cell.[46] *IGHV* has been reported to be associated with several autoimmune diseases.[45] So far, no studies have reported associations between the *IGHV* region and Amb a 1 sensitization. However, in other allergen components, Levin et al[47] reported that Bet v 1–specific IgE antibodies isolated from the nasal mucosa of patients with allergic rhinitis carried the same germline *IGHV* gene, showing the presence of skewed IgE repertories with overrepresentation of the specific VH-encoding transcripts. Recently, it has been reported that the V(D)J gene usage profiles of T-cell receptors are associated with variation in the T-cell receptor β locus.[48] Therefore, our study implies that specific *IGHV* genes, such as those detected by eQTL analysis in relation with the genetic variant of rs35735004, may be associated with Amb a 1 sensitization by influencing the binding property of immunoglobulin genes against allergens. Although previous allergy-related GWASs focused on allergic diseases such as asthma and did not sufficiently detect allergen-specific associations,[13-16] the present study focused on the allergen component and was able to detect Amb a 1 sensitization-specific associations.

Phl p 1 belongs to beta-expansin, a protein involved in plant cell wall growth.[49] The present study showed an association

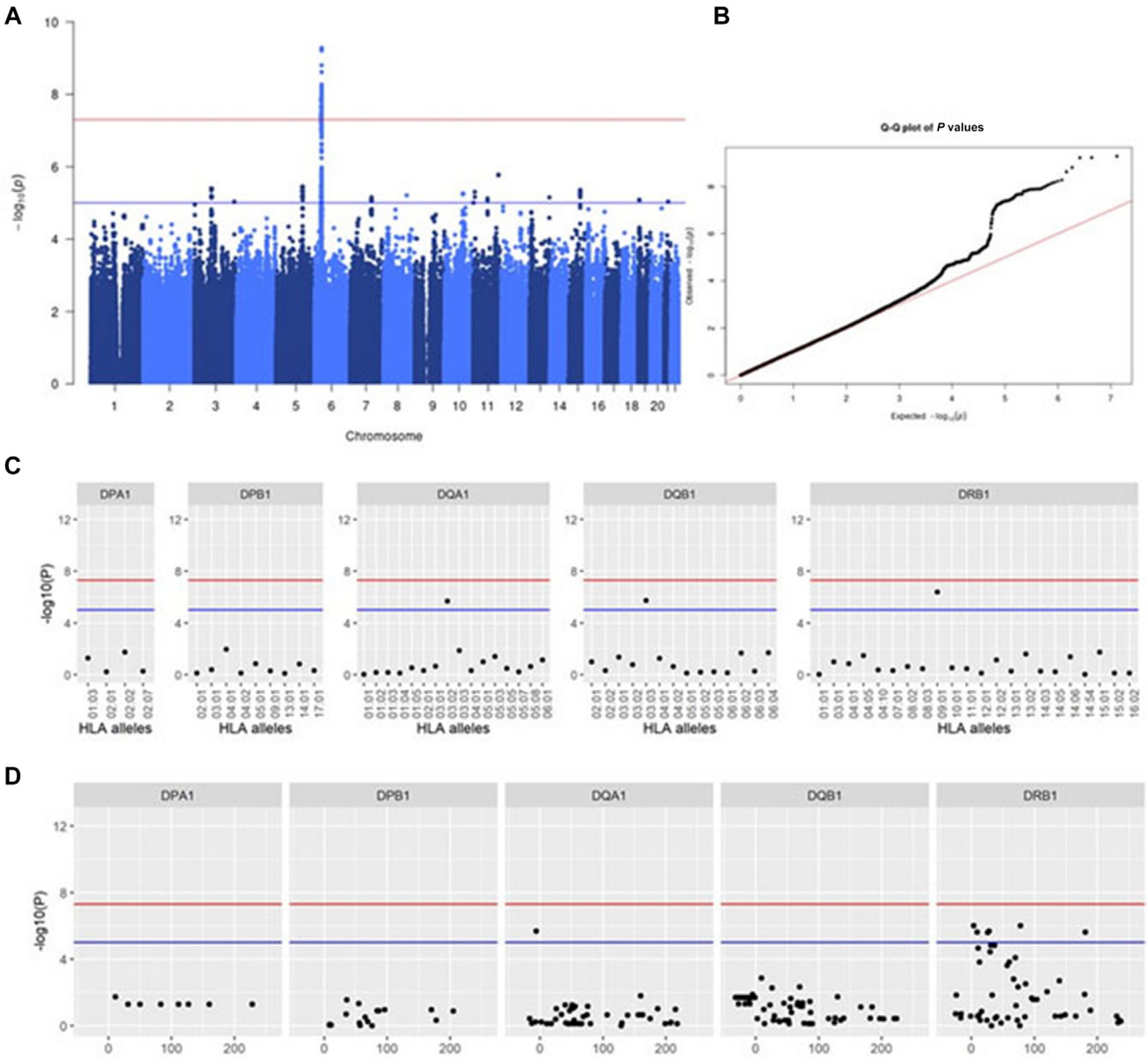


**FIG 5.** GWAS for sensitization against profilin. **A,** Manhattan plot for GWAS with sensitization against profilin. **B,** Association results of *HLA* class II alleles. The x-axis shows the *HLA* alleles, and the y-axis, the $-\log_{10}$ (*P* values). The red line indicates the genome-wide significance level ($P = 5.0 \times 10^{-8}$), and the blue line, the genome-wide suggestive level ($P = 1.0 \times 10^{-5}$). **C,** Association results of 4-digit *HLA* class II alleles. Each circle point shows the $-\log_{10}$ (*P* values) for the 4-digit *HLA* class II alleles. The red line indicates the genome-wide significance level ($5.0 \times 10^{-8}$), and the blue line, the genome-wide suggestive level ($1.0 \times 10^{-5}$). **D,** Association results of *HLA* class II amino-acid polymorphisms. Each circle point shows the $-\log_{10}$ (*P* values) for each amino-acid polymorphism. The x-axis indicates the amino-acid positions. The red line indicates the genome-wide significance level ($5.0 \times 10^{-8}$), and the blue line, the genome-wide suggestive level ($1.0 \times 10^{-5}$).

between rs72847623 and sensitization against Phl p 1. Cyn d 1 belongs to the same protein family as Phl p 1, and the association between Cyn d 1 sensitization and rs72847623 shows the same trend as Phl p 1 association, although it did not reach the GWAS-significant level ($P = 1.02 \times 10^{-4}$; OR, 2.86; 95% CI, 1.67-4.89). One hypothesis is that a unique amino-acid sequence, which is not shared between Phl p 1 and Cyn d 1, may be associated with susceptibility to Phl p 1 sensitization, or else differences in the amount of pollen dispersed may have an impact on the results.

There are many allergen components that belong to the same protein superfamily even though they are derived from different allergens. Cross-reactivity is the phenomenon of sensitization to multiple allergens through components with similar amino-acid structures.[50] In this study, allergen components belonging to the same allergen protein group were analyzed. We found rs1289784088 in the *HLA* region to be associated with profilin sensitization that achieved a GWAS significant level, as well as a GWAS suggestive association for *HLA-DRB1*09:01, HLA-DQB1*03:03, HLA-DQA1*03:02,* 6 amino-acid polymorphisms

in *HLA-DRβ1* and 1 amino-acid polymorphism in *HLA-DQα1*. The amino-acid sequence similarity among allergen components that belong to profilin is estimated to be between 70% and 80%.[51] The GWASs of profilin-associated allergen component sensitizations showed Hev b 8 to have the strongest association (Table E3; $P = 5.27 \times 10^{-10}$; OR, 9.26; 95% CI, 4.65-18.45). The variant with the strongest association for profilin, rs1289784088, is an intronic variant of *HLA-DRB1* with an allele frequency of 0.093 in the Japanese population and of 0.054 in the Korean population, whereas no allele frequency information of the variant was available for Whites or Africans in the dbSNP database (=https://www.ncbi.nlm.nih.gov/snp/?term=rs1289784088; accessed January 31, 2022). The amino-acid polymorphisms that showed suggestive association with profilin sensitization were all located in the allergen-binding region of HLA-DR,[52] and this HLA-DR association disappeared after conditional analysis using rs1289784088. These data suggest that profilin-related HLA alleles, amino-acid polymorphisms, and intronic SNPs are all in linkage disequilibrium, making it difficult to identify which variants/alleles are involved in disease susceptibility.

In the present study, an intronic variant of *LINC01515* was associated with the total serum IgE level at age 9 years. *LINC01515* is a long intergenic nonprotein coding RNA that has been reported to be upregulated in nasopharyngeal carcinoma,[53] but the function of *LINC01515* is largely unknown. We did not find an association between variants in the *HLA* region and total serum IgE levels in the present study. Chang et al[54] compared total serum IgEs with allergen-specific IgE in 3721 patients with allergic diseases, and suggested that allergen-specific IgE can lead to an increased level of total serum IgE, but the level of total serum IgE was not completely determined by the accumulation of allergen-specific IgE. Also, variants associated with sensitization to specific allergen components differ from those of other allergen components on the *HLA* region. Therefore, these factors may have influenced the GWAS association results of total serum IgE levels in the present study.

Gheerbrant et al[44] conducted an association study between *HLA* class II alleles and allergen component sensitizations including 26 aeroallergens in a European adult population. In the present study, association results of Bet v 1, Phl p 5, and Art a 1 sensitizations were available (Table E9). Although some allele frequencies are very low in the Japanese population, we observed the same direction of association and a similar effect size as those of Gheerbrant et al's study, suggesting the possibility of a trans-ethnic association of these HLA alleles with the allergen component sensitization.

There are several limitations to the present study. First, owing to the relatively small sample size, only variants with large effect could be detected, and the OR estimation may be biased. Second, we used $P$ values of $5 \times 10^{-8}$ as GWAS-significant cutoffs despite performing multiple GWASs for each allergen component and protein group; therefore, the cutoff is not stringent enough. IgE values that belong to the same protein family are correlated; therefore, these values are not independent and the Bonferroni correction may be too conservative. Third, although allergen component sensitization measurement has become possible, few studies have been reported that used an allergen microarray immunoassay such as ISAC 112 for a genetic association study, making it difficult to increase the sample size or to have a replication data set; therefore, the possibility of the presence of false positives remains.

Although this was an exploratory study, it investigated a wide range of phenotypes, including allergen component sensitization as well as allergen protein group measurements.

Most previous allergy-related GWASs focused only on allergic diseases and related allergen sensitization; few GWAS analyses are available that used comprehensive allergen sensitization data. The present study was a single-center prospective cohort study, and the IgE measurement was conducted in children of the same age, 9 years; therefore, the degree of allergen exposure was relatively homogeneous, providing information related to the genetics of allergen molecule sensitization in childhood.

We thank the participants of the T-CHILD study, and Flaminia Miyamasu for comments that greatly improved the manuscript.

The Japanese reference panel used for whole genome imputation in the present study was originally obtained by BioBank Japan and the 1000 Genomes Project Phase 3 (v5) and is available at the website of the National Bioscience Database Center/the Japan Science and Technology Agency. The data used for the analyses described in this article were obtained from the Genotype-Tissue Expression Portal on November 3, 2022. The Genotype-Tissue Expression Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by the National Cancer Institute (NCI), National Human Genome Research Institute (NHGRI), National Heart, Lung, and Blood Institute (NHLBI), National Institute on Drug Abuse (NIDA), National Institute of Mental Health (NIMH), and National Institute of Neurological Disorders and Stroke (NINDS).

**Key messages**

- **Amb a 1 sensitization was associated with genetic variants of the *IGHV* on chromosome 14.**
- **Variants on the *HLA* class II region on chromosome 6 were also associated with several allergen components sensitization, including Hev b 8, Mer a 1, Phl p 1, and profilin sensitization.**

## REFERENCES

1. Dierick BJH, van der Molen T, Flokstra-de Blok BMJ, Muraro A, Postma MJ, Kocks JWH, et al. Burden and socioeconomics of asthma, allergic rhinitis, atopic dermatitis and food allergy. Expert Rev Pharmacoecon Outcomes Res 2020;20:437-53.
2. Morikawa E, Sasaki M, Yoshida K, Adachi Y, Odajima H, Akasawa A. Nationwide survey of the prevalence of wheeze, rhino-conjunctivitis, and eczema among Japanese children in 2015. Allergol Int 2020;69:98-103.
3. du Toit G, Tsakok T, Lack S, Lack G. Prevention of food allergy. J Allergy Clin Immunol 2016;137:998-1010.
4. Resch Y, Michel S, Kabesch M, Lupinek C, Valenta R, Vrtala S. Different IgE recognition of mite allergen components in asthmatic and nonasthmatic children. J Allergy Clin Immunol 2015;136:1083-91.
5. Curin M, Garib V, Valenta R. Single recombinant and purified major allergens and peptides: how they are made and how they change allergy diagnosis and treatment. Ann Allergy Asthma Immunol 2017;119:201-9.
6. Steering Committee Authors, Review Panel Members. A WAO - ARIA - GA(2)LEN consensus document on molecular-based allergy diagnosis (PAMD@): Update 2020. World Allergy Organ J 2020;13:100091.
7. Fujimura T, Kawamoto S. Spectrum of allergens for Japanese cedar pollinosis and impact of component-resolved diagnosis on allergen-specific immunotherapy. Allergol Int 2015;64:312-20.
8. Matricardi PM, Kleine-Tebbe J, Hoffmann HJ, Valenta R, Hilger C, Hofmaier S, et al. EAACI Molecular Allergology User's Guide. Pediatr Allergy Immunol 2016;27:1-250.

9. Van Hage M, Schmid-Grendelmeier P, Skevaki C, Plebani M, Canonica W, Kleine-Tebbe J, et al. Performance evaluation of ImmunoCAP® ISAC 112: a multi-site study. Clin Chem Lab Med 2017;55:571-7.
10. Patelis A, Gunnbjornsdottir M, Alving K, Borres MP, Hogman M, Janson C, et al. Allergen extract vs component sensitisation and airway inflammation, responsiveness and new-onset respiratory disease. Clin Exp Allergy 2016;46:730-40.
11. Yamamoto-Hanada K, Borres MP, Aberg MK, Yang L, Fukuie T, Narita M, et al. IgE responses to multiple allergen components among school-aged children in a general population birth cohort in Tokyo. World Allergy Organ J 2020;13:100105.
12. Noguchi E, Akiyama M, Yagami A, Hirota T, Okada Y, Kato Z, et al. HLA-DQ and RBFOX1 as susceptibility genes for an outbreak of hydrolyzed wheat allergy. J Allergy Clin Immunol 2019;144:1354-63.
13. Ramasamy A, Curjuric I, Coin LJ, Kumar A, McArdle WL, Imboden M, et al. A genome-wide meta-analysis of genetic variants associated with allergic rhinitis and grass sensitization and their interaction with birth order. J Allergy Clin Immunol 2011;128:996-1005.
14. Tamari M, Tanaka S, Hirota T. Genome-wide association studies of allergic diseases. Allergol Int 2013;62:21-8.
15. Waage J, Standl M, Curtin JA, Jessen LE, Thorsen J, Tian C, et al. Genome-wide association and HLA fine-mapping studies identify risk loci and genetic pathways underlying allergic rhinitis. Nat Genet 2018;50:1072-80.
16. Zhu Z, Lee PH, Chaffin MD, Chung W, Loh PR, Lu Q, et al. A genome-wide cross-trait analysis from UK Biobank highlights the shared genetic architecture of asthma and allergic diseases. Nat Genet 2018;50:857-64.
17. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res 2017;45:D896-901.
18. Bønnelykke K, Matheson MC, Pers TH, Granell R, Strachan DP, Alves AC, et al. Meta-analysis of genome-wide association studies identifies ten loci influencing allergic sensitization. Nat Genet 2013;45:902-6.
19. Ogawa K, Tanaka S, Limin Y, Arata N, Sago H, Yamamoto-Hanada K, et al. Beta-2 receptor agonist exposure in the uterus associated with subsequent risk of childhood asthma. Pediatr Allergy Immunol 2017;28:746-53.
20. Yang L, Narita M, Yamamoto-Hanada K, Sakamoto N, Saito H, Ohya Y. Phenotypes of childhood wheeze in Japanese children: a group-based trajectory analysis. Pediatr Allergy Immunol 2018;29:606-11.
21. Koseki R, Morii W, Noguchi E, Ishikawa M, Yang L, Yamamoto-Hanada K, et al. Effect of filaggrin loss-of-function mutations on atopic dermatitis in young age: a longitudinal birth cohort study. J Human Genet 2019;64:911-7.
22. Asher MI, Keil U, Anderson HR, Beasley R, Crane J, Martinez F, et al. International Study of Asthma and Allergies in Childhood (ISAAC): rationale and methods. Eur Respir J 1995;8:483-91.
23. Martinez-Aranguren R, Lizaso MT, Goikoetxea MJ, Garcia BE, Cabrera-Freitag P, Trellez O, et al. Is the determination of specific IgE against components using ISAC 112 a reproducible technique? PLoS One 2014;9:e88394.
24. Radauer C, Bublin M, Wagner S, Mari A, Breiteneder H. Allergens are distributed into few protein families and possess a restricted number of biochemical functions. J Allergy Clin Immunol 2008;121:847-52.e7.
25. Guo Y, He J, Zhao S, Wu H, Zhong X, Sheng Q, et al. Illumina human exome genotyping array clustering and quality control. Nat Protoc 2014;9:2643-62.
26. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559-75.
27. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, et al. A global reference for human genetic variation. Nature 2015;526:68-74.
28. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38:904-9.
29. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet 2016;48:1284-7.
30. Okada Y, Momozawa Y, Sakaue S, Kanai M, Ishigaki K, Akiyama M, et al. Deep whole-genome sequencing reveals recent selection signatures linked to evolution and disease risk of Japanese. Nat Commun 2018;9:1631.
31. Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nat Genet 2018;50:1335-41.
32. Hill A, Loh PR, Bharadwaj RB, Pons P, Shang J, Guinan E, et al. Stepwise distributed open innovation contests for software development: acceleration of genome-wide association analysis. Gigascience 2017;6:1-10.
33. George G, Gan S, Huang Y, Appleby P, Nar AS, Venkatesan R, et al. PheGWAS: a new dimension to visualize GWAS across multiple phenotypes. Bioinformatics 2020;36:2500-5.
34. Yin L, Zhang H, Tang Z, Xu J, Yin D, Zhang Z, et al. rMVP: a memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. Genom Proteom Bioinform 2021;19:619-28.
35. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics (Oxford, England) 2010;26:2336-7.
36. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38:e164.
37. Naito T, Suzuki K, Hirata J, Kamatani Y, Matsuda K, Toda T, et al. A deep learning method for HLA imputation and trans-ethnic MHC fine-mapping of type 1 diabetes. Nat Commun 2021;12:1639.
38. Okada Y, Kim K, Han B, Pillai NE, Ong RT-H, Saw W-Y, et al. Risk for ACPA-positive rheumatoid arthritis is driven by shared HLA amino acid polymorphisms in Asian and European populations. Human Mol Genet 2014;23:6916-26.
39. Okada Y, Suzuki A, Ikari K, Terao C, Kochi Y, Ohmura K, et al. Contribution of a non-classical HLA gene, HLA-DOA, to the risk of rheumatoid arthritis. Am J Human Genet 2016;99:366-74.
40. Heinze G, Schemper M. A solution to the problem of separation in logistic regression. Stat Med 2002;21:2409-19.
41. Ishigaki K, Kochi Y, Suzuki A, Tsuchida Y, Tsuchiya H, Sumitomo S, et al. Polygenic burdens on cell-specific pathways underlie the risk of rheumatoid arthritis. Nat Genet 2017;49:1120-5.
42. GTEX CONSORTIUM. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science (New York, N.Y.) 2020;369:1318-30.
43. Aguet F, Brown AA, Castel SE, Davis JR, He Y, Jo B, et al. Genetic effects on gene expression across human tissues. Nature 2017;550:204-13.
44. Gheerbrant H, Guillien A, Vernet R, Lupinek C, Pison C, Pin I, et al. Associations between specific IgE sensitization to 26 respiratory allergen molecules and HLA class II alleles in the EGEA cohort. Allergy 2021;76:2575-86.
45. Bashford-Rogers RJM, Bergamaschi L, McKinney EF, Pombal DC, Mescia F, Lee JC, et al. Analysis of the B cell receptor repertoire in six immune-mediated diseases. Nature 2019;574:122-6.
46. Mikocziova I, Greiff V, Sollid LM. Immunoglobulin germline gene variation and its impact on human disease. Genes Immunity 2021;22:205-17.
47. Levin M, Davies AM, Liljekvist M, Carlsson F, Gould HJ, Sutton BJ, et al. Human IgE against the major allergen Bet v 1—defining an epitope with limited cross-reactivity between different PR-10 family proteins. Clin Exp Allergy 2014;44:288-99.
48. Russell ML, Souquette A, Levine DM, Schattgen SA, Allen EK, Kuan G, et al. Combining genotypes and T cell receptor distributions to infer genetic loci determining V(D)J recombination probabilities. Elife 2022;11:e73475.
49. Grobe K, Becker WM, Schlaak M, Petersen A. Grass group I allergens (beta-expansins) are novel, papain-related proteinases. Eur J Biochem 1999;263:33-40.
50. Ansotegui IJ, Melioli G, Canonica GW, Gómez RM, Jensen-Jarolim E, Ebisawa M, et al. A WAO — ARIA — GA2LEN consensus document on molecular-based allergy diagnosis (PAMD@): Update 2020. World Allergy Organ J 2020;13:100091.
51. del Río PR, Díaz-Perales A, Sánchez-García S, Escudero C, Ibáñez MD, Méndez-Brea P, et al. Profilin, a change in the paradigm. J Investig Allergol Clin Immunol 2018;28:1-12.
52. Bondinas GP, Moustakas AK, Papadopoulos GK. The spectrum of HLA-DQ and HLA-DR alleles, 2006: a listing correlating sequence and structure with function. Immunogenetics 2007;59:539-53.
53. Liu D, Gong H, Tao Z, Chen S, Kong Y, Xiao B. LINC01515 promotes nasopharyngeal carcinoma progression by serving as a sponge for miR-325 to up-regulate CDCA5. J Mol Histol 2021;52:577-87.
54. Chang ML, Cui C, Liu YH, Pei LC, Shao B. Analysis of total immunoglobulin E and specific immunoglobulin E of 3,721 patients with allergic disease. Biomed Rep 2015;3:573-7.