

Letter to the Editor

Epigenome-wide association study links site-specific DNA methylation changes with cow's milk allergy

To the Editor:

IgE-mediated cow's milk allergy (CMA), which affects 2% to 3% of young children, is among the most common types of food allergy. The molecular mechanisms underlying the development, persistence, and resolution of CMA remain largely unknown. Several recent studies have suggested that epigenetic alterations are involved in allergy development.^{1,2} DNA methylation (DNAm), a type of epigenetic mechanism, regulates gene expression. Alterations in DNAm play a key role in T-cell differentiation and in maintaining T_H1/T_H2 balance,^{1,2} and thus may offer a potential mechanistic explanation for the natural history of CMA. However, to date, research on the epigenetics of food allergy,³ or CMA in particular,⁴ has been very limited. Most of the available studies are based on small sample sizes (<60 children), and were not replicated in an independent sample.

We performed the first epigenome-wide association study of CMA in US children using a 2-stage approach. In the discovery stage, we measured DNAm at 485,512 genomic loci in whole blood samples from 106 white children with CMA (cases) and 76 nonallergic and nonatopic whites (controls) using the Illumina HumanMethylation450 arrays. All the cases and controls were enrolled by the Chicago Food Allergy Study using identical protocols. We then sought to validate the top significant DNAm loci in 2 independent replication samples including (1) childhood whole blood samples of 25 white children (5 CMA cases and 20 controls) from the same Chicago Food Allergy Study (Chicago replication sample) and (2) cord blood samples from 140 black children (8 cases and 132 controls) from the Boston Birth Cohort (Boston replication sample) (see this article's [Methods](#) section in the Online Repository at www.jacionline.org). Population characteristics of the discovery and replication samples are presented in [Tables E1](#) and [E2](#) in this article's Online Repository at www.jacionline.org, respectively.

CMA cases were defined if the children met the following criteria: (1) a convincing history of symptoms indicative of an allergic reaction within 2 hours of ingestion of cow's milk and (2) clear evidence of *sensitization* defined as having a specific IgE level of 0.35 kU/L or more to cow's milk and/or a positive skin prick test result to cow's milk with a mean wheal diameter of 3 mm or more greater than that with the saline control. After quality control steps (see this article's [Methods](#) section), 435,642 CpG sites were available for subsequent data analyses. We fit a linear regression model for ComBat-transformed M value at each CpG site as a function of CMA status, adjusting for potential confounders (see this article's [Methods](#) section). Bonferroni correction was applied to account for multiple testing ($P < 1.15 \times 10^{-7}$).

In the discovery stage, we identified 575 significant autosomal differentially methylated positions (DMPs) (469 located within 385 genes and 106 intergenic); 568 hypomethylated and 7 hypermethylated in CMA cases compared with controls ($P < 1.15 \times 10^{-7}$; [Fig 1](#)). Most of these 575 DMPs showed a small DNAm change between cases and controls, and only 5 DMPs had an absolute mean DNAm difference of 5% or more ([Fig 1](#) and [Table I](#)). On the X-chromosome, DMP cg16737869

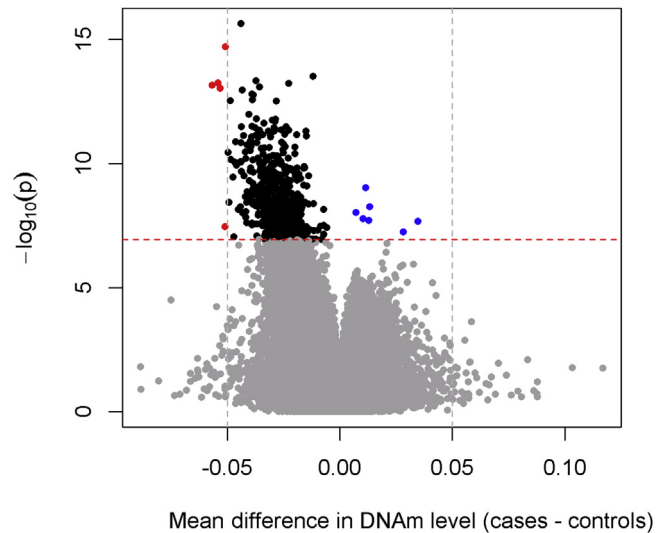


FIG 1. Volcano plot showing the significance and magnitude of autosomal DNAm changes associated with CMA. Mean differences in DNAm (cases-controls) are plotted on the x-axis and corresponding $-\log_{10}(P \text{ value})$ s are plotted on the y-axis. Each point represents a measured CpG site. *Black* and *blue* points represent CpG sites significantly hypomethylated and hypermethylated in CMA cases, respectively. *Red* points represent those significant CpG sites with DNA methylation difference of more than 5% between cases and controls. The *red dashed horizontal line* represents the genome-wide significance threshold ($P = 1.15 \times 10^{-7}$).

met our P value significance threshold, but with a modest DNAm change between cases and controls (2%).

In parallel with individual DMP analyses, we performed a genome-wide screen to identify differentially methylated regions (DMRs) using a “bump hunting” approach. A marginally significant DMR in the 3' untranslated region of the *Enah/Vasp-like* (*EVL*) gene was identified, which was hypomethylated in CMA cases compared with controls (family-wise error rate = 0.085; see [Fig E1, A](#), in this article's Online Repository at www.jacionline.org). This DMR included 5 highly correlated DMPs ($r > 0.8$; [Fig E1, B](#)) that were all significantly associated with CMA at $P < 1.15 \times 10^{-7}$, 2 of which having an absolute mean DNAm difference of more than 5% between cases and controls.

We then performed KEGG pathway-enrichment analysis using WebGestalt on the 386 annotated genes that contained at least 1 significant DMP associated with CMA. Three enriched pathways were identified—“Butirosin and neomycin biosynthesis,” “Starch & sucrose metabolism,” and “Fructose and mannose metabolism”—all of which shared 3 genes (*hexokinase 1-3*) that are critical to glucose metabolism (see [Table E3](#) in this article's Online Repository at www.jacionline.org). This result supports our previous findings on the significant associations between gestational diabetes and increased risk of food sensitization.⁵

Among the 576 DMPs associated with CMA, 10 loci meeting the following criteria were selected for replication: (1) significant DMPs with an absolute DNAm difference of more than 5% ($n = 5$) and/or (2) significant DMPs that annotated to genes relevant to the T_H1-T_H2 pathway ($n = 5$) ([Table I](#)). In the Chicago replication sample, 7 DMPs were validated for their associations with CMA at a false-discovery rate (FDR) of less than 0.05

TABLE I. The top significant differentially methylated positions and their associations with CMA in the discovery and replication samples

DMPs	CHR	Gene	Location	Chicago discovery		Chicago replication		Boston replication	
				β_{diff}^*	P value †	β_{diff}^*	P value ‡	β_{diff}^*	P value §
DMPs with absolute DNAm difference of $\geq 5\%$ in cases vs controls									
cg11770323	13	<i>NDFIP2</i>	Intron	−5.7	2.0×10^{-15}	−4.0	.005	−0.9	.811
cg16409452	14	<i>EVL</i>	3'UTR	−5.4	5.6×10^{-14}	−7.3	.042	−1.4	.322
cg18550847	14	<i>EVL</i>	3'UTR	−5.3	9.6×10^{-14}	−7.4	.034	−1.9	.187
cg09377531	8	<i>TRAPPC9</i>	Intron	−5.1	7.0×10^{-14}	−7.8	.004	−1.5	.260
cg15090899	6	<i>RPS6KA2</i>	Exon	−5.1	3.6×10^{-08}	−7.1	.384	−2.1	.811
DMPs with annotated genes relevant to T _H 1-T _H 2 pathway									
cg16386158	2	<i>IL1RL1</i>	TSS1500	−3.8	1.7×10^{-13}	−8.8	.007	−2.4	.041
cg08404225	3	<i>IL5RA</i>	5'UTR	−3.7	9.3×10^{-10}	−5.5	.062	−2.2	.001
cg13316148	2	<i>STAT4</i>	Intron	−3.2	6.3×10^{-09}	−5.2	.013	−0.4	.301
cg26787239	5	<i>IL4</i>	TSS1500	−2.6	6.7×10^{-09}	−4.8	.006	−1.5	.042
cg06040872	17	<i>CCL18</i>	Intron	−3.2	4.0×10^{-09}	−5.2	.018	−0.6	.256

CHR, Chromosome; TSS, transcription start site; UTR, untranslated region.

*Mean methylation difference (%) between cases and controls.

[†] P value was calculated on the basis of linear regression models with ComBat-transformed M value at each DMP as the outcome, adjusted for age, sex, breast-feeding, parental history of food allergy, estimated cell composition, and genetic ancestry.

[‡] P value was calculated on the basis of linear regression model with ComBat-transformed M value at each DMP as the outcome, adjusted for cell composition and age. Further adjustment for sex and parental history of allergy did not significantly change the associations.

[§] P value was calculated on the basis of linear regression model with ComBat-transformed M value at each DMP as the outcome, adjusted for cell composition, sex, parental history of food allergy, and gestational age group.

||FDR < 5%.

(Table I). In the Boston replication sample, only DMP cg08404225 in the *IL5RA* gene was significantly hypomethylated in CMA cases relative to the normal controls after FDR correction (FDR = 0.01). The associations between the 8 validated DMPs and CMA in the replication samples were in the same directions as were identified in the discovery sample (Table I; see Fig E2 in this article's Online Repository at www.jacionline.org).

The biological significance of the 8 validated DMPs was explored with a Web tool, EpiExplorer. As shown in Table E4 in this article's Online Repository at www.jacionline.org, DMP cg11770323 overlapped with the DNaseI hypersensitive site; DMP cg18550847 was within a CpG island; and DMPs cg13316148, cg08404225, and cg26787239 overlapped with the enhancers and with lamina-associated domains. We also found that DMPs cg16386158, cg13316148, cg26787239, and cg11770323 colocalized with the binding sites of CCCTC-binding factor (CTCF) (see Table E4). Overall, these data indicate that DNAm alteration of these loci may be biologically relevant, but further validation is needed.

This study not only confirmed that DNAm alterations in T_H1 - T_H2 balance (ie, those annotated to *ILIRL1*, *IL5RA*, *IL4*, *CCL18*, and *STAT4* genes) are associated with CMA, it also revealed some novel candidates for CMA. Among the 8 validated DMPs, 3 DMPs, all with an absolute mean DNAm difference of more than 5% between cases and controls, were annotated to *NDFIP2* (cg11770323), *EVL* (cg18550847), and *TRAPPC9* (cg09377531), respectively. These 3 genes have not previously been implicated in allergy, but their role may be biologically plausible. *NDFIP2*, one of the *IL4*-regulated genes, can promote IFN- γ production via polarized human T_H1 lymphocytes.⁶ *NDFIP2* can also activate E3 ubiquitin ligases, which may play an important role in preventing allergic diseases. *EVL* is one of the *IL-13*-regulated genes.⁷ The analyses from GIANT⁸ showed that the *EVL* and *IL4* genes are highly coexpressed in basophils. *TRAPPC9* encodes a protein that likely plays a role in nuclear

factor kappa B signaling, a pathway that may be relevant to the immune system. Further studies are needed to explore the role of these annotated genes in CMA.

The 2 replication analyses in this study, although with limited sample sizes, have provided additional information. Replication in the Chicago sample, where the cases were all allergic to both peanut and cow's milk and the non-CMA controls were all allergic to peanut only, suggested that the replicated DNAm associations may be specific to CMA, whereas replication in the Boston sample, where DNAm was measured in cord blood from black samples, indicate that the replicated DNAm alterations between cases and controls may occur *in utero* and may be common across race groups.

Several limitations are noted. First, we used whole blood DNA to obtain DNAm profiling. Second, questions remain regarding how to interpret small DNAm differences between CMA and controls, mostly less than 5%. One possibility is that such small changes in whole blood may be reflective of larger changes in a specific cell type, as demonstrated in a previous study.⁹ Third, DNAm levels of the identified DMPs were not validated using other techniques such as pyrosequencing, and no functional assays were conducted in this study. However, previous studies have demonstrated that the 450K array correlates well with direct pyrosequencing.¹⁰

In conclusion, this was the first epigenome-wide association study of CMA in US children and it clearly demonstrated that methylation alterations in specific gene loci are associated with CMA. Findings from this study warrant additional validation and functional studies, and, if confirmed, may help to improve our understanding of the epigenetic mechanisms underlying the development and resolution of CMA and offer novel targets for prediction, prevention, and treatment of CMA.

We gratefully acknowledge the individuals and families who participated in the Chicago Food Allergy Study and in the Boston Birth Cohort.

Xiumei Hong, MD, PhD^a
Christine Ladd-Acosta, PhD^b
Ke Hao, PhD^c
Ben Sherwood, PhD^d
Hongkai Ji, PhD^e
Corinne A. Keet, MD, MS, PhD^e
Rajesh Kumar, MD^f
Deanna Caruso, MS^a
Xin Liu, MD, PhD^{g,h}
Guoying Wang, MD, PhD^a
Zhu Chen, PhD^a
Yuelong Ji, MSPH^a
Guanyun Mao, MD, PhD^{a,i}
Sheila Ohlsson Walker, CFA, PhD^{a,j,k}
Tami R. Bartell, BS^l
Zhicheng Ji, BS^l
Yifei Sun, PhD^l
Hui-Ju Tsai, MPH, PhD^{h,m}
Jacqueline A. Pongratic, MD^f
Daniel E. Weeks, PhDⁿ
Xiaobin Wang, MD, MPH, ScD^{a,o}

From ^athe Department of Population, Family and Reproductive Health, Center on the Early Life Origins of Disease, Johns Hopkins University Bloomberg School of Public Health, Baltimore, Md; ^bthe Department of Epidemiology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, Md; ^cthe Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; ^dthe Department of Biostatistics, Johns Hopkins University Bloomberg School of Public Health, Baltimore, Md; ^ethe Division of Pediatric Allergy and Immunology, Johns Hopkins University School of Medicine, Baltimore, Md; ^fthe Division of Allergy and Immunology, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, Ill; ^gthe Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China; ^hthe Department of Pediatrics, Feinberg School of Medicine, Northwestern University, Chicago, Ill; ⁱthe Department of Preventive Medicine, School of Environmental Science & Public Health, Wenzhou Medical University, Wenzhou, China; ^jthe Institute for Interdisciplinary Salivary Bioscience Research, Arizona State University, Tempe, Ariz; ^kThe Johns Hopkins School of Education, Baltimore, Md; ^lMary Ann & J. Milburn Smith Child Health Research Program, Stanley Manne Children's Research Institute, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, Ill; ^mthe Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, Zhunan, Taiwan; ⁿthe Departments of Human Genetics and Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pa; and ^othe Division of General Pediatrics & Adolescent Medicine, Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, Md. E-mail: xwang82@jhu.edu.

This research was supported in part by the Bunning Family and their family foundations, Sacks Family Foundation Fund, Food Allergy Research and Education (FARE), and grants from the National Institute of Allergy and Infectious Diseases (NIAID, grant no. R56AI080627, R21AI088609; and U01AI090727 as part of the consortium of Food Allergy Research) and the Department of Defense (grant no. W81XWH-10-1-0123). H.J. is partly supported by the National Institutes of Health (grant no. R01HG006841). C.A.K. is partly supported by the NIAID (grant no. K23AI103187). H.-J.T. is partly supported by grants from the National Science Council (grant no. 101-2314-B-400-009-MY2) and the Ministry of Science and Technology (grant no. 103-2314-B-400-004-MY3).

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

REFERENCES

1. Hong X, Wang X. Early life precursors, epigenetics, and the development of food allergy. *Semin Immunopathol* 2012;34:655-69.
2. Su RC, Becker AB, Kozyskyj AL, Hayglass KT. Epigenetic regulation of established human type 1 versus type 2 cytokine responses. *J Allergy Clin Immunol* 2008;121:57-63.e3.
3. Martino D, Dang T, Sexton-Oates A, Prescott S, Tang ML, Dharmage S, et al. Blood DNA methylation biomarkers predict clinical reactivity in food-sensitized infants. *J Allergy Clin Immunol* 2015;135:1319-28.e12.
4. Berni Canani R, Paparo L, Nocerino R, Cosenza L, Pezzella V, Di Costanzo M, et al. Differences in DNA methylation profile of Th1 and Th2 cytokine genes are associated with tolerance acquisition in children with IgE-mediated cow's milk allergy. *Clin Epigenetics* 2015;7:38.
5. Kumar R, Ouyang F, Story RE, Pongratic JA, Hong X, Wang G, et al. Gestational diabetes, atopic dermatitis, and allergen sensitization in early childhood. *J Allergy Clin Immunol* 2009;124:1031-8, e1-4.
6. Lund RJ, Loytomaki M, Naumanen T, Dixon C, Chen Z, Ahlfors H, et al. Genome-wide identification of novel genes involved in early Th1 and Th2 cell differentiation. *J Immunol* 2007;178:3648-60.
7. Scotton CJ, Martinez FO, Smelt MJ, Sironi M, Locati M, Mantovani A, et al. Transcriptional profiling reveals complex regulation of the monocyte IL-1 beta system by IL-13. *J Immunol* 2005;174:834-45.
8. Greene CS, Krishnan A, Wong AK, Ricciotti E, Zelaya RA, Himmelstein DS, et al. Understanding multicellular function and disease with human tissue-specific networks. *Nat Genet* 2015;47:569-76.
9. Liang L, Willis-Owen SA, Laprise C, Wong KC, Davies GA, Hudson TJ, et al. An epigenome-wide association study of total serum immunoglobulin E concentration. *Nature* 2015;520:670-4.
10. Roessler J, Ammerpohl O, Gutwein J, Hasemeier B, Anwar SL, Kreipe H, et al. Quantitative cross-validation and content analysis of the 450k DNA methylation array from Illumina, Inc. *BMC Res Notes* 2012;5:210.

<http://dx.doi.org/10.1016/j.jaci.2016.01.056>

METHODS

Study population

For our epigenome-wide screen, we obtained blood from 106 white children with CMA and 77 nonallergic white controls as defined below from the Chicago Food Allergy Study (the discovery sample), which has been previously described in detail.^{E1,E2} Briefly, eligible families had at least 1 biological child with food allergy (FA) and were willing to participate in the study. For each enrolled participant, the following procedures were completed: (1) questionnaire interview by trained research staff to obtain information on home environment, diet, lifestyle, and history of FA and other allergic diseases; (2) allergy skin prick testing (SPT) to 9 food allergens and 6 aeroallergens; and (3) collection of venous blood samples. For each child, we also collected a detailed history of the clinical allergic reaction associated with ingestion of specific foods. The Institutional Review Board of Ann & Robert H. Lurie Children's Hospital of Chicago and Johns Hopkins Bloomberg School of Public Health approved the study protocol.

For replication purposes, we obtained 2 independent samples. The first sample consisted of 5 white cases with allergy to both cow's milk and peanut and 20 positive controls with allergy to peanut but not to cow's milk (see detailed description below for phenotypic classification) enrolled from the same Chicago Food Allergy Study (the Chicago replication sample). These subjects were independent of the discovery sample. The second sample consisted of 140 black children from the prospective Boston Birth Cohort (BBC)^{E3} including 8 cases with CMA and 132 normal controls (the Boston replication sample). Unlike the samples obtained from the Chicago Food Allergy Study, the Boston sample had DNAm measured in cord blood (before clinical expression of disease). The institutional review boards at the Boston Medical Center and Johns Hopkins Bloomberg School of Public Health approved the BBC study protocol.

Phenotype definition

As reported previously,^{E2,E4} we defined CMA cases in this study using the following stringent criteria: (1) a convincing history of symptoms indicative of an allergic reaction within 2 hours of ingestion of cow's milk and (2) clear evidence of sensitization, defined as having a specific IgE (sIgE) level of 0.35 kU/L or more to cow's milk and/or a positive SPT result to cow's milk with a mean wheal diameter (MWD) of 3 mm or more greater than that with the saline control. *Normal controls* were defined as a child who had neither a clinical allergic reaction nor evidence of sensitization to any of the common foods (peanut, egg white, cow's milk, soy, wheat, walnut, fish, shellfish, and sesame seed). In the Chicago replication sample, *positive controls* were defined as a child who had neither a clinical allergic reaction nor evidence of sensitization to cow's milk, but was allergic to peanut.

SPT measurement

In the Chicago Food Allergy Study, SPT to 9 food allergens (cow's milk, egg white, soybean, wheat, peanut, English walnut, sesame seed, fish mix [cod, flounder, halibut, mackerel, tuna], and shellfish mix [clam, crab, oyster, scallops, shrimp]) and 6 aeroallergens (2 dust mites [*Dermatophagoides teronyssinus* and *Dermatophagoides farinae*], cat hair, dog epithelia, cockroach mix, and *Alternaria tenuis*), plus negative (50% glycerinated saline) and positive (histamine, 1.0 mg/mL) controls (Greer, Lenoir, NC), was performed using a Multi-Test II device (Lincoln Diagnostics, Decatur, Ill) as previously described in detail.^{E2,E4} Data were excluded if the MWD for the negative control was 3 mm or more, or if the MWD for the positive control was less than 3 mm, or if the difference of the positive minus the negative control was less than 3 mm. No SPT measurements were performed in the BBC.

sIgE measurement

In the Chicago Food Allergy Study, sIgE levels for 9 food allergens (egg white, sesame, peanut, soy, milk, shellfish, walnut, cod fish, and wheat) and 6 aeroallergens (*Dermatophagoides pteronyssinus* and *Dermatophagoides farinae*, cat dander, dog dander, German cockroach, and *Alternaria alternata*) were measured using the Phadia ImmunoCAP system (Phadia US, Inc, Portage, Mich) by the Clinical Immunology Laboratory at Lurie Children's, a Clinical Laboratory Improvement Amendment certified laboratory for the

ImmunoCAP assay. The lower detection limit for each sIgE was 0.1 kU_A/L and the reporting range was 0.1 to 100 kU_A/L. In the BBC, the sIgE level for 8 food allergens (egg white, peanut, soy, milk, shrimp, walnut, cod fish, and wheat) was measured at Quest Diagnostics using the Phadia ImmunoCAP system. The lower detection limit for food sIgE level was 0.35 kU_A/L.

DNAm measurement and quality control steps

In the discovery stage, case and control DNA samples (50 ng/uL) were evenly distributed across each plate to minimize batch effects, which were then shipped to the Northwestern University Genomics Core for DNAm profiling. Briefly, 0.5 μg of genomic DNA was bisulfite-converted using the EZ-96 DNA Methylation Gold Kit, and DNAm levels at 485,512 loci were measured using the Infinium HumanMethylation450 BeadChips (450K), according to the manufacturer's instructions.

A raw intensity file (.idat) for each sample was processed, and several quality control steps were performed using the "minfi"^{E5} Bioconductor package, as we reported previously.^{E2} First, we examined the 450K control probes to assess bisulfite conversion, extension, hybridization, staining, specificity, and others. Second, we computed the median for both Meth and Unmeth signals for each array and displayed them in a scatter plot. This approach clearly identified one outlier sample with a median log₂ intensity value of less than 10.5, which was removed from further analyses. Third, we removed 748 loci that had a detection *P* value (a measure of probe performance) of more than .01 in 10% or more of the samples. Fourth, we removed 21,524 CpG sites that had an annotated single nucleotide polymorphism (minor allele frequency of >0.01) at the measured and/or neighboring locus (±1bp depending on probe strand orientation), and removed 27,513 CpG sites previously reported to be cross-reactive.^{E6} After these quality control steps, DNAm of 435,642 sites (including 424,420 autosomal sites and 11,222 X-chromosome sites) from 182 samples were available for subsequent data analyses.

Using the minfi package,^{E5} a stratified quantile normalization procedure was applied to the raw data and normalized beta values (β), ranging from 0 to 1 for 0% to 100% methylated, were obtained. *M* values (logit transformed β) were also computed. Both the β and *M* values were ComBat-transformed^{E7} using the "sva" package^{E8} with the array number as the surrogate for the batches. ComBat-transformed *M* values were used for downstream statistical analyses because they represent a more normal distribution than do β values. ComBat-transformed β values were used for plotting purposes because they are more intuitive. The same protocols for DNAm measurement and data cleaning were applied to the 2 replication samples.

Empirical estimation of blood cell composition

To account for potential differences in the proportions of cells that comprised the blood,^{E9-E11} we empirically estimated the proportion of CD8+ T cells, CD4+ T cells, natural killer cells, B cells, monocytes, and granulocytes as previously described by Houseman et al^{E10} for all the discovery and replication samples, using the minfi package.^{E5}

Statistical analyses in the discovery stage

To identify DMPs associated with CMA in the discovery sample, we used the "limma" package^{E12} to fit a linear regression model for ComBat-transformed *M* value at each CpG site as a function of CMA status (1 = case, 0 = control), adjusting for potential confounders including sex, age group (coded as 0 = <2 years, 1 = 2-6 years, 2 = 6-10 years, 3 = ≥10 years), breast-feeding history (yes/no), parental history of FA (0 = none, 1 = one parent, 2 = both parents with FA), cell-type proportions for granulocytes, monocytes, B cells, natural killer cells, CD4+ T cells, and CD8+ T cells, as well as genetic ancestry (represented by the first 2 principal components based on genome-wide single nucleotide polymorphism data, as we reported previously^{E2}). Bonferroni correction was applied to account for multiple testing ($P < 1.15 \times 10^{-7}$). For CpG sites on the X-chromosome, sex-specific analyses were also performed.

As a complementary approach to single-site DMP analyses, we searched for DMRs associated with CMA using the bump hunting() function^{E13} in the minfi package. Briefly, each locus was a priori assigned to a region such that any 2 neighboring loci separated by 300bp or less were assigned to the

same region. Next, for each CpG site, we estimated the difference in average combat-transformed *M* value between CMA cases and controls, adjusting for the same covariates as described above. Significance was assigned with the application of a bootstrapping method (*n* = 1000 times). CMA candidate regions with a family-wise error of less than .05 were considered significant. All statistical analyses were performed using R-3.1.1 and Bioconductor 2.12.

We then performed a Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis to identify biological pathways with enrichment of genes that exhibited differential DNAm between 106 CMA cases and 76 controls, using WebGestalt^{E14} and with the approximately 20,000 annotated genes of the studied 435,642 CpG sites as the background. We used a Bonferroni-adjusted *P* value of less than .05 as our significance threshold.

Statistical analyses in the replication samples

We decided to validate a subset of identified CMA-associated DMPs in our replication analyses because of the small number of samples. These subsets were selected on the basis of 1 or more of the following criteria: (1) significant DMPs ($P < 1.15 \times 10^{-7}$) with an absolute DNAm difference of more than 5% and/or (2) significant DMPs ($P < 1.15 \times 10^{-7}$) that annotated to genes relevant to the T_H1 - T_H2 pathway, a critical pathway in allergy development. In both replication samples, a linear regression model, similar to the one applied in the discovery stage, was fitted to explore the association of each DMP with CMA. Because of a limited sample size, the potential confounders that we considered in the Chicago sample included child's age and cell-type proportions. In the Boston sample, the confounders that we adjusted included cell-type proportions, sex, parental history of FA, and gestational age group (due to a significant association between gestational age and cord blood DNAm level^{E15,E16}). We used an FDR-corrected *P*-value threshold (FDR < 0.05) to account for multiple testing in each replication.

For the CMA-associated DMPs that were validated in at least 1 replication sample, we then explored their biological relevance, using the Web tool EpiExplorer.^{E17}

REFERENCES

- E1. Tsai HJ, Kumar R, Pongracic J, Liu X, Story R, Yu Y, et al. Familial aggregation of food allergy and sensitization to food allergens: a family-based study. *Clin Exp Allergy* 2009;39:101-9.
- E2. Hong X, Hao K, Ladd-Acosta C, Hansen KD, Tsai HJ, Liu X, et al. Genome-wide association study identifies peanut allergy-specific loci and evidence of epigenetic mediation in US children. *Nat Commun* 2015;6:6304.
- E3. Hong X, Wang G, Liu X, Kumar R, Tsai HJ, Arguelles L, et al. Gene polymorphisms, breast-feeding, and development of food sensitization in early childhood. *J Allergy Clin Immunol* 2011;128:374-81.e2.
- E4. Hong X, Caruso D, Kumar R, Liu R, Liu X, Wang G, et al. IgE, but not IgG4, antibodies to Ara h 2 distinguish peanut allergy from asymptomatic peanut sensitization. *Allergy* 2012;67:1538-46.
- E5. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 2014;30:1363-9.
- E6. Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics* 2013;8:203-9.
- E7. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007;8:118-27.
- E8. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 2012;28:882-3.
- E9. Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol* 2014;15:R31.
- E10. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 2012;13:86.
- E11. Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlen SE, Greco D, et al. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS One* 2012;7:e41361.
- E12. Smyth GK. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 2004;3:Article3.
- E13. Jaffe AE, Murakami P, Lee H, Leek JT, Fallin MD, Feinberg AP, et al. Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *Int J Epidemiol* 2012;41:200-9.
- E14. Wang J, Duncan D, Shi Z, Zhang B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res* 2013;41:W77-83.
- E15. Cruickshank MN, Oshlack A, Theda C, Davis PG, Martino D, Sheehan P, et al. Analysis of epigenetic changes in survivors of preterm birth reveals the effect of gestational age and evidence for a long term legacy. *Genome Med* 2013;5:96.
- E16. Lee H, Jaffe AE, Feinberg JI, Tryggvadottir R, Brown S, Montano C, et al. DNA methylation shows genome-wide association of NFIX, RAPGEF2 and MSRB3 with gestational age at birth. *Int J Epidemiol* 2012;41:188-99.
- E17. Halachev K, Bast H, Albrecht F, Lengauer T, Bock C. EpiExplorer: live exploration and global analysis of large epigenomic datasets. *Genome Biol* 2012;13:R96.

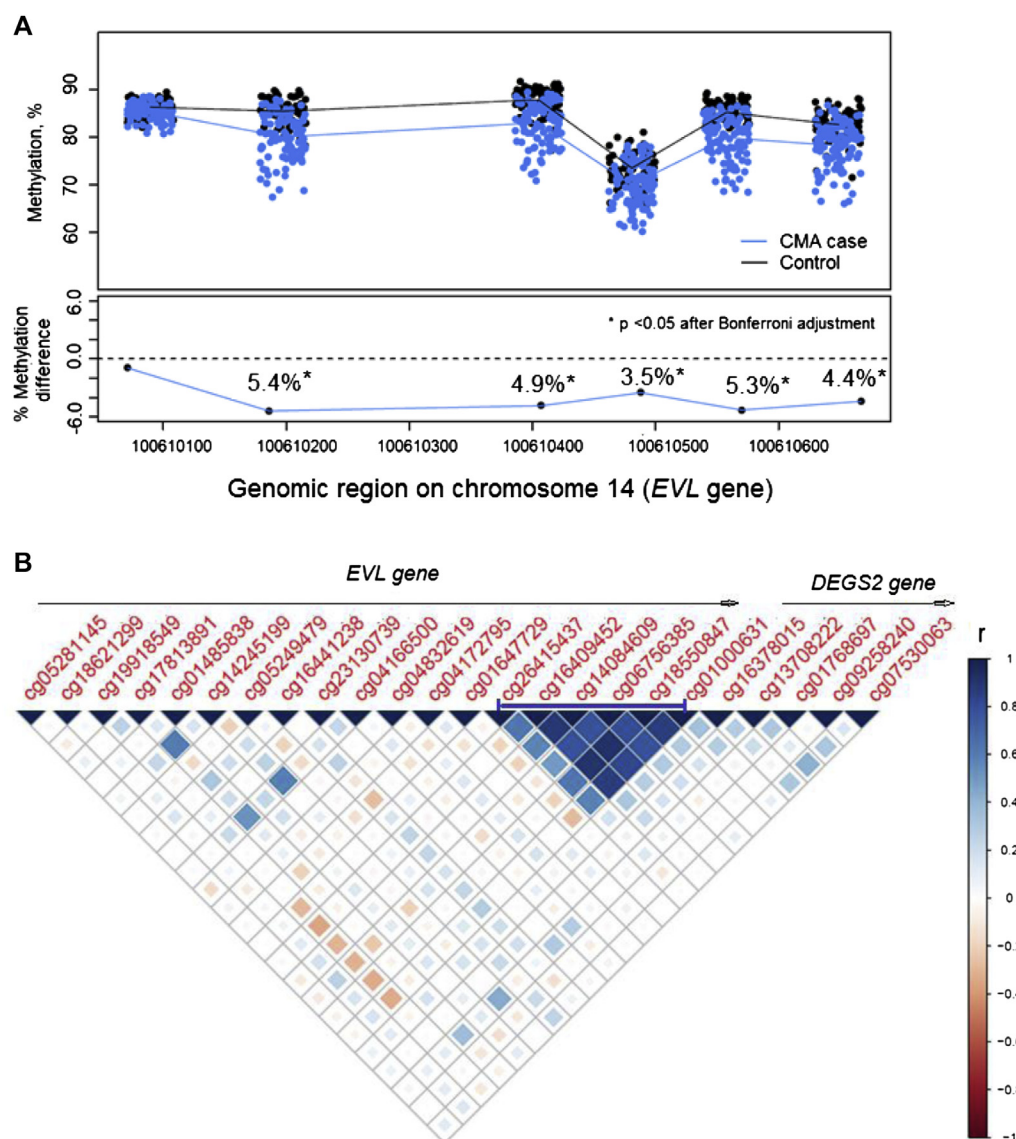


FIG E1. A genomic cluster in the *EVL* gene that hypomethylated in CMA children compared with normal controls. **A**, The upper panel shows DNAm levels on the y-axis and genomic location on the x-axis. Each point represents the methylation level at a specific CpG site for each CMA case (in blue) and control (in black). The lower panel displays the methylation difference between cases and controls for the 6 involved CpG sites (cg26415437, cg16409452, cg14084609, cg06756385, cg18550847, and cg01000631); * $P < 1.15 \times 10^{-7}$ for the methylation difference between cases and controls. **B**, DNAm correlation for the genomic cluster shown in Fig E1, A. The measure of correlation coefficient (r) among each pair of CpG is shown graphically, with blue representing high positive correlation.

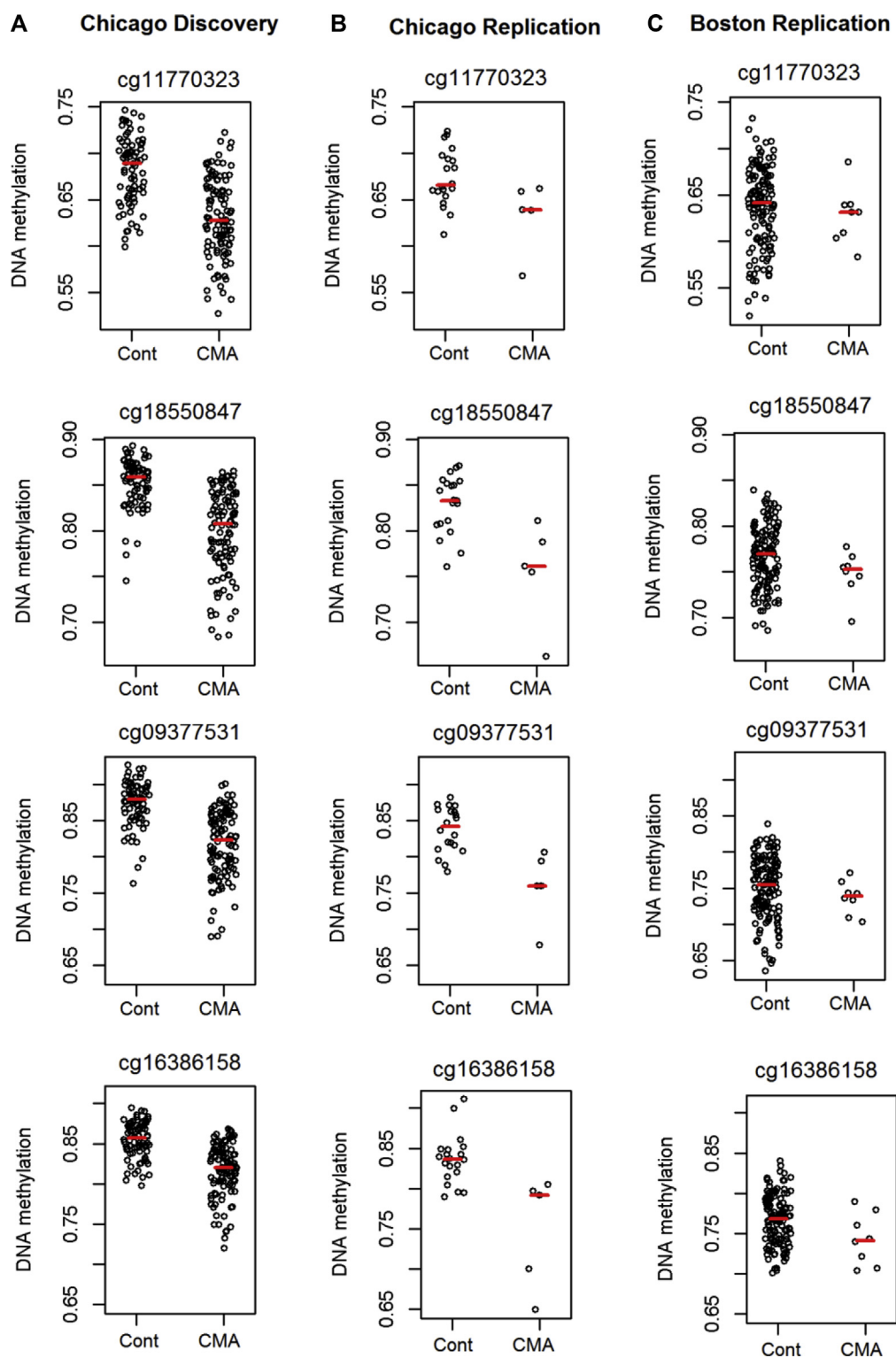


FIG E2. Scatter plots of DNA methylation levels at 8 CMA-associated DMPs that were validated in at least 1 replication sample. **A**, Plot for each DMP in the discovery sample ($n = 182$). **B**, Plot for each DMP in the Chicago replication sample. **C**, Plot for each DMP in the Boston replication sample. The red segment denotes the median DNA methylation levels within each subset of subjects.

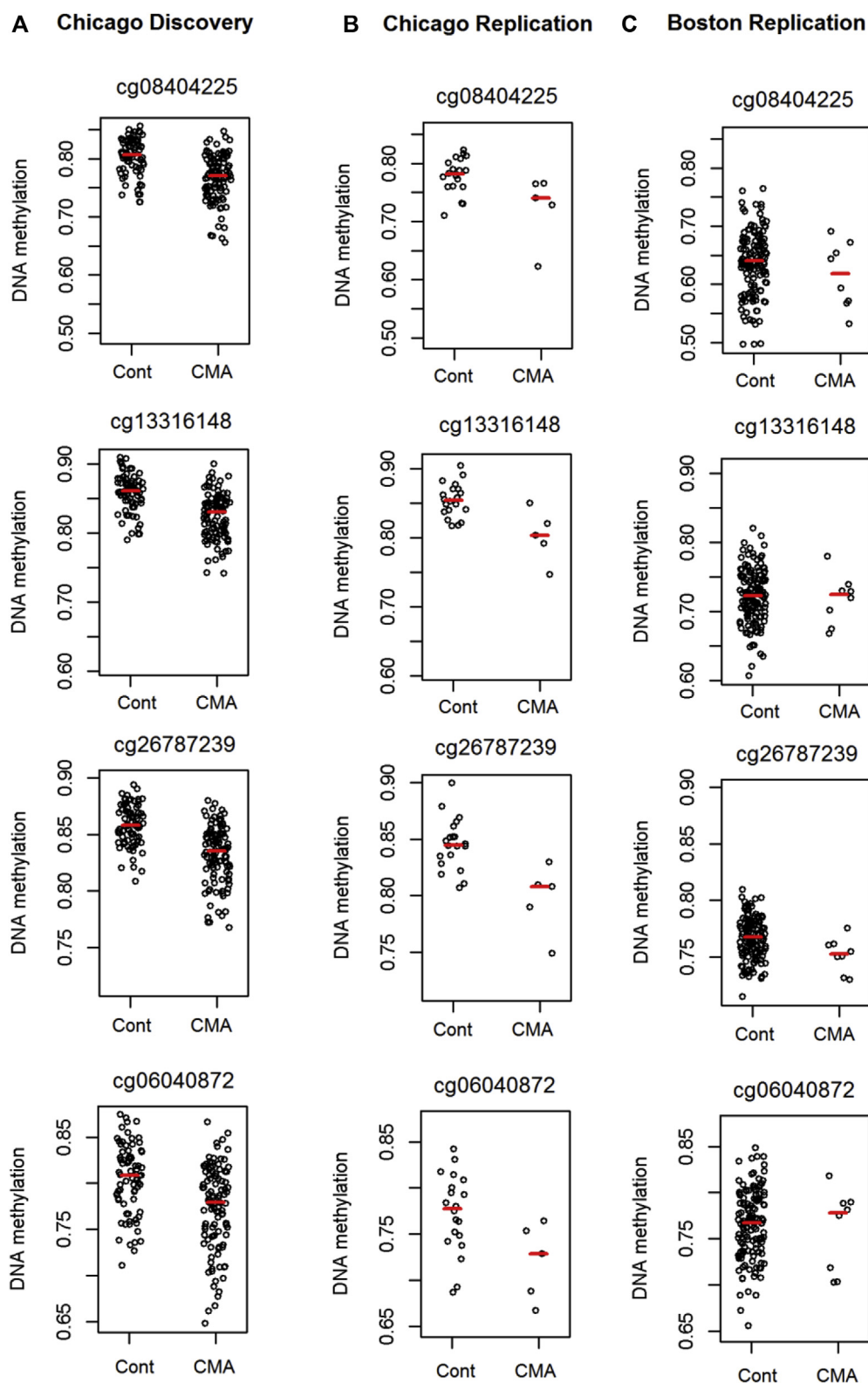


FIG E2. (Continued).

TABLE E1. Population characteristics of the 182 white children included in the discovery stage

Variable	Normal controls (N = 76)	CMA cases (N = 106)	P value*
Child's age (y), mean $\pm$ SD	5.5 $\pm$ 4.0	4.2 $\pm$ 2.7	.010
Boy, n (%)	31 (40.8)	67 (63.2)	.003
Parent history of FA, n (%)			.023
Paternal-only	11 (14.5)	19 (17.9)	
Maternal-only	7 (9.2)	21 (19.8)	
Both	0	5 (4.7)	
Neither	58 (76.3)	61 (57.6)	
Current maternal smoking, n (%)	20 (26.3)	18 (17.0)	.127
Current paternal smoking, n (%)	14 (18.4)	19 (17.9)	.932
Preterm birth, n (%)	8 (10.5)	9 (8.5)	.632
C-section, n (%)	22 (28.9)	35 (33.0)	.559
Pets in the home, n (%)	43 (56.6)	49 (46.2)	.168
Any breast-feeding, n (%)	67 (88.2)	101 (95.3)	.075
Maternal consumption of cow's milk product during breast-feeding, n (%)†			.585
≤ 2 d	2 (4.4)	5 (5.6)	
3-5 d	9 (19.6)	20 (22.2)	
≥ 6 d	35 (76.1)	62 (68.9)	
Stop consumption during the first 3 mo	0	3 (3.3)	
Antibiotic use during the first year of life			.293
Yes	36 (47.4)	38 (35.9)	
No	38 (50.0)	65 (61.3)	
Unsure	2 (2.6)	3 (2.8)	
sIgE to cow's milk (kU/L), median (25 th -75 th)‡	0	23.1 (6.2-59.6)	<.001
SPT to cow's milk (mm), median (25 th -75 th)§	0	10.5 (8.5-14.0)	<.001
Other types of FA	0	49 (46.3)	<.001
Egg allergy	0	36 (34.0)	<.001
Peanut allergy	0	7 (6.6)	<.001

*Comparison between CMA cases and controls was performed via χ^2 test and *t* test, respectively, for categorical and continuous variables.

†This analysis was limited to 146 children (46 normal controls and 90 CMA cases) with any breast-feeding and younger than 7 years.

‡Four children (1 control and 3 cases) had missing data on milk-specific IgE.

§Thirty-four children (6 controls and 28 cases) had missing data on milk SPT.

TABLE E2. Population characteristics of the white replication sample from the Chicago Food Allergy Study and of the black replication sample from the prospective BBC birth cohort

Characteristic	White replication sample			Black replication sample		
	Controls	Cases	P value	Controls	Cases	P value
n	20	5		132	8	
Child's age (y), mean $\pm$ SD*	9.1 $\pm$ 2.5	7.2 $\pm$ 4.1	.331	3.0 $\pm$ 3.3	3.8 $\pm$ 2.1	.312
Boys, n (%)	13 (65.0)	4 (80.0)	.520	81 (61.4)	5 (62.5)	.949
Preterm birth, n (%)	5 (25.0)	0	.211	57 (43.2)	2 (25.0)	.312
Current maternal smoking, n (%)	5 (25.0)	1 (20.0)	.815	24 (18.2)	2 (25.0)	.630
Current paternal smoking, n (%)†	5 (25.0)	1 (20.0)	.461	41 (31.0)†	3 (37.5)	.703
Parental history of FA, n (%)‡	7 (35.0)	4 (80.0)	.070	23 (18.9)‡	4 (57.1)‡	.023
Any breast-feeding, n (%)	19 (95.0)	5 (100.0)	.610	100 (75.8)	6 (75.0)	.961
Eczema, n (%)	13 (65.0)	4 (80.0)	.520	0	4 (50.0)	<.001
Hay fever, n (%)	7 (35.0)	5 (100.0)	.009	24 (18.2)	6 (75.0)	<.001
Asthma, n (%)	8 (40.0)	5 (100.0)	.016	30 (22.7)	5 (62.5)	<.002
Other types of FA, n (%)	20 (100.0)§	5 (100.0)	1.00	0	3 (37.5)	<.001

*Child's age when CMA was defined.

†A total of 15 controls had missing data on current paternal smoking.

‡A total of 10 controls and 1 case had missing data on parental history of FA.

§All the non-CMA controls were allergic to peanut.

TABLE E3. KEGG pathways with enrichment of genes that exhibit differential DNAm in children with CMA

Pathway	No. of genes			Ratio§	Raw <i>P</i> value	Adjusted <i>P</i> value¶	Observed genes
	Total*	O†	E‡				
Butirosin and neomycin biosynthesis	5	3	0.10	31.32	6.8×10^{-5}	.004	<i>HK1, HK2, HK3</i>
Starch and sucrose metabolism	43	6	0.79	7.62	.0001	.003	<i>HK1, HK2, HK3, GPI, AGL, GUSB</i>
Fructose and mannose metabolism	35	5	0.67	7.80	.0004	.011	<i>HK1, HK2, HK3, PFKFB3, FB1</i>

AGL, Amylo-alpha-1, 6-glucosidase, 4-alpha-glucanotransferase; *FB1*, TCF3 (E2A) fusion partner; *HK1*, hexokinase 1; *HK2*, hexokinase 2; *HK3*, hexokinase 3; *GPI*, glucose-6-phosphate isomerase; *GUSB*, glucuronidase, beta; *KEGG*, Kyoto Encyclopedia of Genes and Genomes; *PFKFB3*, 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3.

*The total number of the reference genes in each pathway.

†The total number of genes in each pathway that exhibited differential DNAm in children with CMA.

‡The expected number of genes in each pathway that exhibited differential DNAm in children with CMA.

§Ratio = the number of observed genes (O)/the number of expected genes (E).

||Raw *P* value from hypergeometric test.

¶Bonferroni correction was applied to adjust for multiple testing.

TABLE E4. Biological relevance* of the 8 CMA-associated DMPs that were validated in at least 1 replication sample

DMP	CHR	Gene	Location	CpG island	DHS	Repeat element	LAD†	TFBS‡	Polycomb repressed gene	Enhancer
cg16386158	2	<i>IL1RL1</i>	TSS1500			SINE		CTCF		✓
cg13316148	2	<i>STAT4</i>	Body				✓	CTCF	✓	✓
cg08404225	3	<i>IL5RA</i>	5'UTR				✓			✓
cg26787239	5	<i>IL4</i>	TSS1500				✓	CTCF		✓
cg09377531	8	<i>TRAPPC9</i>	Body			LINE				
cg11770323§	13	<i>NDFIP2</i>	Body		✓			CTCF		✓
cg18550847§	14	<i>EVL</i>	3'UTR	✓						✓
cg06040872	17	<i>CCL18</i>	Body						✓	

CHR, Chromosome; DHS, DNaseI hypersensitive sites; LAD, lamina-associated domains; LINE, long interspersed nuclear element; SINE, short interspersed nuclear element; TFBS, transcription factor binding sites; TSS, transcription start site; UTR, untranslated region.

*Biological relevance was analyzed using the Web tool EpiExplorer.

†An event of hypomethylation of these regions may cause or reflect a fault of the mechanisms that are central to the development and conservation of normal states of differentiation and tissue-specific patterns of gene expression.

‡CTCF, CCCTC binding factor, a transcriptional regulator that may protect downstream DNA from upstream methylation activities.

§Overlapped with a conserved region.

||In mammary epithelial cells.