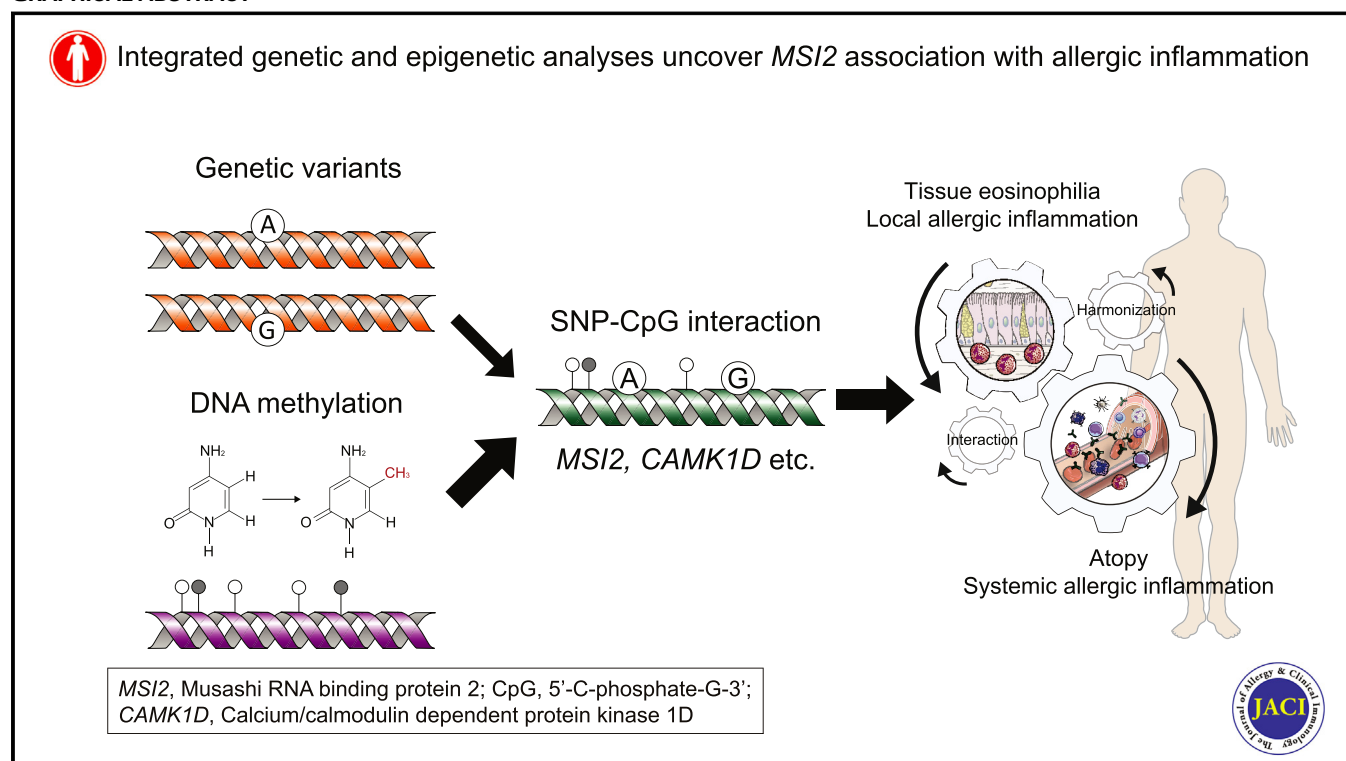


Integrated genetic and epigenetic analyses uncover *MSI2* association with allergic inflammation

Kyung Won Kim, MD, PhD,^{a,*} Sang-Cheol Park, PhD,^{b,*} Hyung-Ju Cho, MD, PhD,^{c,*} Haerin Jang, BS,^a Jaehyun Park, BS,^d Hyo Sup Shim, MD, PhD,^e Eun Gyl Kim, MS,^a Mi Na Kim, PhD,^a Jung Yeon Hong, PhD,^a Yoon Hee Kim, MD, PhD,^f Sanghun Lee, PhD,^g Scott T. Weiss, MD, PhD,^h Chang-Hoon Kim, MD, PhD,^c Sungho Won, PhD,^{b,d,i} and Myung Hyun Sohn, MD, PhD^a *Seoul and Yongin, Korea, and Boston, Mass*

GRAPHICAL ABSTRACT



From ^athe Department of Pediatrics, Severance Hospital, Institute of Allergy, Institute for Immunology and Immunological Diseases, Severance Biomedical Science Institute, Brain Korea 21 PLUS Project for Medical Science, Seoul; ^cthe Department of Otorhinolaryngology, The Airway Mucus Institute, Korea Mouse Phenotyping Center (KMPC), Taste Research Center, Seoul; ^ethe Department of Pathology, Severance Hospital, and ^fthe Department of Pediatrics, Gangnam Severance Hospital, Institute of Allergy, Yonsei University College of Medicine, Seoul; ^gthe Department of Public Health Sciences, ^hInstitute of Health and Environment, and ⁱInterdisciplinary Program for Bioinformatics, College of Natural Science, Seoul National University; ^gDepartment of Medical Consilience, Graduate School, Dankook University, Yongin; and ^hChanning Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston.

*These authors contributed equally to this work.

Supported by National Research Foundation of Korea (NRF) grants funded by the Korean government (grants NRF-2015R1D1A1A01061217, NRF-2018R1A5A2025079, and NRF-2020R1A2B5B02001713); the Bio-Synergy Research Project (grant NRF-2017M3A9C4065964) of the Ministry of Science, ICT and Future Planning through the NRF, the Bio & Medical Technology Development Program of the NRF funded by the Ministry of Science, ICT & Future Planning (grant

NRF-2016M3A9D5A01952414); an NRF grant funded by the Korea government through the Ministry of Science, ICT & Future Planning (grant 2019R1A2C1089841); and grant P01 HL132825 from the National Heart, Lung and Blood Institute, National Institutes of Health (to S.T. Weiss).

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

Received for publication February 16, 2020; revised June 8, 2020; accepted for publication June 26, 2020.

Corresponding authors: Myung Hyun Sohn, MD, PhD, Department of Pediatrics, Yonsei University College of Medicine, 50-1 Yonsei-Ro, Seodaemun-gu, Seoul 03722, Korea. E-mail: mhsohn@yuhs.ac. Or: Sungho Won, PhD, Department of Public Health Sciences, Seoul National University, 1 Kwanak-ro Kwanak-gu Seoul 151742, Korea. E-mail: won1@snu.ac.kr. Or: Chang-Hoon Kim, MD, PhD, Department of Otorhinolaryngology, Yonsei University College of Medicine, 50-1 Yonsei-Ro, Seodaemun-gu, Seoul 03722, Korea. E-mail: ENTMAN@yuhs.ac.

0091-6749/\$36.00

© 2020 American Academy of Allergy, Asthma & Immunology

<https://doi.org/10.1016/j.jaci.2020.06.040>

Background: The relationship between allergic and eosinophilic inflammation, either systemic or local, in allergic diseases remains unclear.

Objective: We performed combined genome-wide association study (GWAS) and epigenome-wide (EWAS) for atopy and tissue eosinophilia to identify both genetic and epigenetic signatures between systemic and local allergic inflammation, and to capture global patterns of gene regulation.

Methods: We included 126 subjects for atopy analysis and 147 for tissue eosinophilia analysis, as well as 18 normal nasal tissue samples. We identified differentially methylated positions (DMPs) and genes associated with atopy and tissue eosinophilia. Furthermore, we performed mendelian randomization analysis and penalized regression along with replication in an independent cohort.

Results: EWAS identified genes, including Musashi RNA binding protein 2 (*MSI2*), associated with atopy, which contained enriched DMPs that genetically affect atopy. A direct association was observed between *MSI2* single-nucleotide polymorphisms and atopy, as was a causal effect of changes in *MSI2* expression and methylation on atopy, which was replicated in a Costa Rican population. Regarding tissue eosinophilia, EWAS identified genes with enriched DMPs directly contributing to tissue eosinophilia at the gene level, including *CAMK1D*. The gene ontology terms of the identified genes for both phenotypes encompassed immune-related terms.

Conclusion: EWAS combined with GWAS identified novel candidate genes, especially the methylation of *MSI2*, contributing to systemic allergic inflammation. Certain genes displayed a greater association with either systemic or local allergic inflammation; however, it is expected that a harmonized effect of these genes influences immune responses. (J Allergy Clin Immunol 2020;■■■:■■■-■■■.)

Key words: Atopy, eosinophilia, GWAS, genetic, epigenetic, methylation

Allergic diseases are complex disorders resulting from both heterogeneous genetic and environmental factors, resulting in phenotypic heterogeneity.¹ Atopy (or allergic sensitization) is a key shared phenotype across allergic diseases, including asthma, allergic rhinitis, and atopic dermatitis. A recent genome-wide association study (GWAS) of allergic diseases identified numerous shared genetic risk variants involved in immune cell function,² which potentially result in the combined phenotype atopy in allergic diseases. Eosinophilic inflammation is a well-known atopic marker. Eosinophils are abundant in the inflammatory infiltrates of late-phase reactions, and they are involved in numerous pathologic processes in allergic diseases.³ The interactive association between allergic inflammation and eosinophilic inflammation, either systemic or local, is not understood clinically or endophenotypically.^{4,5}

Over the past decade, GWASs have identified susceptibility variants that have provided mechanistic insight into the pathogenesis of complex diseases such as allergic diseases; nonetheless, the issues and challenges regarding allergic inflammation remain unclear.⁶ Recently, epigenome-wide association studies (EWASs) of atopy and total serum IgE level reported that epigenetic regulation of genes involved in asthma, immune responses, and mechanistic target of rapamycin signaling

Abbreviations used

ACME:	Average causal mediation effect
ACOT7:	Acyl-CoA thioesterase 7
ADE:	Average direct effect
CAMK1D:	Calcium/calmodulin dependent protein kinase 1D
CpG:	5'-C-phosphate-G-3'
DMP:	Differentially methylated position
EWAS:	Epigenome-wide association study
FDR:	False discovery rate
GEMMA:	Genome-wide efficient mixed-model association
GO:	Gene ontology
GSTT4:	Glutathione S-transferase theta 4
GWAS:	Genome-wide association study
JARID2:	Jumonji and AT-rich interaction domain containing 2
meQTL:	Methylation quantitative trait locus
MR:	Mendelian randomization
MSI2:	Musashi RNA binding protein 2
PC:	Principal component
PPP5D1:	PPP5 tetratricopeptide repeat domain containing 1
RIN3:	Ras and Rab interactor 3
SNP:	Single-nucleotide polymorphism
TRIO:	Trio rho guanine nucleotide exchange factor

are associated with allergic sensitization.⁷⁻⁹ Nonetheless, limited studies have combined GWAS and EWAS to investigate atopy and allergic diseases. We hypothesized that integrated genetic and epigenetic analyses would help identify novel genes that could not have been identified through GWAS or EWAS alone owing to the moderate magnitudes of their associations.^{6,10}

This study aimed to identify epigenetic signatures in atopy (systemic allergic inflammation) and tissue eosinophilia (local allergic inflammation). We used nasal polyp tissue samples to investigate DNA methylation in the local airway tissue from evidence regarding the association between the upper and lower airways.¹¹ We performed combined GWAS and EWAS analyses to identify unique genetic signatures for atopy and tissue eosinophilia and to evaluate shared genetic signatures and biologic pathways between the 2 conditions. We aimed to capture global patterns of genetic risk and the effects of gene-environment interactions on allergic diseases.

METHODS

Fig 1 provides a schematic representation of the combined GWAS and EWAS analyses for atopy and tissue eosinophilia herein. Among the 134 enrolled patients, samples from 108 and 129 patients were used for atopy analyses and eosinophilia analyses, respectively. Furthermore, we used 18 normal nasal tissue samples for these analyses. Clinical definitions of atopy and tissue eosinophilia and detailed sample compositions are provided in the Online Repository materials for this article (available at jacionline.org). All subjects provided written informed consent, and the study was approved by the institutional review board of Severance Hospital (Seoul, Korea; IRB no. 4-2015-1034).

DNA methylation profiling and differences in atopy and tissue eosinophilia

Differences in DNA methylation were assessed by using the Infinium MethylationEPIC BeadChip assay (Illumina, Inc, Calif). More than 850,000

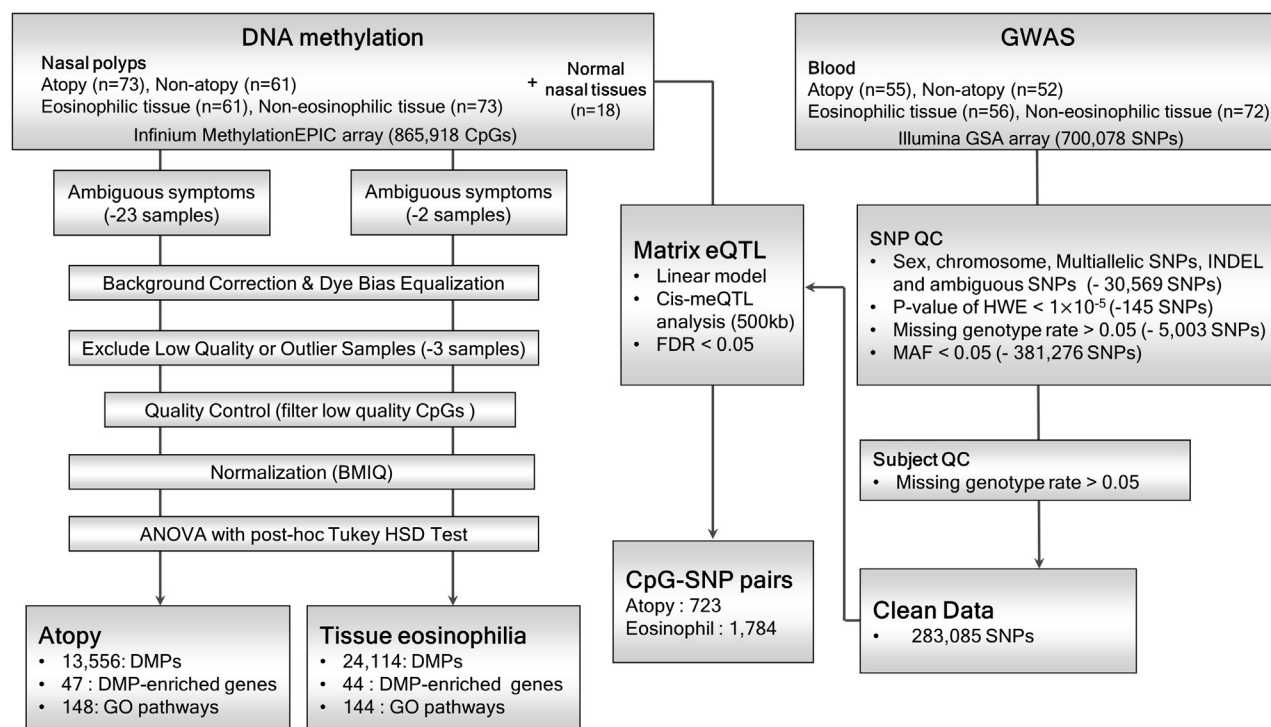


FIG 1. Flowchart of the combined analyses of GWAS with EWAS for atopy and tissue eosinophilia in this study. The epigenetic signature of atopy and tissue eosinophilia was identified in nasal polyps, and genetic variants associated with each phenotype were investigated. These results were integrated by using Matrix expression quantitative trait loci (Matrix eQTL) to capture the causal interaction between epigenetics and DNA sequence variation. *BMIQ*, Beta-mixture quantile; *GSA*, Global Screening Array; *HSD*, honest significant difference; *HWE*, Hardy-Weinberg equilibrium; *INDEL*, insertion/deletion; *MAF*, minor allele frequency; *QC*, quality control.

5'-C-phosphate-G-3' (CpG) sites in the EPIC array were initially assessed, and 808,934 methylation sites from 126 subjects for atopy analysis and 808,934 methylation sites from 147 subjects for tissue eosinophilia analysis remained after quality control (see the [Online Repository materials](#)).

SNP genotyping, quality control, imputation, and genetic analyses

All subjects were genotyped for single-nucleotide polymorphisms (SNPs) by using the Infinium Global Screening Array-24 v1.0 (Illumina, Inc). General quality control procedures were conducted, and genome-wide imputation was performed for all autosomes with IMPUTE2 v.2.32,¹² with all participants' genotypes for the 1000 Genome project phase 3 used as a reference (see the [Online Repository material](#)).

Mapping of methylation quantitative trait loci

For each methylation site, we investigated significant *cis*-methylation quantitative trait loci (meQTLs). Trans-meQTL analysis was not performed in view of the possibility of obtaining false-positive associations between CpG-SNP pairs. CpG-SNP pairs were considered significant when they passed the 5% false discovery rate (FDR) threshold in mappings (see the [Online Repository material](#)).

One-sample mendelian randomization

We analyzed instrumental variables only for significant *cis*-meQTL pairs on the differentially methylated position (DMP)-enriched genes by using the ivprobit R package (v1.1)¹³ (see the [Online Repository material](#)).

Replication with the Costa Rican cohort data

Significantly associated changes in methylation levels were replicated by assessing the association of their meQTL SNPs with use of data from a Costa Rican family-based cohort. Detailed information regarding genome sequencing in this data is provided in the [Online Repository material](#). Phenotypes associated with atopy, including serum total and dust mite-specific IgE levels and peripheral blood eosinophil count, were analyzed for genome-wide efficient mixed-model association (GEMMA) with significant SNPs by using ONETOOL software (v2.0.1).¹⁴

Analysis of genetic and epigenetic variation with *MSI2* expression and atopy

Associations between atopy and SNP, CpG, and Musashi RNA binding protein 2 (*MSI2*) gene expression levels were examined through a mediation test and logistic regression analysis. In addition to an SNP-CpG pair that was significant on mendelian randomization (MR) analysis and the analysis of data from the Costa Rican cohort, 20 DMPs on *MSI2* and their effect on *MSI2* expression levels were considered. Consequently, some of the 20 DMPs revealed major correlations, and to investigate the overall effect of nongenetic DMPs on *MSI2* by overcoming multicollinearity, we performed principal component (PC) analysis for β -values of 20 DMPs. Here, SNPs were coded in a binary manner (0/1) in accordance with risk allele inclusion.

Estimation of relative deviance based on SNPs, CpGs, and clinical variables among total variations

Prediction models with SNPs, CpGs, and 4 clinical variables, including age, sex, total serum IgE level, and blood eosinophils for atopy status and tissue eosinophilia prediction were generated with the penalized regression

TABLE I. Subject characteristics

Characteristic	Atopy analysis (n = 108)		Tissue eosinophilia analysis (n = 129)		Normal nasal tissue (n = 18)
	Atopy (n = 56)	Nonatopy (n = 52)	Eosinophilic tissue (n = 57)	Noneosinophilic tissue (n = 72)	
Male, no. (%)	44 (78.6)	31 (59.6)	45 (78.9)	47 (65.2)	6 (33.3)
Age (y), mean $\pm$ SD	44.7 $\pm$ 13.1	44.9 $\pm$ 14.8	46.8 $\pm$ 12.6	44.1 $\pm$ 14.4	50.1 $\pm$ 17.9
Total serum IgE level (kU/L), mean $\pm$ SD	446.2 $\pm$ 697.6	24.1 $\pm$ 18.0	196.3 $\pm$ 288.2	225.6 $\pm$ 620.5	32.0 $\pm$ 36.7
Atopy, no. (%)	56 (100)	0 (0)	30/48 (62.5)	25/58 (43.1)	0 (0)

model. To evaluate the relative importance of all variables, we determined the McFadden R^2 improvement on the basis of the deviance with respect to their potential to accurately predict phenotypes^{15,16} (see the [Online Repository material](#)).

RESULTS

Subjects' characteristics

Atopy analysis was performed for 126 subjects, and tissue eosinophilia analysis was performed for 147 subjects, including 18 subjects with normal tissues; their clinical characteristics are summarized in [Table I](#).

Differences in DNA methylation in atopy and tissue eosinophilia

To investigate the airway epigenetic signatures associated with atopy and tissue eosinophilia, we identified differences in methylation between atopy and nonatopy and between eosinophilia and noneosinophilia in the nasal polyp. For more than 850,000 methylation probes across the genome, DMPs, DMP-associated genes, and gene ontology (GO) terms were determined with the criteria of 20%, 10%, 5%, and 1% of methylation differences. At a 20% difference level, we found 1 CpG (cg20790648) (ie, predicted enhancer element) for tissue eosinophilia and none for atopy. At a 10% difference level, we found 4 significant CpGs on glutathione S-transferase theta 4 (*GSTT4*) and PPP5 tetratricopeptide repeat domain containing 1 (*PPP5D1*) for atopy and 21 CpGs on 15 genes for tissue eosinophilia, including *HLA-DPA1*, acyl-CoA thioesterase 7 (*ACOT7*), *ZFPM1*, and *CISH*. The results from the other categories are summarized in [Tables E1](#) and [E2](#) (in this article's [Online Repository at www.jacionline.org](#)). Thereafter, we focused on a 1% methylation difference to secure a sufficient number of CpGs for DMP enrichment analysis. Consequently, we identified 25,132 DMPs associated with atopy and 46,503 DMPs associated with tissue eosinophilia between the cases and controls (ANOVA; FDR < 0.05 and Tukey honest significant difference; $P < .05$). Furthermore, we clarified the epigenetic effect of target phenotypes by comparing their methylation levels with the levels in normal nasal tissues. The DMPs analyzed herein were significant among all groups or only between the case and control groups. In the latter case, DMPs associated with polyp formation were excluded before analysis. Consequently, 13,556 CpG sites (1.68%) and 24,144 CpG sites (2.98%) were identified for atopy and tissue eosinophilia, respectively, and postanalysis was conducted for those sites (see [Tables E1](#) and [E2](#)). A higher DMP concentration was observed in gene bodies from either atopy or tissue eosinophilia analyses (see [Fig E1](#) in this article's [Online Repository at](#)

[www.jacionline.org](#)). Furthermore, differential methylation occurred in non-CpG island regions such as open sea.

Thereafter, we examined DMP-enriched genes and GO terms overrepresented among genes with the methylation profiles assessed herein. In all, 47 and 44 genes were identified on atopy and tissue eosinophilia analyses, respectively, with a threshold FDR of less than 0.05 (see [Tables E3](#) and [E4](#) in this article's [Online Repository at www.jacionline.org](#)). The top 9 genes displaying the same methylation pattern (hypomethylation/hypermethylation) are displayed in [Fig 2, A](#) (atopy) and [Fig 2, C](#) (tissue eosinophilia). The genes harboring atopy DMPs were *MSI2*, *CEBPE*, *PRR5L*, *WNT10B*, and *TREML2*, whereas those harboring tissue eosinophilia DMPs were *IL13*, *Ras* and *Rab* interactor 3 (*RIN3*), *CEBPE*, *IL3*, and *ACOT7*. For each gene, the average methylation levels were lower in cases than in controls and normal nasal tissue on both the atopy and tissue eosinophilia analyses.

Thereafter, we examined the functions presenting the largest differences in DNA methylation for the 2 target phenotypes. When CpGs were mapped to genes and GO enrichment analyses were performed by using missMethyl, 148 and 144 biologic processes were significantly different in the atopy ([Fig 2, B](#)) and tissue eosinophilia analyses ([Fig 2, D](#)), respectively. For the atopy analysis, categories including neutrophil-mediated immunity, cell-cell adhesion, and T-cell activation were identified as biologically enriched functions and were more strongly associated with hypomethylated CpGs than with hypermethylated CpGs. As in atopy analysis, categories including granulocyte activation, T-cell activation, and leukocyte migration were significantly enriched on tissue eosinophilia analysis. In contrast, leukocyte migration and regulation of cell morphogenesis were observed on tissue eosinophilia analysis, whereas categories including cell-cell adhesion and cell activation were frequently observed on atopy analysis. Furthermore, the contribution of hypermethylation was slightly higher in tissue eosinophilia ([Fig 2](#)). The entire list of GO pathways can be found in [Tables E5](#) and [E6](#) (in this article's [Online Repository at www.jacionline.org](#)).

Effect of genetic variants on methylation levels

To investigate SNPs associated with DNA methylation, we performed Matrix-eQTL analysis by using both GWAS data and EWAS data on the nasal polyps. After prescreening procedures for genotype data, 283,085 SNPs were investigated for their association with all CpG sites. Within a 500-kb range, 54,791 and 68,509 CpG-SNP pairs displayed significant genetic-epigenetic interactions in the atopy and tissue eosinophilia analyses, respectively (FDR-adjusted P value < .05). To clarify

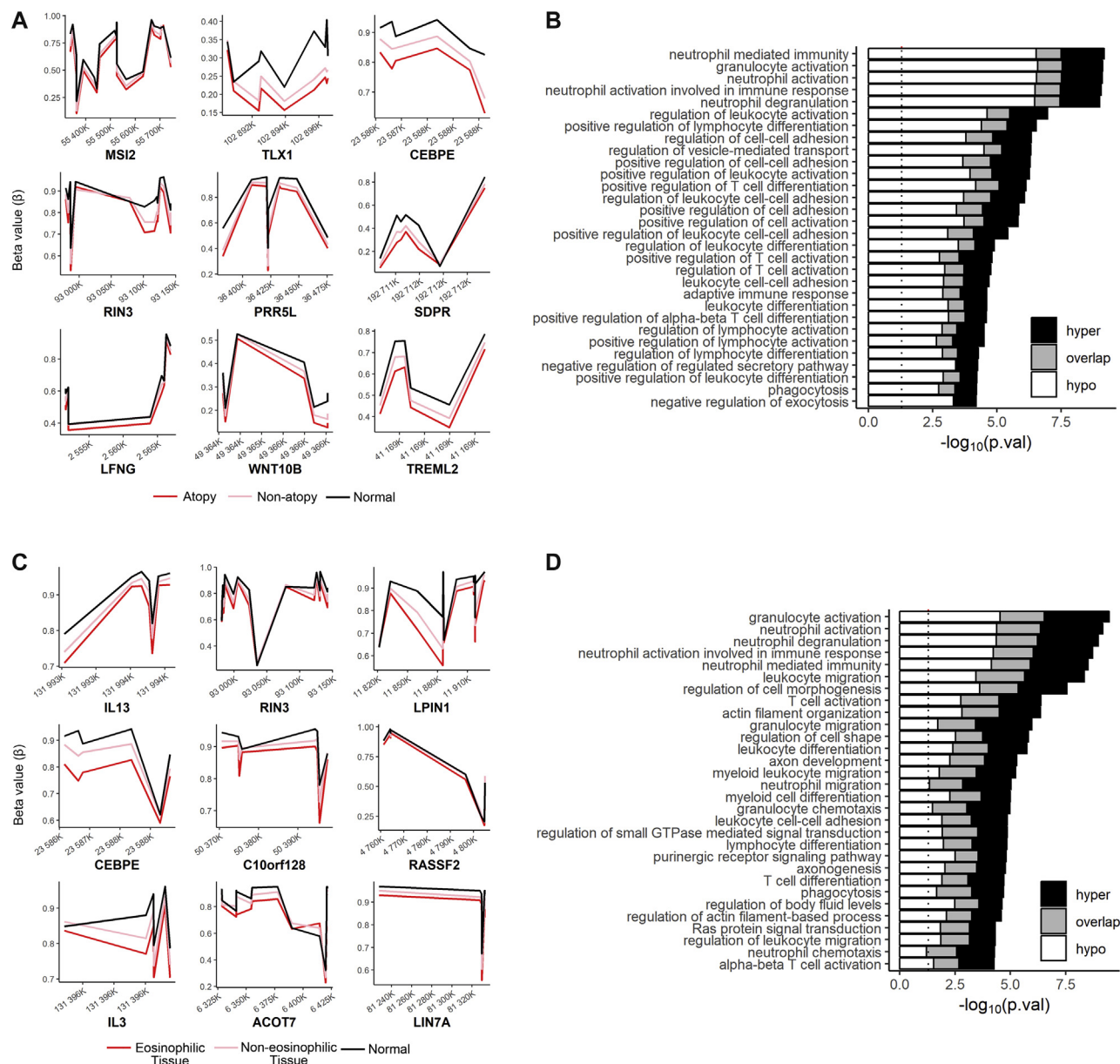


FIG 2. DNA methylation variations are associated with atopy and tissue eosinophilia in nasal polyp. The methylation levels of the top 9 significant DMP-enriched genes (**A**) in atopy and (**C**) eosinophilic tissue. Methylation values for each of the following groups are displayed using β -values: normal (black), nonatopy or noneosinophilic tissue (pink), and atopy or eosinophilic tissue (red). The top 30 GO terms enriched in each methylation profile for atopy (**B**) and tissue eosinophilia (**D**). The degree of significance was obtained from a hypergeometric test and adjusted via the Benjamini-Hochberg method. The color represents whether hypermethylation (black) or hypomethylation (white) has occurred in the CpG site belonging to the GO term.

associations, those pairs that were not among the previously defined DMPs were filtered out because such associations represent CpG sites that are not contingent on each phenotype. Consequently, 723 and 1,784 CpG-SNP pairs were detected in the atopy and tissue eosinophilia analyses, respectively. We further investigated pairs occurring in DMP-enriched genes. On atopy analysis, 4 genes (*MSI2*, *Jumonji* and AT-rich interaction domain containing 2 [*JARID2*], Trio rho guanine nucleotide exchange factor [*TRIO*], and *ACOT7* [8.51%]) displayed

significant changes in DNA methylation, mediated through genetic effects (see Table E7 in this article's Online Repository at www.jacionline.org). On tissue eosinophilia analysis, 10 genes (*RNF19A*, calcium/calmodulin dependent protein kinase 1D [*CAMK1D*], *SLC45A4*, *ACOT7*, *PRDM16*, *NOTCH1*, *GJB2*, *ATAD3C*, *JAZF1*, and *TG* [22.72%]) displayed SNP-mediated changes in methylation (see Table E8 in this article's Online Repository at www.jacionline.org). Risk variants in *MSI2*, *TRIO*, and *ACOT7* (atopy) and variants in *CAMK1D*, *RNF19A*,

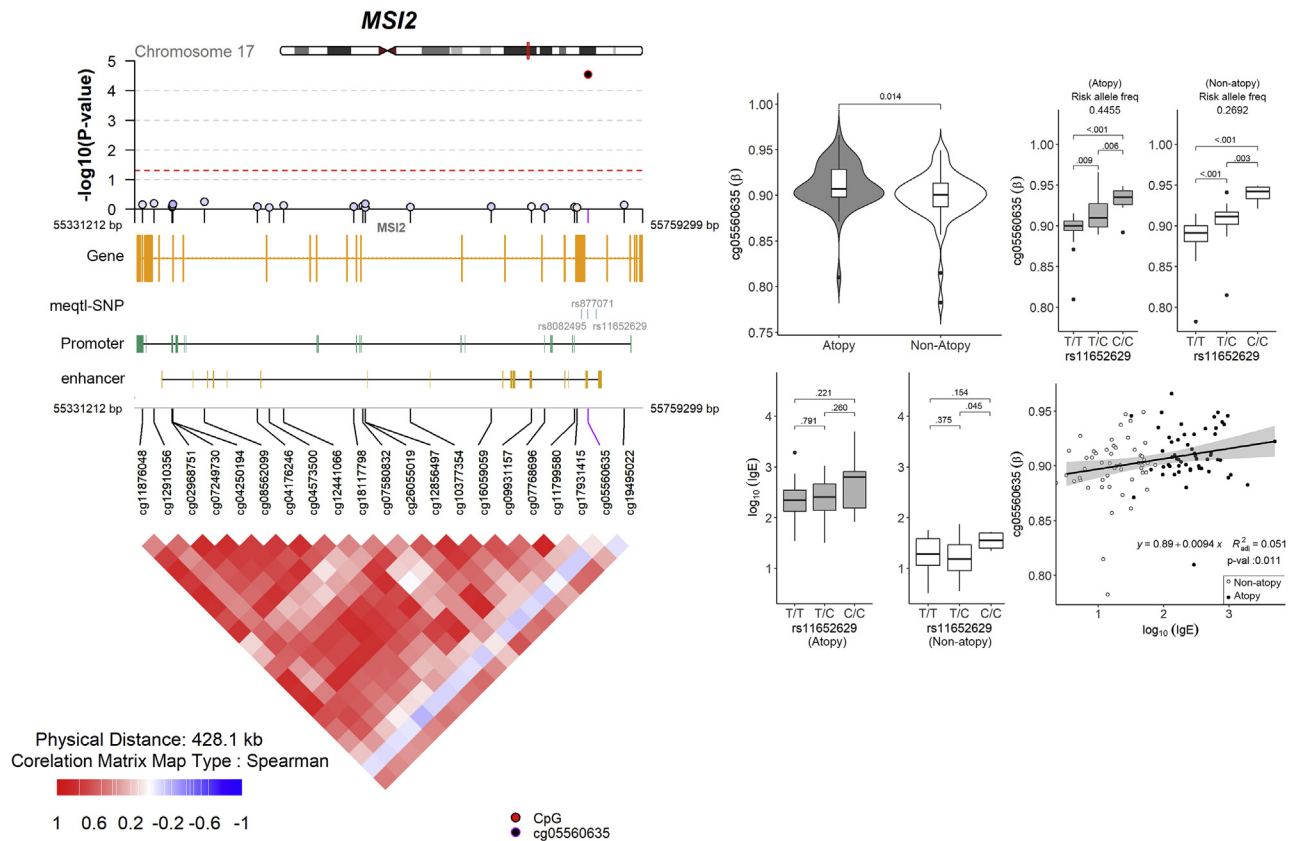


FIG 3. Detailed genomic characterization of *cis*-meQTLs at *MSI2* locus in atopy. The plot displays the genomic region of epigenome-wide association and comethylation profiles. The *top panel* represents negative $\log_{10}$ -transformed meQTL *P* values, the *middle panel* represents the genomic annotation (gene symbols and promoters and genomic enhancers from Ensemble; CpG islands, from the University of California Santa Cruz genome browser) of *cis*-meQTL-SNPs, and the *lower panel* represents comethylation patterns obtained by using the Spearman correlation coefficient. The relationships between methylation level at CpG, genotype, and serum IgE levels are shown on the *right*.

NOTCH1, *ACOT7*, and *GJB2* (tissue eosinophilia) were significantly associated with increased methylation, whereas the other genes were associated with decreased methylation.

Furthermore, we examined genotype-phenotype associations. Two SNPs (rs877071 and rs11652629; LD $r^2 = 0.587$) in *MSI2* were significantly associated with atopy ($P = .0091$ and $.0179$; odds ratio = 2.574 and 2.326, respectively), and the methylation level at cg05560635 in *MSI2* increased in atopy rather than in nonatopy (Tukey $P = .0185$) (Fig 3). In tissue eosinophilia, 2 SNPs (rs7077423 and rs7894890; LD $r^2 = 0.326$) in *CAMK1D* and 3 SNPs (rs753778, rs10102728, and rs7014062; in the case of the last 2 SNPs, LD $r^2 = 0.913$) in *SLC45A4* were significantly associated with the phenotype. The risk allele at rs7077423 displayed the strongest association with tissue eosinophilia ($P = .008$; odds ratio = 2.310), and the methylation level at cg04836423 in *CAMK1D* increased in eosinophilic tissue rather than in noneosinophilic tissue (Tukey $P = 2.31 \times 10^{-7}$).

Moreover, we assessed 12 CpG-SNP pairs for atopy and 29 pairs for tissue eosinophilia to identify the causal relationship between methylation levels and each phenotype (see Table E9 in this article's Online Repository at www.jacionline.org). MR analyses revealed that all 3 pairs in *MSI2* displayed significant or almost significant causal evidence, and significant results were

obtained for rs877071-cg05560635 (coefficient = 1.2374; 95% CI = 0.7219-1.529; $P = .0184$) and rs11652629-cg05560635 (coefficient = 1.1921; 95% CI = 0.5934-1.7908; $P = .0494$). However, no causal associations among methylation levels were observed on tissue eosinophilia analysis.

Replication in the Costa Rican cohort

Every locus significantly associated with a change in methylation levels was independently assessed for atopy-related variables in blood by using GEMMA in a Costa Rican population data set.¹⁷ There were 12 significant CpG-SNPs for atopy at an FDR-adjusted significance level of 0.05. We eliminated the 2 SNPs with minor allele frequencies less than 5% and considered the remaining 10 different CpG-SNPs, for which we calculated the Sidak-adjusted *P* to control the family-wise error rates. For the most of significant loci in atopy, we observed a replication of our results, revealing a marked association between these genes and allergic inflammation. More specifically, total serum IgE level in the blood was significantly associated with *MSI2* (rs8082495 and rs11652629 [$P = .002$ and $.007$, respectively]) SNPs. Among the significant loci in tissue eosinophilia, *CAMK1D* (rs3995694 and rs7077423 [$P = .034$ and $.009$, respectively]) displayed an association with blood eosinophils. Total serum IgE (rs7077423) and dust

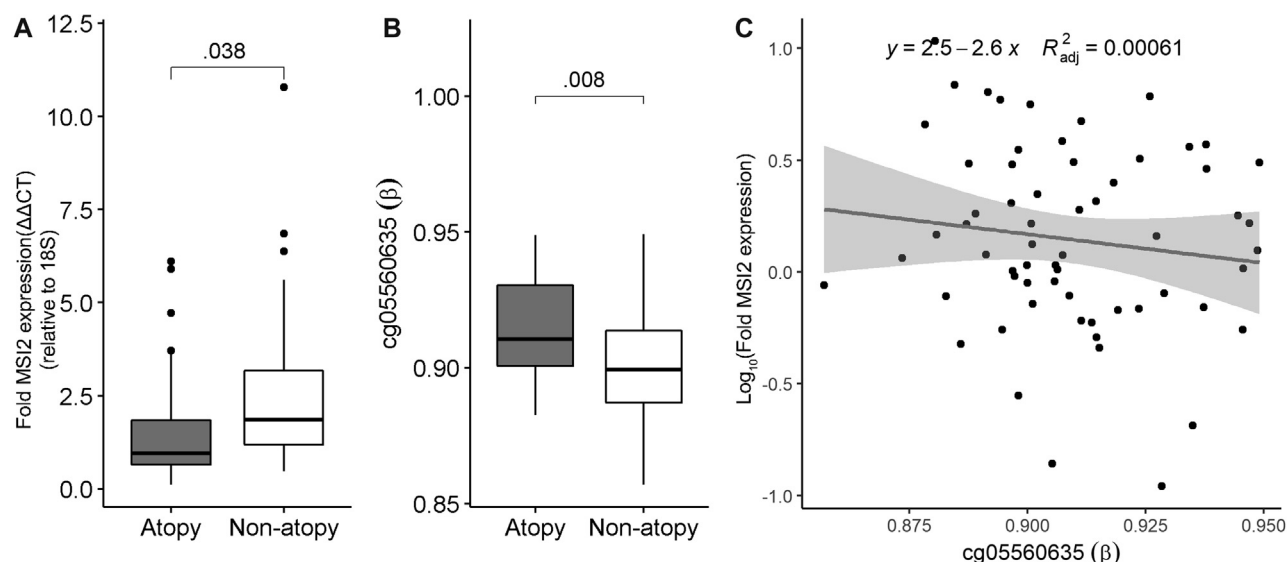


FIG 4. Relationship between the methylation signature and gene expression in *MSI2*. **A**, The expression levels (RT-qPCR) of *MSI2*, which were normalized to those of 18S rRNA (**B**), and the methylation levels of cg05560635 are presented as box plots based on atopy and nonatopy, and the differences between each group are represented ($n = 62$). **C**, The distribution and correlation between the methylation and $\log_{10}$ -transformed expression levels are displayed. All P values shown herein were obtained by using the Student t test.

mite-specific IgE (rs3995694 and rs7077423) levels were significantly associated with *CAMK1D* (see Table E10 in this article's Online Repository at www.jacionline.org).

Association between genetic and epigenetic variations with *MSI2* expression and atopy

RT-PCR experiments revealed that *MSI2* expression is significantly associated with the incidence of atopy ($P = .038$) (Fig 4, A-C). Mediation analysis revealed that cg05560635 methylation levels mediated the association between rs11652629 and atopy (average causal mediation effect [ACME] = 0.200 [$P = .044$]; average direct effect (ADE) = 0.038 [$P = .812$]) (see Table E11 in this article's Online Repository at www.jacionline.org). A nonsignificant ADE indicated no alternative pathway from SNP to atopy for MR. We conducted additional mediation analysis to identify the mediating effect of *MSI2* expression between cg05560635 and atopy (see Table E12 in this article's Online Repository at www.jacionline.org), and we found that methylation directly influences atopy and exerts a nonsignificant effect on *MSI2* expression (ACME = 0.256 [$P = .316$]; ADE = 0.667 [$P = .018$]) (see Table E12). Furthermore, the correlation between cg05560635 and *MSI2* expression was not significant, thus potentially accounting for the nonsignificance of the ACME. However, the sample size was very small, thus potentially leading to a nonsignificant ACME of cg05560635. Furthermore, regression analysis based on PC scores revealed that the other 20 *MSI2* CpGs were highly associated with *MSI2* expression (adjusted $R^2 = 0.394$; F-test P value = .002) (see Table E13 in this article's Online Repository at www.jacionline.org). Furthermore, with no adjustment for *MSI2* expression and SNPs, some PCs were significantly associated with atopy; however, after adjustment for *MSI2* expression, the effect of 20 PCs and cg05560635 became less significant (*MSI2* P value = .039; cg05560635 P value = .046; model

pseudo- R^2 value = 0.532; model P value = .002) (Table II and see Tables E14 and E15 in this article's Online Repository at www.jacionline.org). These results indicate that changes in *MSI2* methylation levels lead to changes in gene expression levels, which may be significantly associated with atopy.

Global effect of genetic and epigenetic effects on phenotypes

PC component analysis of DNA methylation and genotype was used to compare the global differences in sample distributions between cases and controls; however, no marked differences were observed for either phenotype (see Fig E2 in this article's Online Repository at www.jacionline.org).

To determine the relationship between the effects of SNPs and DNA methylation on the phenotypes, we represented the meQTL pairs as a function of effect differences in accordance with the allelic effect (Fig 5, A and C). For 723 CpG-SNP pairs derived from meQTL on atopy analysis (1784 pairs in tissue eosinophilia analysis), we examined the association between the effect of the SNPs on the target phenotype and on methylation levels through kernel density estimation. Consequently, the absolute value of the CpG effects tended to be greater than that of the SNPs (Fig 5, B and D). Moreover, these 2 effects were inversely correlated in atopy; the SNPs exerting markedly positive effects on atopy displayed marked negative effects on methylation levels and *vice versa*.

A prediction model was generated with SNPs, CpGs, and the clinical variables for atopy and tissue eosinophilia, and we compared the McFadden R^2 improvement in accordance with each feature based on a test set with 5-fold cross-validation. Although most proportions remained unexplained, variations of 17.31% on atopy analysis and 23.15% on tissue eosinophilia analysis were explained by the full model. Moreover, on atopy

TABLE II. Logistic regression analysis of the effects of features in *MSI2* on the atopy

Model	Odds ratio	95% CI	P value
Model 1*			
Atopy ~ CpG (cg05560635) + 20 PCs†			
CpG (cg05560635)	9.24×10^{18}	$3.17 - 6.29 \times 10^{42}$	.072
Principal components 1	4.46×10^{-03}	3.33×10^{-05} to 1.17×10^{-01}	.007
Model 2			
Atopy ~ Expression‡ (<i>MSI2</i>)	0.13	2.27×10^{-2} to 5.46×10^{-1}	.009
Model 3*			
Atopy ~ 1 CpGs + 20 PCs + Expression (<i>MSI2</i>)‡			
CpG (cg05560635)	6.95×10^{30}	2.45×10^{-45} to 1.64×10^{18}	.046
Expression‡ (<i>MSI2</i>)	6.45×10^{-3}	1.76×10^{-5} to 3.72×10^{-1}	.039
Principal components 5	5.92×10^{12}	1.09×10^3 to 3.67×10^{26}	.026

*Significant features. Details are summarized in Tables E13 and E15.

†PC values obtained from β -values of 20 CpG sites that are not related to genetic factors.

‡Expression level of *MSI2* was $\log_{10}$ -transformed.

and tissue eosinophilia analyses, CpGs, SNPs, and clinical variables accounted for variations of 11.66% and 18.09%, 5.48% and 4.93%, and 0.17% and 0.13%, respectively. Variations in CpGs yielded better predictions than SNPs or clinical variables for both phenotypes, although the prediction performance was not good with respect to global methylation patterns.

DISCUSSION

Our study describes airway epigenetic signatures associated with atopy and tissue eosinophilia through an integrative approach, including DMP identification and functional analyses with DMP-enriched genes, meQTL analysis, instrumental variable analysis, and variance estimation, along with replication in an independent cohort. This study revealed a novel candidate gene, *MSI2*, associated with atopy through genetic-epigenetic interactions and another candidate gene, *CAMK1D*, associated with tissue eosinophilia.

Classical EWAS using nasal polyps identified 2 genes, *GSTT4* and *PPP5D1*, to be associated with atopy. After expansion of the criteria for methylation differences, more than 100 DMP-enriched genes were identified; they included *MSI2*, *ACOT7*, and *JARID2*. *MSI2* participates in the IL-6, TGF- β , eukaryotic initiation factor 2, hepatocyte growth factor, and epidermal growth factor signaling pathways.^{18,19} Recent blood EWASs have reported associations between *MSI2* methylation with diabetes²⁰ and obesity-related traits,²¹ which are suggested to contribute to allergic diseases.²² *ACOT7* is upregulated by lipopolysaccharide and modifies prostaglandin D2 and E2 biosynthesis.²³ The epigenetic characteristics of nasal *ACOT7* are associated with asthma- and allergy-related traits.²⁴ Furthermore, previous blood EWASs have reported associations between CpG sites at *ACOT7* and total serum IgE.^{8,9} *JARID2* upregulation occurs during T-cell receptor stimulation and during positive selection in the thymus. *JARID2*-deficient inducible natural killer T cells perturb T_H17 cell differentiation, leading to a reduced T_H17 cell-driven autoimmune pathology.²⁵ The GO terms for overrepresented genes in the methylation profiles displayed an increased proportion of hypomethylated genes, potentially indicating the upregulation of these immune-related genes. *MSI2*, *JARID2*, *TRIO*, and *ACOT7* contained enriched DMPs interacting to affect atopy at the genetic level.

We identified a *cis*-meQTL SNP, rs11652629, of cg05560635 at *MSI2* to estimate effect of *MSI2* methylation on atopy. A recent study estimated that almost 20% of reliably assayed variation in blood DNA methylation is heritable and that 50% of CpG sites showed evidence of a significant genetic component.²⁶ The minor allele frequencies of rs11652629 are 0.3359 and 0.17 for East Asians and non-Finnish Europeans, respectively, indicating a clear difference among different populations.²⁷ A robust correlation between rs11652629 and cg05560635 indicates that cg05560635 methylation levels may differ at the population level. Furthermore, the rs11652629 SNP facilitated MR and mediation analysis. Those analyses suggest that changes in *MSI2* expression may be an intermediate risk factor for atopy, and CpG methylation contributes to changes in *MSI2* expression. MR is potentially based on 3 different assumptions: (1) a robust association between SNPs and CpGs, (2) the absence of an alternative pathway from SNP to atopy, and (3) independence between SNPs and the potential confounders associated with atopy and CpGs. Our results indicate robust, stable correlations between the variant and CpGs ($P = 5.66 \times 10^{-9}$ on linear regression analysis). Mediation analysis based on these 3 variables confirmed that ACME is significant ($P = .044$); however, ADE was not ($P = .828$), indicating no alternative pathway from atopy to CpG, concurrent with the second assumption. Finally, SNPs are generally assumed to be independent, with potential confounders associated with atopy and CpG. However, this assumption may be potentially opposed. For instance, the present SNP potentially harbors a trans me-QTL CpG marker associated with potential confounders, and further evidence is warranted to ensure the causal effect of cg05560635 on atopy. Fortunately, *MSI2* variants were identified on the basis of total serum IgE level, which is a marker of systemic allergic inflammation, from a Costa Rican cohort. Epigenetic alterations in *MSI2* in whole blood²⁸ and nasal samples²⁴ are reportedly associated with asthma in children.

Regarding tissue eosinophilia, classical EWAS using nasal polyps identified 15 genes, including *HLA-DPA1* and *ACOT7*. There were more than 40 DMP-enriched genes, including *IL13*, *RIN3*, *ACOT7*, and *JARID2*, after expansion of the criteria for differences in methylation. *IL13* is a well-known cytokine gene that is central to the regulation of the pathogenesis of atopic

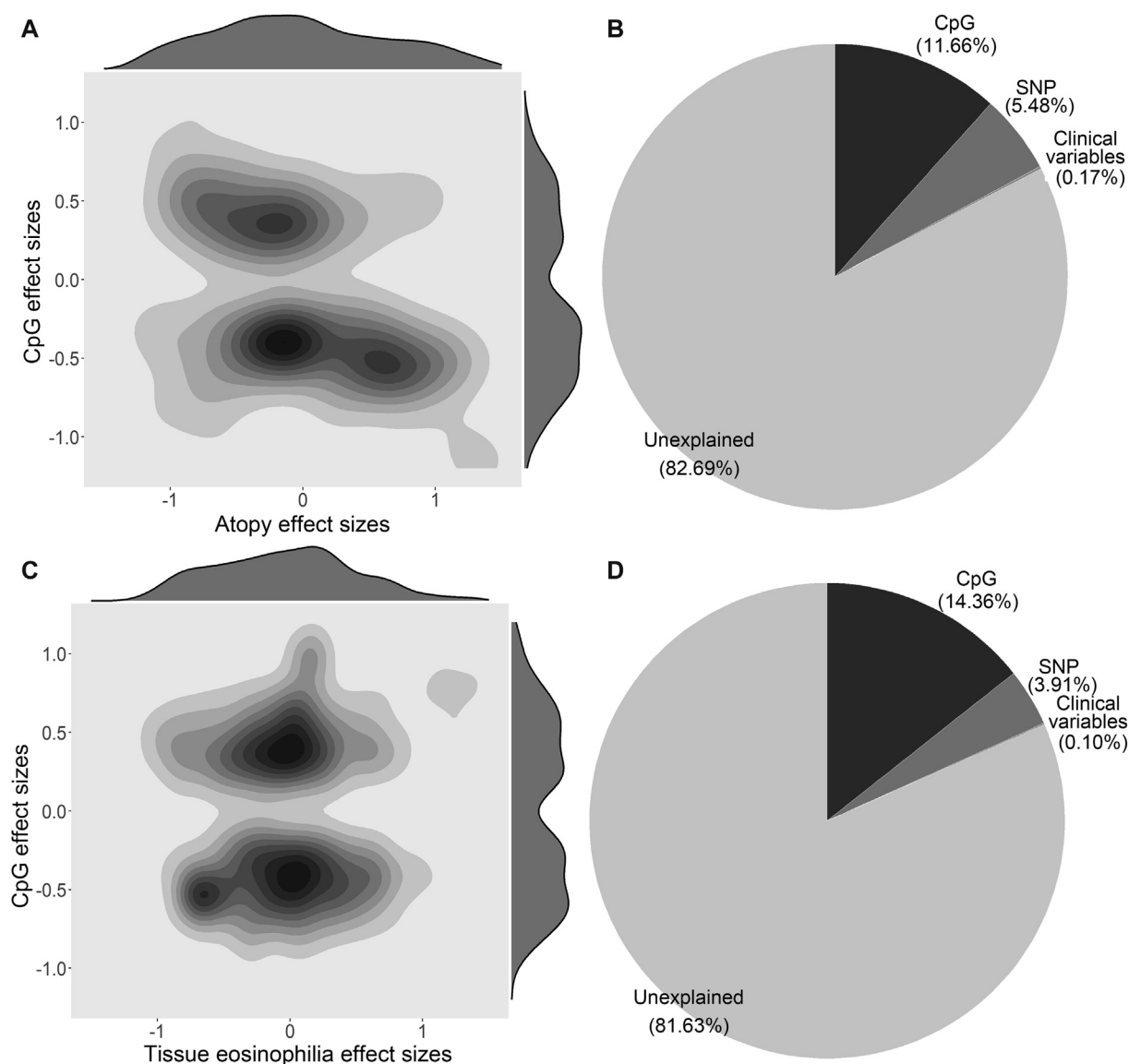


FIG 5. Differences in the distribution of DNA methylation and genotype and its explanatory power for each phenotype. The contour plots show the effect of the alleles derived from meQTLs on phenotype and DNA methylation in atopy (**A**) and tissue eosinophilia (**C**). The associations between β -coefficients from logistic regression (x-axis) and γ -coefficients from the meQTL linear model (y-axis) are plotted on the basis of kernel probability density estimates. The marginal distribution of the 2 variables, logistic β , and meQTL γ is displayed. Percentage variation explained by each data type in the final model in atopy (**B**) and tissue eosinophilia (**D**). For methylation, genotype, and clinical variables, a penalized regression model was applied. The individual proportions of variation are displayed with use of the McFadden R^2 values.

diseases.²⁹ *RIN3*, a negative regulator of mast cell responses to stem cell factors,³⁰ is associated with chronic obstructive pulmonary disease and lung function.^{31,32} Most genes displayed low methylation levels in eosinophilic rather than noneosinophilic polyps, and the highest methylation level was observed in normal nasal tissues, suggesting that the expression of these genes are potentially upregulated in eosinophilic rather than noneosinophilic polyps or normal nasal tissues. Furthermore, GO terms were enriched for hypomethylated genes, thus indicating the up-regulation of these immune-related genes. We also identified 10 genes, including *CAMK1D* and *ACOT7*, with enriched DMPs interacting to affect tissue eosinophilia at the genetic level. These genes exerted a genetic effect mediated by methylation changes.

Among the genes, we identified a direct association between SNPs in *CAMK1D* and tissue eosinophilia. *CAMK1D* is associated with immune responses in dendritic cells.³³ Because the Costa Rican cohort lacked information on nasal tissue eosinophilia, the association with *CAMK1D* variants could not be replicated; however, associations with blood eosinophil, total serum IgE, and dust mite-specific IgE levels were found in the cohort.

GO terms of the identified DMP-enriched genes associated with atopy and tissue eosinophilia encompassed immune-related terms, including innate immunity, adaptive immunity, and their regulation. However, each profile for atopy and tissue eosinophilia comprised different GO terms. The compositions of the sets of immune-related genes associated with atopy and tissue

eosinophilia were comprised quite differently owing to the different set of DMPs identified for each phenotype. They might interact to induce T_H2 cell inflammation, even though each gene plays a unique role in each phenotype. Given that atopy represents systemic allergic inflammation whereas tissue eosinophilia is involved in local allergic inflammation, the configuration of immune responses might differ between systemic and local allergic inflammation.

Herein, we have identified significant genes for systemic (atopy) and local (tissue eosinophilia) allergic inflammation by using EWAS combined with GWAS, despite the limited statistical power of EWAS *per se*. Results from both the EWAS and GWAS tended to display small effect sizes and account for only a small proportion of allergic inflammation. Therefore, we extended the criteria for methylation differences to encompass “low-hanging fruit”³⁴ and found epigenetic signatures that interact with genetic effects. Integrated GWAS and EWAS analyses, which often have complementary and synergistic outcomes, potentially reflect several significant but small signals and provide novel biologic insight into allergic inflammation. We believe that unlike analysis of a single data set that primarily focuses on comparing the most significant hits among studies, our integrated study elucidates the complex associations within and between these data sets. Of note, the larger proportion of the epigenetic markers compared with the genetic markers identified through GWAS in the joint prediction model potentially indicates that epigenetic changes due to environmental effects are better associated with atopy than with genetic effects. And the large unexplained proportion is probably due to the fact that atopy is extremely sensitive to environmental factors. In addition, interactions between genes, gene-environment interactions, and genome-epigenetic interactions, which are unexplored factors in this study, may also be the causes.³⁵ To examine these factors, it will be necessary to yield more various omics data using larger samples. An understanding of the underlying associations is essential for understanding the pathogenesis of systemic and local allergic inflammation.

Indeed, nasal polyps comprise different cell types. Theoretically, differences in the cellular composition of polyps would affect the outcomes of DNA methylation analysis. However, differences in the cell type composition of polyps do not affect the validity of the analysis if the magnitude of the difference is similar between cases and controls. Furthermore, we have replicated the significant genes in an independent cohort and confirmed differential gene expression to validate the association with atopy. Herein, we have focused on understanding the pathogenesis of allergic inflammation, and not on formation of the nasal polyp. To that end, we eliminated the effect of the morphologic and molecular phenomena associated with polyp formation by using normal nasal tissue from subjects without allergic sensitization, whose methylation pattern revealed a cluster that was distinct from those of nasal polyps.

To our knowledge, this is the first integrated EWAS and GWAS focused on systemic and local allergic inflammation. EWAS can be affected by reverse causation, and our results reinforce a causal relationship between methylation and allergic inflammation. Furthermore, we have identified novel candidate genes and focused on the role of *MSI2* methylation in causing systemic allergic inflammation with its decreased expression. Certain

genes appear to be more associated with either systemic or local allergic inflammation; however, it is expected that a harmonized effect of these genes contributes to immune responses. Finally, although the epigenetic signature is mediated by genetic effects, it can explain both systemic and local allergic inflammation better than the genetic signature can.

Key messages

- A genetic variant resulted in an alteration of methylation level at the CpG site, which regulated *MSI2* expression in the airway of patients with atopy.
- The epigenetic signature is more important than the genetic signature in predicting allergic inflammation, notwithstanding numerous important unknown features.

REFERENCES

- Bonini S, Rasi G, Torre A, D'Amato M, Matricardi PM. The heterogeneity of allergic phenotypes: genetic and environmental interactions. *Ann Allergy Asthma Immunol* 2001;87:48-51.
- Ferreira MA, Vonk JM, Baurecht H, Marenholz I, Tian C, Hoffman JD, et al. Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. *Nat Genet* 2017;49:1752-7.
- Fulkerson PC, Rothenberg ME. Targeting eosinophils in allergy, inflammation and beyond. *Nat Rev Drug Discov* 2013;12:117-29.
- Arasi S, Pajno GB, Lau S, Matricardi PM. Local allergic rhinitis: a critical reappraisal from a paediatric perspective. *Pediatr Allergy Immunol* 2016;27:569-73.
- Chawes BL, Kreiner-Moller E, Bisgaard H. Objective assessments of allergic and nonallergic rhinitis in young children. *Allergy* 2009;64:1547-53.
- Kim KW, Ober C. Lessons learned from GWAS of asthma. *Allergy Asthma Immunol Res* 2019;11:170-87.
- Peng C, Van Meel ER, Cardenas A, Rifas-Shiman SL, Sonawane AR, Glass KR, et al. Epigenome-wide association study reveals methylation pathways associated with childhood allergic sensitization. *Epigenetics* 2019;14:445-66.
- Chen W, Wang T, Pino-Yanes M, Forno E, Liang L, Yan Q, et al. An epigenome-wide association study of total serum IgE in Hispanic children. *J Allergy Clin Immunol* 2017;140:571-7.
- Ek WE, Ahsan M, Rask-Andersen M, Liang L, Moffatt MF, Gyllenstein U, et al. Epigenome-wide DNA methylation study of IgE concentration in relation to self-reported allergies. *Epigenomics* 2017;9:407-18.
- Reese SE, Xu CJ, den Dekker HT, Lee MK, Sikdar S, Ruiz-Arenas C, et al. Epigenome-wide meta-analysis of DNA methylation and childhood asthma. *J Allergy Clin Immunol* 2019;143:2062-74.
- Yi ACA, Tay TR, Choo XN, Koh MSY, Tee AKH, Wang DY. Precision medicine in united airways disease: a “treatable traits” approach. *Allergy* 2018;73:1964-78.
- van Leeuwen EM, Kanterakis A, Deelen P, Kattenberg MV. Genome of the Netherlands Consortium, Slagboom PE, et al. Population-specific genotype imputations using minimac or IMPUTE2. *Nat Protoc* 2015;10:1285-96.
- Zaghdoudi T. ivprobit: An R package to estimate the instrumental variables probit model. *The Journal of Open Source Software* 2018;3:523.
- Song YE, Lee S, Park K, Elston RC, Yang HJ, Won S. ONETOOL for the analysis of family-based big data. *Bioinformatics* 2018;34:2851-3.
- Gim J, Kim W, Kwak SH, Choi H, Park C, Park KS, et al. Improving disease prediction by incorporating family disease history in risk prediction models with large-scale genetic data. *Genetics* 2017;207:1147-55.
- Gim J, An J, Sung J, Silverman EK, Cho MH, Won S. A between ethnicities comparison of chronic obstructive pulmonary disease genetic risk. *Front Genet* 2020;11:329.
- Hunninghake GM, Soto-Quiros ME, Avila L, Ly NP, Liang C, Sylvia JS, et al. Sensitization to *Ascaris lumbricoides* and severity of childhood asthma in Costa Rica. *J Allergy Clin Immunol* 2007;119:654-61.
- Duggimpudi S, Kloetgen A, Maney SK, Munch PC, Hezaveh K, Shaykhalishahi H, et al. Transcriptome-wide analysis uncovers the targets of the RNA-binding protein *MSI2* and effects of *MSI2*'s RNA-binding activity on IL-6 signaling. *J Biol Chem* 2018;293:15359-69.

19. Park SM, Deering RP, Lu Y, Tivnan P, Lianoglou S, Al-Shahrour F, et al. Musashi-2 controls cell fate, lineage bias, and TGF-beta signaling in HSCs. *J Exp Med* 2014;211:71-87.
20. Jeon JP, Koh IU, Choi NH, Kim BJ, Han BG, Lee S. Differential DNA methylation of *MSI2* and its correlation with diabetic traits. *PLoS One* 2017;12: e0177406.
21. Dhana K, Braun KVE, Nano J, Voortman T, Demerath EW, Guan W, et al. An epigenome-wide association study of obesity-related traits. *Am J Epidemiol* 2018;187:1662-9.
22. Perez MK, Piedimonte G. Metabolic asthma: is there a link between obesity, diabetes, and asthma? *Immunol Allergy Clin North Am* 2014;34:777-84.
23. Forwood JK, Thakur AS, Guncar G, Marfori M, Mouradov D, Meng W, et al. Structural basis for recruitment of tandem hotdog domains in acyl-CoA thioesterase 7 and its role in inflammation. *Proc Natl Acad Sci U S A* 2007;104: 10382-7.
24. Cardenas A, Sordillo JE, Rifas-Shiman SL, Chung W, Liang L, Coull BA, et al. The nasal methylome as a biomarker of asthma and airway inflammation in children. *Nat Commun* 2019;10:3095.
25. Pereira RM, Martinez GJ, Engel I, Cruz-Guilloty F, Barboza BA, Tsagaratou A, et al. *Jarid2* is induced by TCR signalling and controls iNKT cell maturation. *Nat Commun* 2014;5:4540.
26. McRae AF, Powell JE, Henders AK, Bowdler L, Hemani G, Shah S, et al. Contribution of genetic variation to transgenerational inheritance of DNA methylation. *Genome Biol* 2014;15:R73.
27. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020;581:434-43.
28. Xu CJ, Soderhall C, Bustamante M, Baiz N, Gruziova O, Gehring U, et al. DNA methylation in childhood asthma: an epigenome-wide meta-analysis. *Lancet Respir Med* 2018;6:379-88.
29. Gandhi NA, Pirozzi G, Graham NMH. Commonality of the IL-4/IL-13 pathway in atopic diseases. *Expert Rev Clin Immunol* 2017;13:425-37.
30. Janson C, Kasahara N, Prendergast GC, Colicelli J. RIN3 is a negative regulator of mast cell responses to SCF. *PLoS One* 2012;7:e49615.
31. Cho MH, McDonald ML, Zhou X, Mattheisen M, Castaldi PJ, Hersh CP, et al. Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *Lancet Respir Med* 2014;2:214-25.
32. Lutz SM, Cho MH, Young K, Hersh CP, Castaldi PJ, McDonald ML, et al. A genome-wide association study identifies risk loci for spirometric measures among smokers of European and African ancestry. *BMC Genet* 2015;16:138.
33. Lin J, Chen YT, Xia J, Yang Q. MiR674 inhibits the neuraminidase-stimulated immune response on dendritic cells via down-regulated Mbnl3. *Oncotarget* 2016; 7:48978-94.
34. Ober C. Asthma genetics in the Post-GWAS Era. *Ann Am Thorac Soc* 2016;13: S85-90.
35. Potaczek DP, Harb H, Michel S, Alhamwe BA, Renz H, Tost J. Epigenetics and allergy: from basic mechanisms to clinical applications. *Epigenomics* 2017;9: 539-71.

Supplementary Texts

METHODS

Clinical evaluation

Atopy was defined by total serum IgE level of greater than 150 kU/L or specific IgE levels of greater than 0.7 kUA/L to at least 1 of the 6 common allergens: egg white, milk, *Dermatophagoides pteronyssinus*, *Dermatophagoides farinae*, *Alternaria* species, or *Blattella germanica*. Non-atopy was defined by both total serum IgE level of less than 100 kU/L and specific IgE levels of less than 0.35 kUA/L to all the allergens above. Tissue eosinophilia was defined as ≥ 70 eosinophils/HPF.^{1,2}

Sample compositions

We enrolled 134 patients with nasal polyps which require removal, including 59 atopic and 52 non-atopic patients, while 23 patients with ambiguous phenotypes were excluded. We considered atopy and eosinophilia as response variables, and 111 and 134 patients were considered for atopy analyses and eosinophilia analyses, respectively. Out of the 134 patients, 59 patients had eosinophilic nasal polyps and 73 patients had non-eosinophilic nasal polyps. Two patients whose tissue eosinophilia could not be confirmed were excluded. Nasal polyp tissues were obtained during endoscopic sinus surgery and the freshly excised tissue was contained within the transfer media for prompt delivery to the laboratory. Blood samples were also collected from each patient. The blood derived genomic DNA was genotyped and the tissue derived genomic DNA was bisulfite converted. We used 18 normal nasal tissues from patients with a negative history of both allergic diseases and allergic sensitization who underwent septoplasty with turbinoplasty because of anatomical problems. All patients were recruited from the Otorhinolaryngology clinic of Severance Hospital. Patients with an established immunodeficiency, pregnancy, coagulation disorder, or diagnosis of aspirin hypersensitivity, classic allergic fungal sinusitis, eosinophilic granulomatous polyangiitis (Churg-Strauss syndrome), or cystic fibrosis were excluded from the study.

DNA methylation profiling and differences in atopy and tissue eosinophilia

Genomic DNA was bisulfite converted using Zymo kit according to the manufacturer's protocol. The methylation levels of the converted DNA were measured using the Infinium MethylationEPIC BeadChip assay (Illumina, Inc., CA, USA). More than 850,000 CpG sites in the EPIC array were measured to obtain the differential methylation profile of each sample groups. Preprocessing and

normalization were performed to maximize sensitivity for detection of differences between interest groups. Firstly, background correction and dye bias equalization were conducted with lumi R package³. Secondly, samples determined as low quality or outlier from meffil R package were excluded⁴. Next, biologically or technically biased probes were removed by the following criteria: (i) located on the Sex chromosomes, X and Y, (ii) detection P-value > 0.05 in at least 25% of samples (iii) associated with single nucleotide polymorphisms (SNPs) for any minor allele frequency (iv) identified as low-quality CpG site at least once from three R/Bioconductor packages (meffil, ENmix, and wateRmelon)^{5,6}. After QC, normalization of the DNA methylation data was performed used BMIQ methods. Finally, 808,934 methylation sites from 126 subjects for atopy analysis and 808,934 methylation sites from 147 subjects for tissue eosinophilia analysis remained for further analysis.

To identify methylated sites associated with atopy and tissue eosinophilia, we applied one-way ANOVA and post hoc Tukey's honest significance difference (HSD) test on M-values of preprocessed methylation probes. We also identified differentially methylated position (DMP)-enriched genes in each phenotype by carrying out a Fishers' exact test for all genes that contained at least five methylation probes, at the BH adjusted 0.05 significance level. Last, gene ontology (GO)-based gene set enrichment analysis was performed to identify the biological processes harboring differentially methylated probes.

Identification of differentially methylated positions

We used nasal polyp tissues to investigate the differences of genome-wide methylation in airway between atopy (case) and non-atopy (control) and between eosinophilia (case) and non-eosinophilia (control) in tissue. Their amount of methylation was also compared with normal nasal tissue to eliminate the effect of the morphological and molecular processes associated with polyp formation. We conducted ANOVA to compare differentially methylated positions (DMPs) among nasal polyp tissues of patients with and without non-atopy and normal nasal tissue. The multiple testing problem of ANOVA analyses was adjusted by Benjamini Hochberg (BH) method. The Tukey's HSD test was conducted for the post-hoc analyses of the probes showing the adjusted *P* values < 0.05. From the Tukey's HSD test results, two cases of combinations were chosen for analysis. The first case contains DMPs with polyp formation related effects, comprising the intersection of DMPs with significant differences (Tukey's HSD *P* value < 0.05) between the case group and normal group, control group and normal group, case group and control group. The second group of DMPs are independent of polyp formation, consisting DMPs with significant differences between the case group and control group but not between the case group and normal group or the control group and normal group. Here, we also considered only probes with a difference of mean β value between case and control at least 1%.

Identification of DMP-enriched genes and gene ontology-based enrichment analysis.

DMP-enriched genes were identified in each phenotype by carrying out a Fishers' exact test for all genes that contain at least five methylation probes, at the BH adjusted 0.05 significance level. The test was conducted on previously determined probes by ANOVA and Tukey test without considering the direction of methylation level. Among these results, we have also selected only those genes with a clear direction of methylation pattern by calculating the ratio of hyper-methylation to hypo-methylation. The ratio was to be greater than 0.9 or less than 0.1. Next, gene ontology (GO)-based gene set enrichment analysis was performed to identify the biological processes harboring differentially methylated probes. With gometh function in missMethyl R package, a hypergeometric test was used for GO term pathway enrichment⁷. The entire methylation probes were used as background, and previously selected DMPs were tested for functional analyses. The enriched GO terms containing overrepresentation in the genes were selected based on BH adjusted P value < 0.05 . For each enriched pathway, the contribution of hyper- and hypo-DMPs was calculated separately. When the genes included both hyper- and hypo-DMPs, they were categorized as overlaps.

Quality control and imputation of genotype data

The overall quality check and evaluation of genotype data were carried out using PLINK. The SNPs was excluded from the analysis if any of the following criteria were met: (i) they were on sex chromosome, multi-allelic, or INDELs (ii) Hardy–Weinberg equilibrium P -value $< 10^{-5}$, (iii) genotype call rates were less than 95%, or (iv) minor allele frequencies were < 0.05 . Also any subjects whose missing genotype rates were $> 5\%$, or showed identity-by-state $> 80\%$ were discarded. This preprocessing steps resulted in 107 individuals for atopy analysis and 128 individuals for tissue eosinophilia analysis with high-quality genotype data for 283,085 SNPs. To improve the power and resolution of associations, genome-wide imputation was performed for all autosomes with IMPUTE2 v.2.32⁸ by using all participants' genotypes for 1000 Genome project phase 3 as reference.

Mapping of methylation quantitative trait loci

For each methylation site, we investigated statistically significant associations between methylation levels and nearby genetic variants. Subjects that are missing either methylation or genotype data were excluded. Then β -values of methylation level were converted to M-values through a logit transformation, and the standard additive linear model was used for testing an association between CpG and SNP.⁹ In addition to adjusting for age and sex, we included the first 10 principal component (PC) scores from the genotype data. We only considered local associations in CpG-SNP pairs, which were highly located

(within ± 500 k base pairs). For the associations, CpG-SNP pairs were considered significant when passing the 5% FDR threshold on mappings. We further investigated pairs occurring among DMP-enriched genes to clarify CpGs with a strong signal for the gene and elucidate causality. Trans-methylation quantitative trait loci (meQTL) analysis was not performed due to the possibility of finding of false positive association of CpG-SNP pairs and the insufficiency of the sample size to supply adequate power.

One-sample Mendelian randomization

Association between CpGs and phenotypes can be found by the reverse causation, and to reveal causal effect of methylation on each phenotype, instrumental variable analysis was performed using IV regression from ivprobit R package that compute the instrumental variables probit model with generalized least squares estimator¹⁰. We conducted the IV analysis for the only significant meQTL pairs on the DMP-enriched genes, and the methylation M-value was used as endogenous variables. Here, the probit model was adjusted by age, sex and the first 10 genetic principal components.

Replication with Costa Rica cohort data

The previous reports described the detailed whole procedure of subject screening, enrollment and phenotyping protocols¹¹. Whole genome sequencing data were produced as the part of the NHLBI Trans-Omics for Precision Medicine (TOPMed) Whole Genome Sequencing Program. Details regarding the laboratory methods, data processing and quality control are described on the TOPMed website (see URL https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000988.v3.p1) and in documents included in TOPMed accession released on the database of Genotypes¹². A total of 3,773 subjects were available and we calculated their genome-wide identity-by-descent (IBD) estimates using KING (Kinship-Based Inference for Genome-wide Association Studies).

Quantitative real-time PCR for MSI2 expression

Total RNA was extracted from nasal polyp using TRIzol™ Reagent (Invitrogen, Carlsbad, USA) followed by complementary DNA synthesis using the ReverTra Ace® qPCR RT Master Mix Kit (Toyobo Co., Ltd., Osaka, Japan). Quantitative real-time PCR was then performed on *StepOnePlus™ Real-Time PCR System* (Applied Biosystems) with TaqMan™ Fast Advanced Master Mix (Applied Biosystems, Foster City, USA). Customized TaqMan primer/probes for MSI2

(Hs00292670_m1) and 18s rRNA (Hs99999901_s1) (Applied Biosystems) were used. MSI2 expression was normalized to 18s rRNA and the levels were calculated by $\Delta\Delta CT$ method.

Prediction model building with SNPs, CpGs and clinical variables for atopy and tissue eosinophilia

To build the prediction model with SNPs, CpGs and four clinical variables including age, sex, total serum IgE level, and blood eosinophils, we applied the penalized regression model, and calculated the deviance improvement by comparing the predicted and true outcomes. Atopy status and tissue eosinophilia were used as response variables respectively. All patients were divided into a training set and a test set by the five-fold cross validation. The approximately same numbers of patients with and without atopy were used for each test and train set, and patients selected for each test set did not overlap. The train set was used to build a predictive model, and the model was applied to predict each phenotype (atopy or tissue eosinophilia) of patients in the test set. Then, by comparing the predicted phenotypes from the test set and true phenotypes, the deviance was obtained. Predicted phenotypes was obtained from test set to minimize the overfitting. We also calculated the deviance when SNPs, CpG or four clinical variables was separately removed, and by comparing those deviances, the deviance improvement for each covariate was estimated using McFadden's R^2 by using the same procedures as Gim et al¹³.

Methylation patterns in immune response related genes

To identify the overall patterns of methylation level according to the immune response, we utilized the reference gene categories from the other study.¹⁴ We considered the 6 known reference immune gene categories (Basal cell signaling, IFN- α , IFN- γ , IL-13, IL-4, and IL-4 and IL-13), and chose genes with more than two differentially methylated positions for each category. The averages of β -values from significant probes in each gene for three groups were calculated.

RESULTS

DNA methylation changes in immune response

We summarized the global pattern of methylation in the categorized reference gene lists to investigate methylation level change in immune response. For the six categories including basal cell, IFN- α , IFN- γ , IL-4, IL-13 and IL-4/IL-13, methylation beta values from the genes in each immune related category were represented as boxplot using the significant DMPs. For atopy, the most of genes were hypomethylated in cases than in controls and normal group (Fig E3). In the case of tissue eosinophilia, a greater number of genes were considered due to more findings at the DMPs, but the methylation patterns tend to be heterogeneous compared to the atopy analysis (Fig E4). Nonetheless, the set of immune related genes which contains different DMPs were quite differently comprised between atopy and tissue eosinophilia. For instance, in both phenotypes, the number of genes containing DMP belonging to the IL13 category was equally seven, but only two of them overlap.

REFERENCES

1. Nakayama T, Yoshikawa M, Asaka D, Okushi T, Matsuwaki Y, Otori N, et al. Mucosal eosinophilia and recurrence of nasal polyps - new classification of chronic rhinosinusitis. *Rhinology* 2011;49:392-6.
2. Tokunaga T, Sakashita M, Haruna T, Asaka D, Takeno S, Ikeda H, et al. Novel scoring system and algorithm for classifying chronic rhinosinusitis: the JESREC Study. *Allergy* 2015;70:995-1003.
3. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* 2008;24:1547-8.
4. Min JL, Hemani G, Davey Smith G, Relton C, Suderman M. Meffil: efficient normalization and analysis of very large DNA methylation datasets. *Bioinformatics* 2018;34:3983-9.
5. Xu Z, Niu L, Li L, Taylor JA. ENmix: a novel background correction method for Illumina HumanMethylation450 BeadChip. *Nucleic Acids Res* 2016;44:e20.
6. Pidsley R, CC YW, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics* 2013;14:293.
7. Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics* 2016;32:286-8.
8. van Leeuwen EM, Kanterakis A, Deelen P, Kattenberg MV, Genome of the Netherlands C, Slagboom PE, et al. Population-specific genotype imputations using minimac or IMPUTE2. *Nat Protoc* 2015;10:1285-96.
9. Shabalín AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* 2012;28:1353-8.
10. Zaghdoudi T. ivprobit: An R package to estimate the instrumental variables probit model. *J. Open Source Software* 2018;3:523.
11. Hunninghake GM, Soto-Quiros ME, Avila L, Ly NP, Liang C, Sylvia JS, et al. Sensitization to *Ascaris lumbricoides* and severity of childhood asthma in Costa Rica. *J Allergy Clin Immunol* 2007;119:654-61.
12. Regier AA, Farjoun Y, Larson DE, Krasheninina O, Kang HM, Howrigan DP, et al. Functional equivalence of genome sequencing analysis pipelines enables harmonized variant calling across human genetics projects. *Nat Commun* 2018;9:4038.
13. Gim J, Kim W, Kwak SH, Choi H, Park C, Park KS, et al. Improving Disease Prediction by Incorporating Family Disease History in Risk Prediction Models with Large-Scale Genetic Data. *Genetics* 2017;207:1147-55.
14. Ordovas-Montanes J, Dwyer DF, Nyquist SK, Buchheit KM, Vukovic M, Deb C, et al. Allergic inflammatory memory in human respiratory epithelial progenitor cells. *Nature* 2018;560:649-54.

FIGURE LEGENDS

Figure E1. Genomic location of differentially methylated probes in DNA methylation profiles.

Each pie chart represents the percentage of hypo-methylated and hyper-methylated probes in (A) atopy and (B) tissue eosinophilia. Top, the percentage at protein coding region. Middle, the percentage at gene level (TSS1500, 200–1,500 bases upstream of transcription start site; TSS200, 0–200 bases upstream of transcription start site; 5' untranslated regions (UTR); first exon; Body; ExonBnd, within 20 bases of an exon boundary; and 3'UTR). Bottom, the percentage across different genomic regions (CpG islands; Shore, probes located less 2 kb from CpG islands; Shelf, probes located over 2 kb from CpG islands; and open sea).

Figure E2. The distribution of DNA methylation and genotype data. Multidimensional scaling (Euclidian distance) of the DNA methylation profiles and genetic variants for all individuals. The two MDS plots show global landscape of methylation and genetic variants by atopy (A, B) and tissue eosinophilia (C, D). Circles and triangles in the top panels represent individuals. The colors indicate each phenotype group; non-atopy or non-eosinophilic tissue (white), and atopy or eosinophilic tissue (black).

Figure E3. Differential methylation contribution of the immune-related gene set in atopy. For six categorized reference genes (basal cell; genes induced by IL-4, IL-13, IL-4/IL-13, IFN- α , and IFN- γ), the methylation levels in genes with at least three DMPs show the differences between groups.

Figure E4. Differential methylation contribution of the immune-related gene set in tissue eosinophilia. In the same manner as atopy, the methylation levels of genes contain at least three DMPs from tissue eosinophilia are shown in six pre-defined gene list (basal cell; genes induced by IL-4, IL-13, IL-4/IL-13, IFN- α , and IFN- γ).