

Washington University School of Medicine

Digital Commons@Becker

2020-Current year OA Pubs

Open Access Publications

6-1-2023

A functional genomics pipeline to identify high-value asthma and allergy CpGs in the human methylome

Andréanne Morin

Katherine Rivera-Spoljaric

et al.

Follow this and additional works at: https://digitalcommons.wustl.edu/oa_4



Part of the [Medicine and Health Sciences Commons](#)

Please let us know how this document benefits you.

A functional genomics pipeline to identify high-value asthma and allergy CpGs in the human methylome



Andréanne Morin, PhD,^{a*} Emma E. Thompson, PhD,^{a*} Britney A. Helling, PhD,^{a*} Lyndsey E. Shorey-Kendrick, PhD,^b Pieter Faber, PhD,^c Tebeb Gebretsadik, MPH,^d Leonard B. Bacharier, MD,^e Meyer Kattan, MD,^f George T. O'Connor, MD,^g Katherine Rivera-Spoljaric, MD,^h Robert A. Wood, MD,ⁱ Kathleen C. Barnes, PhD,^j Rasika A. Mathias, PhD,^k Matthew C. Altman, MD,^{l,m} Kasper Hansen, PhD,ⁿ Cindy T. McEvoy, MD,^o Eliot R. Spindel, MD,^b Tina Hartert, MD,^d Daniel J. Jackson, MD,^p James E. Gern, MD,^p Chris G. McKennan, PhD,^{q,‡} and Carole Ober, PhD,^{a,‡} on behalf of program collaborators for Environmental Influences on Child Health Outcomes and Children's Respiratory and Environmental Workgroup
Chicago, Ill; Beaverton and Portland, Ore; Nashville, Tenn; New York, NY; Boston, Mass; St Louis, Mo; Baltimore, Md; Aurora, Colo; Seattle, Wash; Madison, Wis; and Pittsburgh, Pa

Background: DNA methylation of cytosines at cytosine-phosphate-guanine (CpG) dinucleotides (CpGs) is a widespread epigenetic mark, but genome-wide variation has been relatively unexplored due to the limited representation of variable CpGs on commercial high-throughput arrays.

Objectives: To explore this hidden portion of the epigenome, this study combined whole-genome bisulfite sequencing with *in silico* evidence of gene regulatory regions to design a custom array of high-value CpGs. This study focused on airway epithelial cells from children with and without allergic asthma because these cells mediate the effects of inhaled microbes, pollution, and allergens on asthma and allergic disease risk.

Methods: This study identified differentially methylated regions from whole-genome bisulfite sequencing in nasal epithelial cell DNA from a total of 39 children with and without allergic asthma of both European and African ancestries. This study selected CpGs from differentially methylated regions, previous allergy or asthma epigenome-wide association studies (EWAS), or genome-wide association study loci, and overlapped them with functional annotations for inclusion on a custom Asthma&Allergy array. This study used both the custom and EPIC arrays to perform EWAS of allergic sensitization (AS) in nasal epithelial cell DNA from children in the URECA (Urban Environment and Childhood Asthma) birth cohort and using the custom array in the

From ^athe Department of Human Genetics, and ^cthe Genomics Core, University of Chicago; ^bthe Division of Neuroscience, Oregon National Primate Research Center, Oregon Health and Science University, Beaverton, and ^othe Department of Pediatrics, Oregon Health and Science University, Portland; ^dthe Department of Medicine, Vanderbilt University School of Medicine, and ^ethe Department of Pediatrics, Monroe Carell Jr Children's Hospital at Vanderbilt University Medical Center, Nashville; ^fthe Department of Pediatrics, Columbia University Medical Center, New York; ^gthe Pulmonary Center, Boston University School of Medicine; ^hthe Department of Pediatrics, Washington University, St Louis; ⁱthe Department of Pediatrics, and ^kthe Department of Medicine, Johns Hopkins University, and ⁿthe Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore; ^lthe Department of Medicine, University of Colorado Denver, Aurora; ^mthe Systems Immunology Division, Benaroya Research Institute Systems, and ^mthe Department of Medicine, University of Washington, Seattle; ^pthe Department of Pediatrics, University of Wisconsin School of Medicine and Public Health, Madison; and ^qthe Department of Statistics, University of Pittsburgh.

*These authors contributed equally to this work.

‡These authors jointly supervised this work.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Research reported in this publication was supported by the Environmental Influences on Child Health Outcomes (ECHO) program, Office of the Director, National Institutes of Health, under awards U2COD023375 (Coordinating Center), U24OD023382 (Data Analysis Center), U24OD023319 (Patient reported outcome [PRO] Core), UH3 OD023282 (COAST [Childhood Origins of Asthma], INSPIRE [Infant Susceptibility to Pulmonary Infections and Asthma Following RSV Exposure], URECA [Urban Environment and Childhood Asthma]), and UH3 OD023288 (VCSIP [Vitamin C to Decrease the Effects of Smoking in Pregnancy on Infant Lung Function]) and grants UM1 AI114271, UM1 AI160040, U19 AI62310, UL1 TR002243, P51 OD011092, and R01 HL104608.

Disclosure of potential conflict of interest: L.B. Bacharier reports personal fees from GlaxoSmithKline, Genentech/Novartis, Merck, DBV Technologies, Teva, Boehringer Ingelheim, AstraZeneca, WebMD/Medscape, Sanofi/Regeneron, Vectura, Circassia, Elsevier, Kinaset, and Vertex outside the submitted work. G.T. O'Connor has received consulting fees from Dicerna and Menarini unrelated to this work. R.A. Wood has received grants from Aimmune, Astellas, DBV Technologies, Genentech, Novartis, Regeneron, and Sanofi. K.C. Barnes is employed by Tempus Labs. M.C. Altman has received consulting fees from Regeneron outside the submitted work. D.J. Jackson reports grants and personal fees from GlaxoSmithKline and Regeneron and personal fees from Pfizer, Sanofi, AstraZeneca, and Vifor Pharma, outside the submitted work. J.E. Gern is a paid consultant for AstraZeneca and Meissa Vaccines Inc. and has stock options in Meissa Vaccines Inc. The rest of the authors declare that they have no relevant conflicts of interest.

Received for publication August 16, 2022; revised November 24, 2022; accepted for publication December 20, 2022.

Available online February 6, 2023.

Corresponding author: Carole Ober, PhD, Department of Human Genetics, University of Chicago, 920 E. 58th Street, Chicago, IL 60637. E-mail: c-ober@genetics.uchicago.edu. Or: Christopher Gordon McKennan, PhD, Department of Statistics, University of Pittsburgh, 1826 Wesley W. Posvar Hall, 230 S Bouquet Street, Pittsburgh, PA 15260. E-mail: CHM195@pitt.edu.

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

0091-6749

© 2023 The Authors. Published by Elsevier Inc. on behalf of the American Academy of Allergy, Asthma & Immunology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

<https://doi.org/10.1016/j.jaci.2022.12.828>

INSPIRE [Infant Susceptibility to Pulmonary Infections and Asthma Following RSV Exposure] birth cohort. Each CpG on the arrays was assigned to its nearest gene and its promotor capture Hi-C interacting gene and performed expression quantitative trait methylation (eQTM) studies for both sets of genes.

Results: Custom array CpGs were enriched for intermediate methylation levels compared to EPIC CpGs. Intermediate methylation CpGs were further enriched among those associated with AS and for eQTMs on both arrays.

Conclusions: This study revealed signature features of high-value CpGs and evidence for epigenetic regulation of genes at AS EWAS loci that are robust to race/ethnicity, ascertainment, age, and geography. (J Allergy Clin Immunol 2023;151:1609-21.)

Key words: EWAS, DNA methylation array, gene expression, airway epithelial cells

Epigenetics refers to modifications of DNA molecules that do not alter the DNA sequence but play important roles in regulating gene expression. Environmental exposures can directly modify epigenetic marks in the human genome,¹ and epigenetic responses can mediate the effects of exposures on gene expression² and disease risk.³ Thus, the epigenome may contribute directly to disease risk or may be a site of gene-environment interactions, providing both complementary data and mechanistic insights, respectively, to genome-wide association studies (GWAS). The most common epigenetic mark in the human genome is methylated cytosines at cytosine-phosphate-guanine (CpG) dinucleotides (CpGs) and the availability of high-throughput array-based platforms to measure DNA methylation has led to an explosion of epigenome-wide association studies (EWAS).⁴ The most used commercial array, the Infinium MethylationEPIC BeadChip (hereafter EPIC; Illumina, Inc, San Diego, Calif), interrogates up to 850,000 CpGs, but this comprises <5% of CpGs in the human genome. Because the selection of CpGs on the EPIC array was agnostic with respect to noncancer diseases or tissue types, findings to date may represent just the tip of the iceberg with an abundance of yet unknown functional CpG sites contributing to disease risk.

The goal of our study was to develop and validate a custom Asthma&Allergy DNA methylation array enriched for CpGs in airway epithelial cells that are likely to be functional and associated with asthma and allergic diseases in populations of diverse ancestries. We focused on airway epithelial cells because of their critical role in responding to inhaled microbes, pollution, and allergens and mediating their downstream effects on asthma and allergic disease risk.^{5,6} An overview of our study design is shown in Fig 1.

METHODS

Cohorts included in WGBS studies or EWAS

Three longitudinal birth cohorts were included in these studies: the URECA (Urban Environment and Childhood Asthma)⁷ cohort (n = 20) and the COAST (Childhood Origins of Asthma)⁸ cohort (n = 19) were included in the whole genome bisulfite sequencing (WGBS) studies; URECA (n = 280) and the INSPIRE (Infant Susceptibility to Pulmonary Infections and Asthma Following RSV Exposure)⁹ study (n = 474) were included in EWAS of allergic sensitization (AS). Twenty individuals (COAST, n = 10;

Abbreviations used

AS:	Allergic sensitization
DMC:	Differentially methylated CpG
CpG:	Cytosine-phosphate-guanine
CpGs:	CpG dinucleotides
DMR:	Differentially methylated region
eQTM:	Expression quantitative trait methylation
EWAS:	Epigenome-wide association study
FDR:	False discovery rate
FET:	Fisher exact test
GWAS:	Genome-wide association study
HDAC:	Histone deacetylase
IM:	Intermediate methylation
LF:	Latent factor
NEC:	Nasal epithelial cell
pcHi-C:	Promoter capture Hi-C
QC:	Quality control
SNP:	Single nucleotide polymorphism
SPT:	Skin prick test
TFBS:	Transcription factor binding site
WGBS:	Whole-genome bisulfite sequencing

URECA, n = 10) were selected to represent the extremes of the asthma/allergy spectrum. We first selected individuals with parent-reported race as Black (in URECA) or White (in COAST) and then confirmed African admixed or European ancestries by principal components analysis. The participants who were allergic and asthmatic were required to have both atopy and wheezing in the first 3 years of life and asthma diagnosed at age 6 (COAST) or age 7 (URECA) years and at the time of sampling (age 18-20 in COAST and age 11 in URECA). The participants who were nonallergic and nonasthmatic were required to have none of the 3 phenotypes (atopy, wheeze, asthma) at any age until the age at sampling. We selected from these individuals those with sufficient DNA for methylation studies on the 2 platforms (EPIC and Booster) and then selected 20 from each cohort so that the participants who were asthmatic and those who were nonasthmatic were balanced for sex. All studies were approved by the Institutional Review Boards of each of the institutions recruiting subjects. For each cohort, 1 parent or guardian provided informed consent for their child's participation. Informed consent was obtained from participants in the COAST cohort after their 18th birthday. The characteristics of the 280 URECA and 474 INSPIRE participants in the EWAS are described in Table 1; detailed descriptions of the 3 cohorts are included in the Online Repository (available at www.jacionline.org).

Cohorts included in cross-tissue comparison studies

In addition to studying nasal epithelial cell (NEC) DNA with the Asthma&Allergy custom array in URECA, we included results from studies of the custom array in DNA from nasal lavage cells (n = 96), buccal cells (n = 96), placenta (n = 96), and cord blood (n = 96). The nasal lavage cell DNA was isolated from subjects in the URECA study at mean age 13.4 years (range 12.8-14.9 years). Methylation values for the same 96 children in the URECA study are shown for the epithelial and nasal lavage cells (custom) and the epithelial cells (EPIC). The buccal, placenta, and cord blood cell DNA was collected from participants in the VCSIP (Vitamin C to Decrease the Effects of Smoking in Pregnancy on Infant Lung Function) cohort.¹⁰ The methylation levels shown for these 3 cell/tissue types (EPIC and custom) are from a largely overlapping samples of the participants in this study.

WGBS and DMR studies

DNA (250 ng) from participants in the URECA and COAST cohorts was transferred to the University of Chicago Genomics Core Facility. Bisulfite

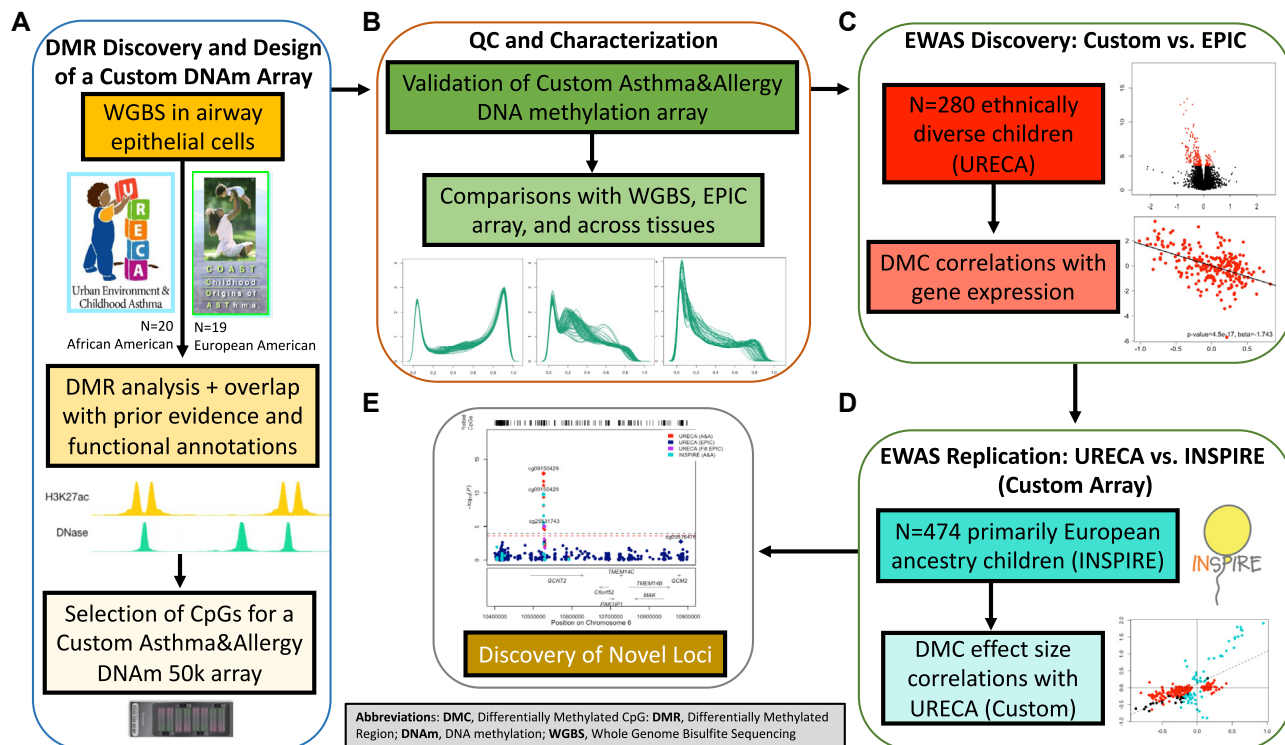


FIG 1. Overview of study design. **A**, WGBS and differential methylation were performed in airway epithelial cell DNA from birth cohorts composed of African American children or European American young adults (half of each cohort with allergic asthma and half without asthma or allergies). To select CpGs for the custom array, we identified CpGs based on prior evidence of association with asthma or allergic disease from 3 sources (see main text and Fig E1). The CpGs were then prioritized based on their overlap with functional annotations to ultimately design a custom array with 42,192 probes for 37,863 CpGs. **B**, Airway epithelial cell DNA was hybridized to both the custom array and the EPIC array, and β value distributions on the arrays were compared to the WGBS data and across tissues using the custom and EPIC arrays. **C-D**, We validated the custom array by performing an EWAS of AS, examined correlations with the gene expression in the same cells, and replicated findings in a second cohort and with allergic asthma. **E**, The combined results of the custom and EPIC arrays in the discovery cohort and of the custom array in the replication cohort for known and novel loci were examined in regional association plots.

TABLE I. Demographic characteristics of URECA and INSPIRE samples

Characteristic	URECA	INSPIRE
Sample size (n)	280	474
Age, mean (range), y	11.1 (10.9-12.4)	6.3 (5.0-7.9)
Female sex	131 (46.8)	223 (47.0)
Ethnicity		
Non-Hispanic White	1 (<1)	307 (65.4)
Non-Hispanic Black	201 (71.8)	86 (18.1)
Hispanic	56 (20.0)	45 (9.5)
Other	22 (7.9)	36 (7.6)
Clinical definitions		
Asthma*	89 (31.8)	45 (9.5)
Allergic asthma*	62 (22.1)	22 (4.6)
Allergic sensitization (≥ 1 +SPT)	177 (63.2)	179 (37.8)
Nonallergic/nonasthmatic	76 (27.1)	271 (57.2)

Also see Fig E5 for details on allergic sensitizations. Values are n (%) unless otherwise indicated.

*Asthma status was missing for 1 INSPIRE subject.

conversion and sequencing were conducted using the EZ DNA Methylation-Gold Kit (Zymo Research, Irvine, Calif) library prep and the Swift Biosciences Accel-NGS Methyl-seq DNA library kit (Ann Arbor, Mich),

respectively. Samples were sequenced to a minimum of 330 million reads on the Illumina NovaSeq 6000 (S4 flow cell). Adapters were removed from sequence reads using trimgalore¹¹ and then mapped using Bismark (version 0.18.2)¹² and the hg19 reference assembly. Bismark was also used to remove duplicate reads and to call methylation values. Prior to differential methylated region (DMR) analyses, CpGs were removed if they overlapped with a common single nucleotide polymorphism (SNP) (minor allele frequency >0.05) in 1000 Genomes European (CEU) or African (YRI) populations¹³ or in black-listed regions.¹⁴ Variant calling from WGBS was performed using the biscuit algorithm, which also identified sample swaps and contamination. The variant calls from WGBS were compared to either array-based genotypes (COAST) or whole genome-sequence-based genotypes (URECA). All variant calls matched with an accuracy of >95%. One sample in COAST was a duplicate and was excluded from further analysis, leaving 19 samples for analysis (9 from participants who were allergic and asthmatic and 10 from controls who were nonasthmatic and nonallergic). The DMR analyses are described in the Online Repository.

Selection of CpGs and design of the custom array

Our pipeline for selecting high-value CpGs for a custom array was implemented in 3 steps (see Fig E1 in this article's Online Repository at www.jacionline.org). Additional details are described in the Online Repository. Our goal was to develop a custom array with 50,000 probes for CpGs that would complement and serve as a "booster" for the EPIC array in studies

of asthma and allergic diseases. Briefly, we first identified a set of CpGs based on prior evidence of association with asthma or allergic diseases (atopic dermatitis/eczema, allergic rhinitis/hay fever, and food allergy) from 3 categories of studies. The first category included the 199,473 CpGs within DMRs from our WGBS study described above (see Table E1 in this article's Online Repository at www.jacionline.org), the second included 19,057 CpGs associated with asthma or allergic disease in 17 previous EWAS (see Table E2 in this article's Online Repository at www.jacionline.org for the platforms used and phenotypes included), and the third included 570,350 CpGs at 140 GWAS loci (see Table E3 in this article's Online Repository at www.jacionline.org for GWAS phenotypes included). We then removed CpGs on the EPIC array, in ENCODE blacklisted regions,¹² and those that overlapped with a common SNP in 1000 Genomes European (CEU) or African (YRI) populations. In the second step, we further prioritize the remaining 696,225 CpGs, by considering their overlap with 6 functional annotations: (1) ENCODE¹⁵ transcription factor binding sites (TFBSs) from all cell types; (2-4) Roadmap Epigenetics Consortium¹⁶ (2) transcriptional start sites, (3) poised enhancers, and (4) active enhancers from smooth muscle (E078, E076, E103, E111), epithelial (E055, E056, E059, E061, E058), and blood cells (E062, E034, E045, E044, E043, E039, E041, E042, E040, E037, E048, E038, E047, E029, E050, E032, E046); (5) assay for transposase-accessible chromatin using sequencing in human cultured bronchial epithelial cells exposed to rhinovirus or a vehicle control from individuals who are asthmatic and nonasthmatic,¹⁷ and (6) promoter capture Hi-C (pcHi-C) from *ex vivo* human bronchial epithelial cells.¹⁷ We then required that CpGs within DMRs overlap with at least 3 annotations or other category of prior evidence, CpGs from prior EWAS overlap with at least 1 annotation or other category of prior evidence, and CpGs at GWAS loci overlap with at least 4 annotation categories or other category of prior evidence. After removing duplicates, 92,024 high-value CpGs remained and 53,700 of these CpGs were amenable to the Illumina design algorithm. We selected 44,047 CpGs targeted by 49,999 probes for manufacturing, of which 14% subsequently failed, resulting in a custom array with 38,541 CpGs targeted by 43,605 probes. After removing probes with unknown origins and those without genomic coordinates, 37,863 CpGs passed array quality control (QC). The locations of the 37,863 CpGs, their selection among the 3 primary criteria, and their overlap with the annotation categories are shown in Dataset E1 in this article's Online Repository at www.jacionline.org. Finally, for our analyses we further removed probes that included a common SNP (minor allele frequency $\geq 5\%$) within 3 bp of the CpG interrogation site, probes that failed the minfi QC, and probes on the X chromosome, leaving 37,256 CpGs following both array and processing QC.

Processing custom and EPIC array methylation data in URECA and INSPIRE

DNA methylation was assessed using the Allergy&Asthma Custom BeadChip version 1.0, manufactured by Illumina, or the Illumina Infinium MethylationEPIC BeadChip following bisulfite conversion at the University of Chicago Genomics Facility. Methylation data from both arrays were processed using minfi version 1.29.1.¹⁸ The annotation and manifest packages for the Allergy&Asthma array version 1.0 can be found on GitHub (<https://github.com/hansenlab/IlluminaHumanMethylationAllergymanifest>). The details of our protocols for processing the DNA methylation data on each array are described in the Online Repository.

EWAS of AS in the URECA and INSPIRE cohorts

We defined AS as the percentage of positive skin prick tests ([SPTs]; 14 airborne and oral allergens in URECA and 13 airborne allergens in INSPIRE). For each AS EWAS we used the following models:

URECA: $\text{DNAm} \sim \text{proportion positive SPTs} + \text{sex} + \% \text{epithelial cells} + \text{AncPCs}_{1-3} + \text{LFs}_{1-n}$

INSPIRE: $\text{DNAm} \sim \text{proportion positive SPTs} + \text{sex} + \text{race/ethnicity} + \text{DNA concentration} + \text{LFs}_{1-n}$,

in which AncPCs = ancestry principal components, DNAm = DNA methylation, and LFs = latent factors.

For the URECA EWAS, we included 14 and 5 latent factors for the custom¹⁹ and EPIC²⁰ arrays, respectively, after removing technical variables (see Table E4 in this article's Online Repository at www.jacionline.org). Cell counts (300 cells/slide) were available for each sample.²¹ Because we did not have information on cell proportion or genetic ancestry for INSPIRE, we included sex, parent-reported race/ethnicity, DNA concentration, and 21 LFs as covariates to adjust for cell composition, as well as other unwanted variation in the EWAS. Analyses were performed in R (version 4.1.0; R Foundation, Vienna, Austria) using limma version 3.50.0.^{22,23} To control the false discovery rate (FDR), we used a *q*-value threshold of 0.05. To assess the overlap of AS differentially methylated CpGs (DMCs) with allergic asthma DMCs, we tested the AS DMCs for association with allergic asthma in URECA and INSPIRE. For these studies, we defined allergic asthma cases as children with a diagnosis of asthma at age 10 (URECA) or age 6 (INSPIRE) years and sensitization to at least 1 allergen ($n = 62$ and 22, respectively). For these analyses, controls were defined as children without a diagnosis of asthma (ever) and without any positive SPTs ($n = 76$ and 271, respectively). We calculated *q* values based on the 193 AS DMCs in URECA and the 85 AS DMCs in INSPIRE.

RNA-sequencing and eQTM studies in URECA

Protocols for processing samples for RNA-sequencing in NECs from the URECA children have been described.²¹ Gene expression data were available in 249 of the children with DNA methylation data. In these samples, 15,643 genes were detected as expressed. To test for cis associations between DNA methylation and gene expression (eQTM studies), we used a linear model that included sex, percentage of epithelial cells, and 3 ancestry principal components as covariates (FDR < 0.05).

Data sharing plans

The gene expression data used in this study are available on the GEO (Gene Expression Omnibus) repository (accession number GSE145505). The DNA methylation array data used in the EWAS are available in GEO (accession numbers GSE217337 and GSE220874 for the EPIC and custom array, respectively). The WGBS data are available in dbGaP (in progress).

RESULTS

Identifying DMRs in WGBS

We performed WGBS in airway epithelial cell DNA from 20 African American children (10 with allergic asthma, 10 without asthma or allergies; 11 years old) from the URECA cohort⁷ and 19 European American young adults (9 with allergic asthma, 10 without asthma or allergies; 18-20 years old) from the COAST cohort.⁸ After QC (see the Methods section), analyses for DMRs between the asthma/allergy cases and nonasthma/nonallergy controls were performed in the African American sample, the European American sample, and the combined sample (see the Methods section for additional details). Overall, we identified 16,611 DMRs that included 199,473 CpGs (see Table E5 in this article's Online Repository at www.jacionline.org).

Selecting high-value CpGs for an Asthma&Allergy custom DNA methylation array

Our goal was to develop a custom array with 50,000 probes for high-value CpGs that would complement and serve as a "booster" for the EPIC array for studies of asthma and allergic diseases, as described in the Methods section and in the Online Repository. After manufacture and array QC, 37,863 CpGs targeted by 42,192 probes remained (Table E5). The distributions of the 37,256 CpGs on the custom array and the 789,290 CpGs on the EPIC arrays that passed final QC are shown by annotation category in Fig 2, A. Of note, 90% of CpGs on the custom array

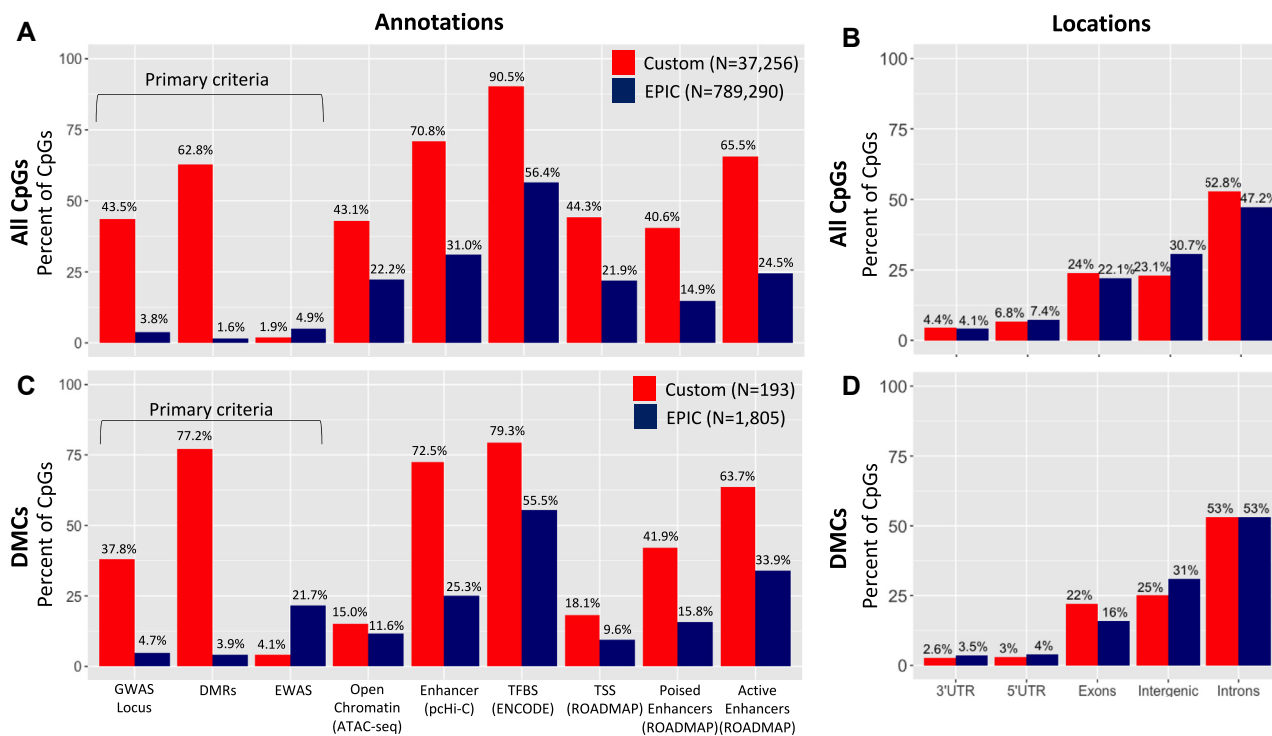


FIG 2. Proportions of custom and EPIC CpGs by primary criteria, functional annotation category, and genomic location. All comparisons were performed with a FET. **A**, Proportions of custom CpGs that passed processing QC in URECA on each array by 3 primary criteria and 6 functional annotation categories. CpGs on the custom array were significantly enriched ($P < 2.2 \times 10^{-16}$) in all categories compared to the EPIC array with the exception of prior EWAS, in which they were depleted on the custom compared to the EPIC ($P < 2.2 \times 10^{-16}$). **B**, The distributions of CpGs by genomic location. Compared to the EPIC, CpGs on the custom array were enriched in introns and exons and depleted in intergenic regions and 5' untranslated regions (UTRs); ($P < 10^{-4}$) but did not differ in any other categories. **C**, Proportions of custom and EPIC DMCs from each array by 3 primary criteria and 6 functional annotation categories. Compared to all CpGs, DMCs from both arrays were depleted at transcription start sites and in areas of open chromatin (custom $P = 1.14 \times 10^{-7}$ and $P = 4.11 \times 10^{-9}$, EPIC $P < 2.2 \times 10^{-16}$ for both, respectively). DMRs were marginally enriched for DMCs compared to all CpGs on the custom array ($P = .066$) but significantly enriched on the EPIC array ($P = 1.54 \times 10^{-6}$). DMCs in prior EWAS of asthma and allergic diseases were modestly enriched on the custom array (FET $P = .059$) and significantly enriched on the EPIC array ($P < 2.2 \times 10^{-16}$). In contrast, CpGs at GWAS loci for asthma and allergic diseases were not enriched among DMCs on the custom array ($P = .32$) and only modestly enriched on the EPIC array ($P = .042$). **D**, The distribution of DMCs by genomic location. The distribution of DMCs on both arrays did not differ from the distributions of all CpGs. ATAC-seq, assay for transposase-accessible chromatin using sequencing.

overlapped with a TFBS, and 94% overlapped with a predicted enhancer, compared to 56% and 54% of the EPIC array CpGs, respectively (Fisher exact test [FET] $P < 10^{-16}$ for both comparisons). Moreover, compared to EPIC CpGs, custom CpGs were enriched in introns (52.8% vs 47.2%) and exons (24.0% vs 22.1%) and depleted in intergenic regions (23.1% vs 30.7%) and 5' untranslated regions (6.8% vs 7.4%) (FET $P < 10^{-4}$ for all comparisons) (Fig 2, B). The CpGs on the custom array proportionally represented DMR-CpGs from the studies in URECA and COAST and combined analysis in WGBS and CpGs from the 3 categories of prior evidence (see Fig E2 in this article's Online Repository at www.jacionline.org).

Evaluating methylation patterns of CpGs on the custom array

To assess the reproducibility of methylation measures of CpGs on the custom array, we compared methylation levels (measured as β values) in airway epithelial cell DNA in the WGBS to the

same CpGs on the custom and the EPIC arrays for the subset of individuals with all 3 (19 URECA subjects) or 2 (17 COAST subjects) measures. We included CpGs with at least 10 WGBS reads that overlapped with an array CpG that passed QC. This resulted in 502,229 CpGs for the WGBS-EPIC comparison in URECA and 24,744 and 20,861 CpGs for the WGBS-custom array comparisons in URECA and COAST, respectively. The β distribution plots were similar and highly correlated between CpGs measured by both WGBS and an array (Spearman $P < 2.2 \times 10^{-16}$ for each) (Fig 3). Notably, however, compared to the EPIC array, CpGs on the custom array were depleted for hypermethylated CpGs and enriched for intermediate methylated (IM) CpGs (β values 20%-80%), an observation that was corroborated by the WGBS data.

To more broadly assess EPIC and custom array DNA methylation patterns, we compared the distributions of β values across different tissues using available data for the EPIC array in DNA from airway epithelial cells (this study): airway smooth muscle cells²⁴ and buccal swabs, placenta cells, and cord blood;¹⁰ for the

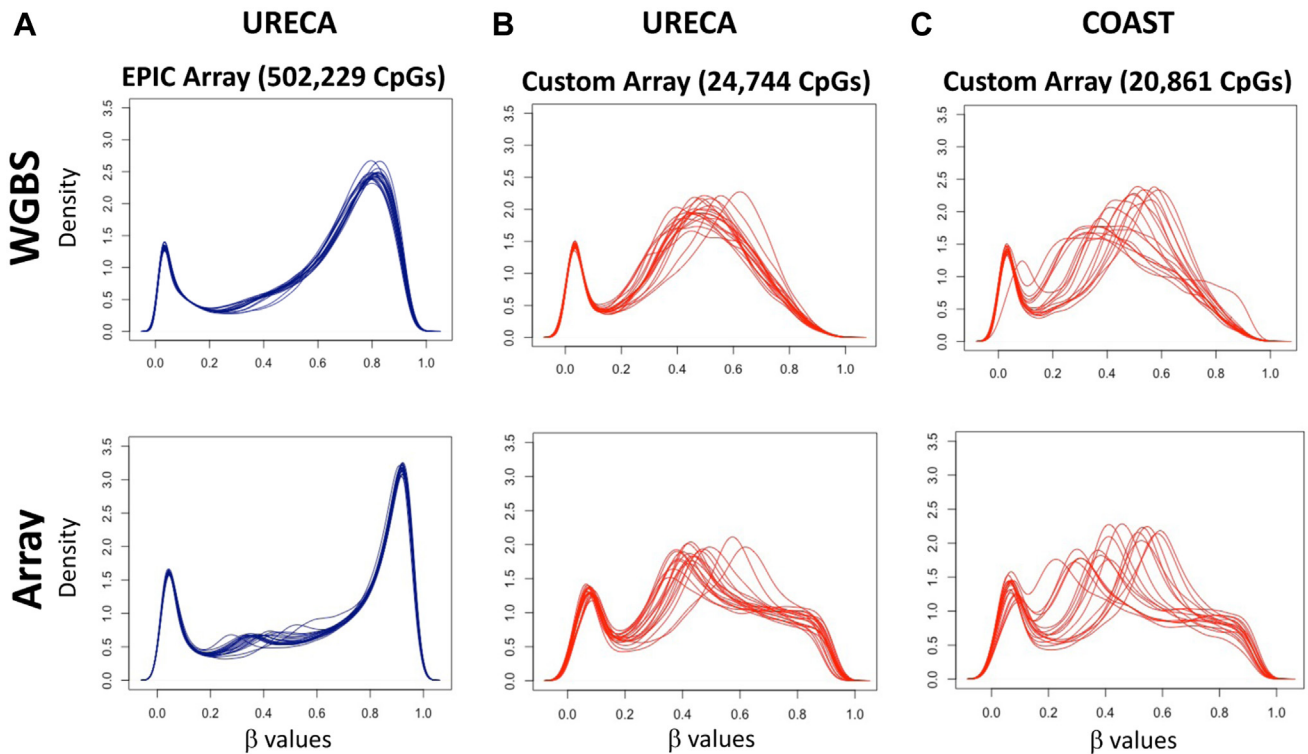


FIG 3. Comparison of methylation level (β value) distributions from the EPIC and custom arrays to the WGBS data. Percentage of methylation is shown on the x-axis and density is shown on the y-axis. The number of CpGs in each comparison is shown at the top of each pair, which includes the number of sites with at least 10 mapped reads in the WGBS data. **A**, β value distribution for the WGBS data compared to the EPIC array for the 19 URECA subjects assayed using both platforms that passed QC. Spearman $\rho = 0.82$ ($P < 2.2 \times 10^{-16}$). **B**, β value distribution for the WGBS data compared to the custom array for the same 19 URECA subjects. Spearman $\rho = 0.83$ ($P < 2.2 \times 10^{-16}$). **C**, β value distribution for the WGBS data compared to the custom array for the 17 COAST samples that passed QC. Spearman $\rho = 0.82$ ($P < 2.2 \times 10^{-16}$).

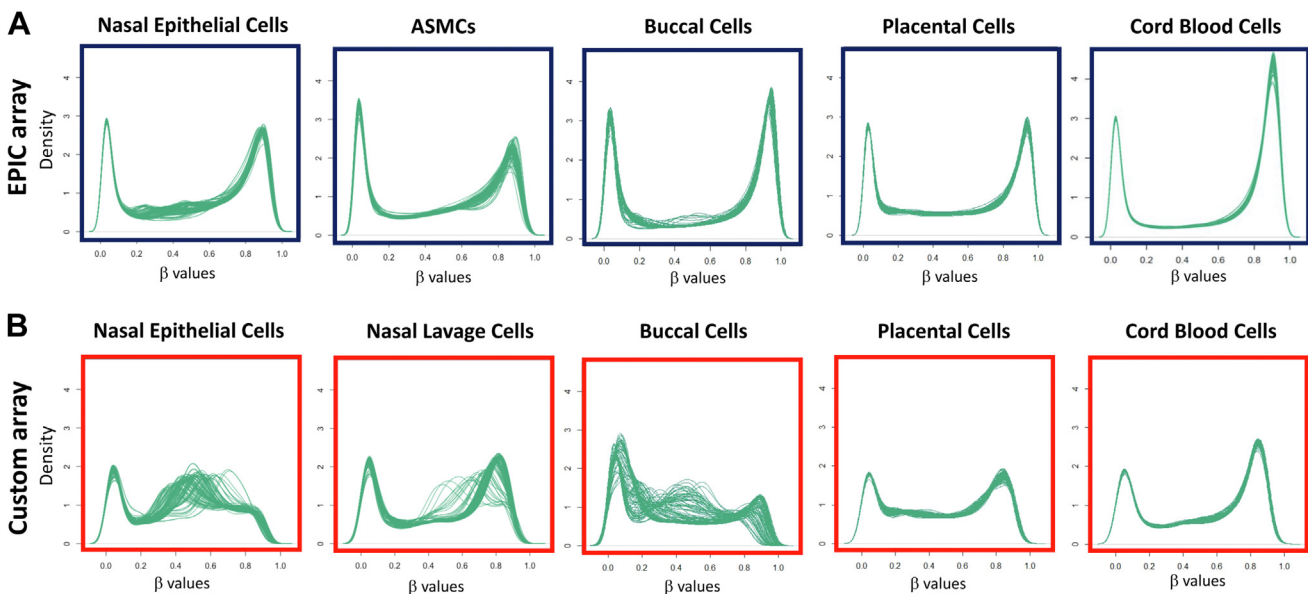


FIG 4. Cross-tissue comparisons of methylation levels for CpGs on the EPIC and custom arrays. Density distribution plots of methylation levels, measured as β values, in 96 individuals after quantile (custom) or quantile + SWAN (EPIC) normalizations (see the Methods section). The x-axis shows the proportion of methylation at CpG sites on the EPIC (**A**) and custom (**B**) arrays for individuals. All DNA methylation data were processed using the same pipeline. The NEC DNA (EPIC and custom) and nasal lavage cell DNA (custom) are shown for the same randomly selected URECA children. The buccal, placenta, and cord blood cells were from infants in the VCSIP study. SWAN, Subset-quantile Within Array Normalization.

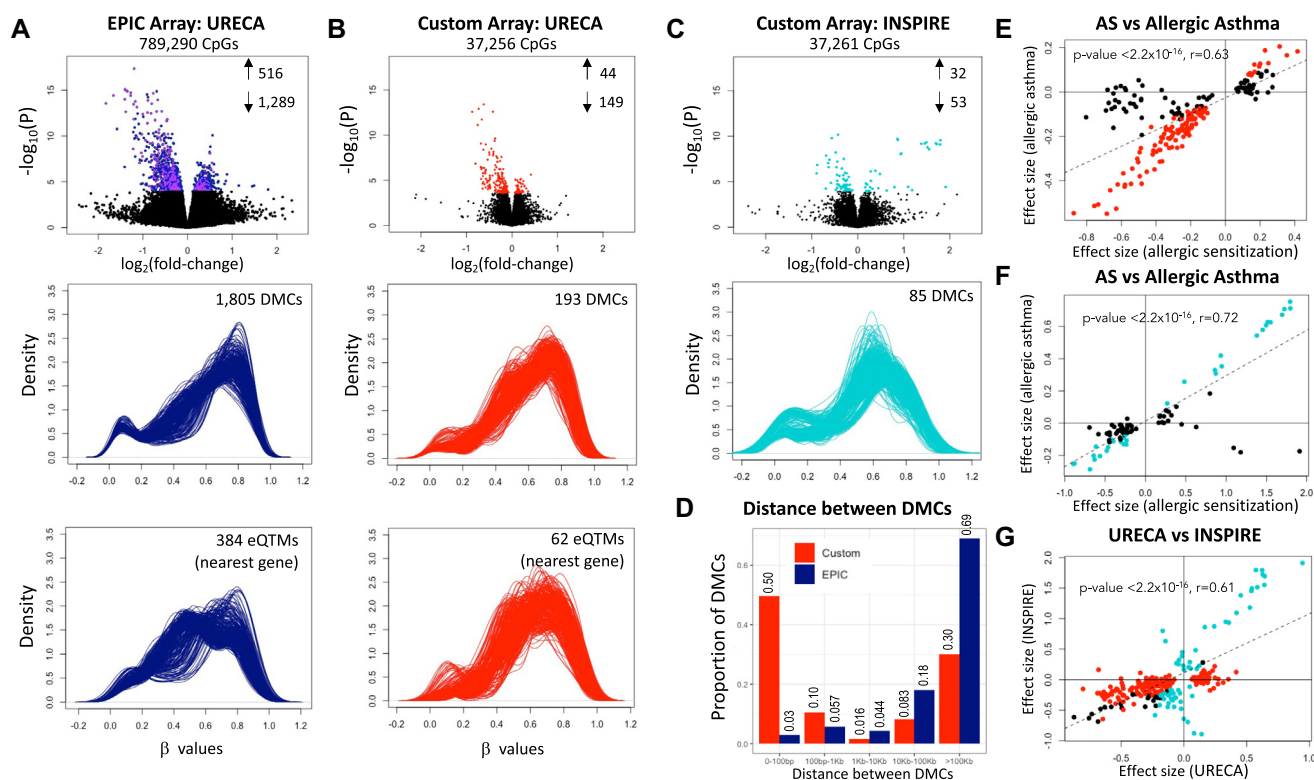


FIG 5. AS EWAS results in URECA and INSPIRE children using the EPIC and custom arrays. **A-C, upper,** Volcano plots showing the \log_2 -fold change in methylation (M values) by proportion positive SPTs (x -axis) and the $-\log_{10}(P)$ value from the EWAS (y -axis). Significant DMCs at a q -value threshold of 0.05 are shown in blue or purple (EPIC and high-value EPIC, respectively) in URECA, and custom in red (URECA) or turquoise (INSPIRE). The number of AS DMCs that were hypermethylated or hypomethylated are shown as up and down arrows, respectively. **A-C, middle,** Density plots of the DMC β values in each EWAS. **A-B, lower,** Density plots of β values for DMCs that are eQTMs for their nearest genes (also see Fig E9). **D,** Distributions of the distances of DMCs to the next nearest DMC on the custom (red) and EPIC (blue) arrays. The distance to nearest DMC is shown on the x -axis; the proportion of DMCs in each distance bin is shown on the y -axis. **E-F,** Correlation of effect sizes (log-fold change) of AS DMCs and allergic asthma in URECA and INSPIRE. The effect size of DMCs for AS (x -axis) and allergic asthma (y -axis) are shown. Red or turquoise dots were associated with allergic asthma at $q \leq 0.05$ in URECA or INSPIRE, respectively; black dots are AS-only DMCs. **G,** Correlation of effect sizes (log-fold change) between DMCs identified in the INSPIRE AS EWAS (x -axis) and the URECA AS EWAS (y -axis). Red or turquoise points are AS DMCs only in URECA or INSPIRE, respectively, at a q -value threshold of 0.05; black points are AS DMCs in both at $q \leq 0.05$. The dashed lines in E-G are the correlation lines.

custom array in DNA from NECs and nasal lavage cells from URECA children; and buccal cells, cord blood cells, or placenta tissue from infants in the VCSIP cohort.¹⁰ The β distributions were similar across cell sources for the EPIC array (Fig 4, A) but showed varying patterns between cell types on the custom array (Fig 4, B). Whereas an average of 68% (range 60%-79%) of CpGs on the EPIC array were either hypomethylated (0%-20%) or hypermethylated (80%-100%) in all 6 cell types, most CpGs on the custom array (66% in NECs and 52% on average across cell types) were IM CpGs with depletions of β values between both 0%-10% (24% in NECs and 29% on average) and 90%-100% (10% in NECs and 20% on average). In fact, these β value distributions of CpGs on the EPIC array are remarkably similar to those in 9 GTEx (Genotype-Tissue Expression Project) tissues (see Fig E3 in this article's Online Repository at www.jacionline.org).²⁵ To confirm that differences between the 2 arrays were a result of the selection criteria used for designing the array, we used the same criteria for the CpGs on the EPIC array. The 26,905 high-value EPIC CpGs also showed an enrichment for IM CpGs and a depletion of both

hypomethylated and hypermethylated CpGs (see Fig E4 in this article's Online Repository at www.jacionline.org).

EWAS of AS in 280 Multiancestry children using the custom and EPIC arrays

Our ultimate goal was to develop a DNA methylation array that could detect asthma- or allergy-associated differential methylation that is missed by the content on the EPIC array. AS is an IgE response to allergens²⁶ and a crucial step in the development of allergic diseases and common phenotypes of asthma.^{21,27,28} Moreover, as a quantitative trait, it provides more power to detect DMCs compared to binary traits (such as asthma). We conducted an EWAS of AS (see the Methods section) in airway epithelial cell DNA collected from 280 11-year-old URECA children using the custom and the EPIC arrays. The demographic and clinical description of the URECA children is shown in Table I and Fig E5 in this article's Online Repository at www.jacionline.org.

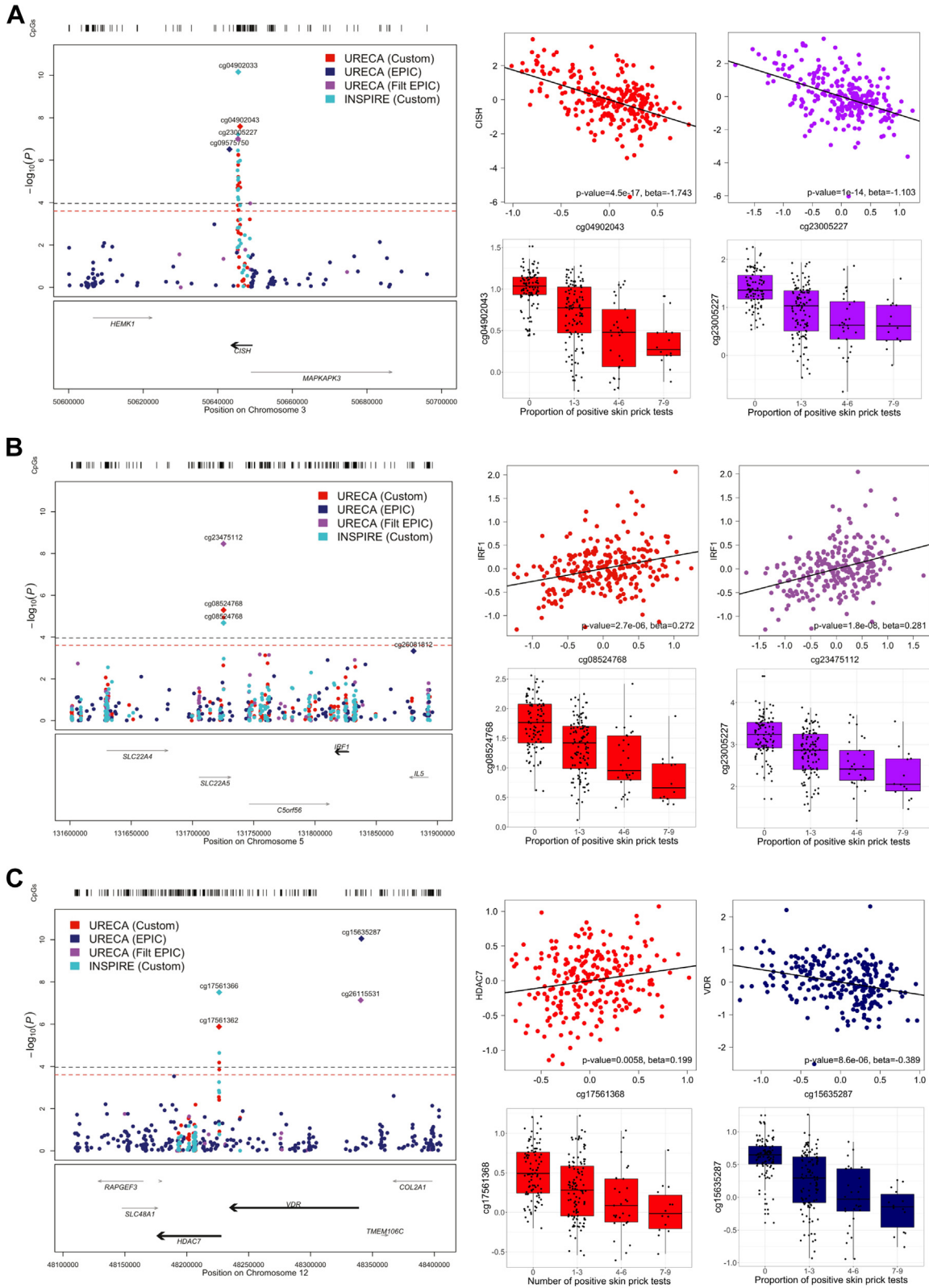


FIG 6. DMCs and eQTMs at 3 exemplary loci. In each panel, locus zoom plots show the DMCs from the 3 AS EWAS using the EPIC array (URECA; blue points = non-high-value CpGs; purple points = high-value [filtered] CpGs) and custom array (URECA, red points; INSPIRE, turquoise points). The genomic locations and genes are shown on the x-axis; EWAS P values are shown on the y-axis. The dashed horizontal lines show the 0.05 q -value threshold for the custom array in URECA (red) and for the EPIC array in URECA and custom array in INSPIRE (blue and purple overlaid). The bar code at the top of the plot shows the

Following QC of the array, we performed 2 EWAS of AS in the same URECA children using the EPIC and custom arrays (see the Methods section). Using a q -value threshold of 0.05, we identified 1805 DMCs using the EPIC array and 193 DMCs using the custom array (Fig 5, A and B), revealing an enrichment for AS DMCs on the custom compared to the EPIC array (0.50% vs 0.23% of CpGs on each array, respectively; FET $P < 2.2 \times 10^{-16}$). Among the DMCs on the EPIC array, 295 (16%) were high-value EPIC CpGs, a significant enrichment compared to 3.5% of high-value CpGs among all CpGs on the EPIC array (FET $P < 2.2 \times 10^{-16}$). In contrast to the different β distributions of CpGs on each array (Fig 4, A), the β distributions of DMCs were similar (Fig 5, A and B, middle panels), showing a nearly complete depletion of both hypomethylated and hypermethylated CpGs and an enrichment of IM CpGs among AS DMCs in airway epithelial cells. The DMCs on both arrays were distributed across the autosomes (see Fig E6, A and B in this article's Online Repository at www.jacionline.org), however, custom array DMCs showed spikes of association signals compared to the sparse distribution of EPIC array DMCs. In fact, 50% of DMCs on the custom array were within 100 bp of the next nearest DMC compared to only 3% of DMCs on the EPIC array (Fig 5, D). Regional clustering of DMCs on the custom array provides greater confidence in and internal validation of EWAS results.

Most of the functional annotation categories of DMCs were proportional to all CpGs on each array (Fig 2, A and C; see Fig 2 legend for all statistical test results). However, DMCs from both arrays were depleted at transcription start sites and in areas of open chromatin compared to all CpGs on the arrays. Among the primary criteria (prior evidence), CpGs in DMRs were modestly enriched for DMCs compared to all CpGs on the custom array and DMCs in prior EWAS studies were modestly enriched on the custom array but significantly enriched on the EPIC array. In contrast, CpGs at GWAS loci were not enriched among DMCs on the custom array and only modestly enriched on the EPIC array compared to all CpGs on the array. The genomic locations of the DMCs were proportional to all CpGs on the arrays (Fig 2, B and D).

Because AS is an important step in the development of childhood asthma, we next asked whether AS DMCs were also associated with allergic asthma. Considering only the AS DMCs on the EPIC array, 1155 (64%) were also associated with allergic asthma at $q < 0.05$ (1805 tests). Among the AS DMCs on the custom array, 115 (60%) were also associated with allergic asthma at $q < 0.05$ (193 tests). The effect sizes were significantly correlated between AS DMCs and allergic asthma DMCs (Spearman $\rho = 0.63$). Among the CpGs that were DMCs for both AS and allergic asthma, all showed the same direction of effect; among all AS DMCs, 93% showed the same direction of effect with allergic asthma ($r = 0.84$) (Fig 5, E). The sample distributions of AS and allergic asthma are shown in Fig E7 in this article's Online Repository at www.jacionline.org.

Validating the custom array in the INSPIRE cohort

The URECA cohort is diverse with respect to ancestry (see Fig E8 in this article's Online Repository at www.jacionline.org) but includes <1% non-Hispanic White children.⁷ Therefore, to both replicate results of the EWAS in URECA children and assess the performance of the custom array in a primarily non-Hispanic White population, we studied 5- to 7-year-old children from the INSPIRE cohort.⁹ An EWAS was performed using the custom array and NEC DNA from 474 children with measures of AS (see the Methods section). See Table I for demographic and clinical characteristics, Fig E5 in this article's Online Repository at www.jacionline.org for sensitization rates by allergen, and the Online Repository for details on processing and QC of the array data.

At a q -value threshold of <0.05 (37,261 tests), 85 CpGs on the custom array were DMCs (0.2% of CpGs). The \pm distribution of the DMCs (Fig 5, C) and the patterns of associations (Fig E6, C in this article's Online Repository at www.jacionline.org) were similar to those observed in URECA. Among the AS DMCs in INSPIRE, 38% were also allergic asthma DMCs ($q < 0.05$ based on 85 tests), all with the same direction of effect. Among all AS DMCs in INSPIRE, 93% had the same direction of effect with allergic asthma (Fig 5, F). Moreover, among the AS DMCs in URECA, 25 (13%) were also AS DMCs in INSPIRE, which was a highly significant enrichment ($P < 2.2 \times 10^{-16}$), and all AS DMCs had concordant directions of effect. Finally, the effect sizes of the 253 CpGs that were DMCs in either the URECA or INSPIRE AS EWAS were highly correlated ($r = 0.61$; $P < 2.2 \times 10^{-16}$) (Fig 5, G). The high reproducibility of EWAS results in 2 cohorts demonstrates that the high-value CpGs on the custom array identified AS DMCs that are robust to ancestry, ascertainment strategies, age at sampling, and geography. The sample distributions of AS and allergic asthma are shown in Fig E7 in this article's Online Repository at www.jacionline.org.

Evaluating the biological significance of DMCs in airway epithelial cells

Because nearly all CpGs on the custom array were selected to overlap with a TFBS or enhancer mark, we expected these CpGs to be correlated with the expression of genes in airway epithelial cells more often than CpGs on the EPIC array.²⁹ To test this, we used gene expression data²¹ in the same cells as those used for DNA methylation studies in 249 of the URECA children and defined 2 sets of genes from among the 15,551 genes detected as expressed in these cells: the nearest gene to each CpG and the pChI-C—defined target gene in airway epithelial cells (see the Methods section). The latter identifies putative enhancers that physically interact with target gene promoters and account for the 3-dimensional structure of the genome that allows for distal regulatory elements to interact with and regulate the activity of promoters over large distances (eg, up to 1 Mb or more).

location of CpGs at this locus on both arrays. The most significant DMC in the region for a given EWAS (cohort and platform; see legend) is illustrated by a *diamond*; additional DMCs appear as *circles*. Correlations between methylation levels (x-axis) of the lead custom and high-value EPIC CpGs in URECA and expression of relevant genes (y-axis) in epithelial cells from URECA children are shown in the *upper right* of each panel, and box plots of the proportion of positive SPTs (x-axis) and methylation levels for the lead custom and high-value EPIC CpGs in URECA (from the EWAS) are shown in the *lower right* of each panel. **A**, *CISH* locus on chromosome 3. **B**, *SLC22A5-IRF1* locus on chromosome 5. **C**, *HDAC7-VDR* locus on chromosome 12.

Among the nearest expressed genes to DMCs on the custom (98 genes) and EPIC (1,449 genes) arrays, 63 were the nearest gene to CpGs on both arrays. Among the pcHi-C target genes for DMCs on the custom (245 genes) or EPIC (1,155 genes) arrays, 95 were target genes of CpGs on both arrays. Thus, although no CpGs overlapped between the arrays, 101 of nearest or pcHi-C target genes were shared on both arrays (see Table E6 in this article's Online Repository at www.jacionline.org). The 318 nearest or pcHi-C target genes on the custom array that were expressed in airway epithelial cells were enriched in 16 pathways (FDR < 0.05), including "T_H1 and T_H2 cell differentiation," "JAK-STAT signaling," "T_H17 cell differentiation," and "Viral protein interaction with cytokine and cytokine receptor" (see Table E7 in this article's Online Repository at www.jacionline.org and the Online Repository). In contrast, the 2,366 nearest or pcHi-C target genes on the EPIC array that were expressed in airway epithelial cells were not enriched for any pathways (smallest FDR = 0.13). These findings further demonstrate that the CpGs on the custom array are enriched in pathways relevant to asthma and allergic disease.

For each CpG-gene pair, we tested for correlation between DNA methylation and expression levels of their nearest and pcHi-C target gene(s) to identify eQTM. Significantly more CpGs on the custom array were eQTMs with their nearest gene (23%) or pcHi-C target gene (16%) compared to CpGs on the EPIC array (11% and 9%, respectively; FET $P < 2.2 \times 10^{-16}$ in both analyses) (see the Online Repository). The high-value EPIC CpGs were enriched for eQTMs compared to all EPIC CpGs (20% and 12%, respectively; FET $P < 2.2 \times 10^{-16}$ in both analyses). Although more AS DMCs were eQTMs compared to non-DMCs on both arrays, significantly more were on the custom compared to the EPIC array (nearest gene 35% vs 20% [FET $P = .0019$] and pcHi-C 22% vs 15% [FET $P = .0082$], respectively) (see Table E8 in this article's Online Repository at www.jacionline.org). The β distributions for DMCs on the custom and EPIC arrays that were eQTMs were further enriched for IM CpGs and depleted of CpGs at the extremes of the distribution (Fig 5 and see Fig E9 in this article's Online Repository at www.jacionline.org).

Further insights into the epigenetic regulation of gene expression at EWAS loci

To examine the DMCs and their associations with gene expression more closely, we generated regional association plots for the 10 most significant DMCs in the URECA EWAS using the EPIC and custom arrays. The 10 most significant EPIC DMCs were at 10 loci (see Fig E10 in this article's Online Repository at www.jacionline.org). Seven were "solitary" with no other DMCs within 500 kb and 3 had 1 other DMC within 6.5 kb. Among the solitary DMCs, 6 were high-value EPIC CpGs and all 10 were reported as DMCs in airway epithelial cells in previous EWAS of asthma or allergic phenotypes. Six of the 10 DMCs were within genes and 4 were intergenic. CpGs from the custom array were present in 4 of these regions, but none were AS DMCs. In contrast, the 10 most significant custom DMCs were at 6 loci and only 2 were solitary (see Fig E11 in this article's Online Repository at www.jacionline.org). One solitary DMC in *ALOX15* was near a high-value and a non-high-value EPIC DMC; the other solitary CpG in *PDE6A* was also a DMC in the INSPIRE EWAS. The

other 4 loci had spikes of association, often with a combination of DMCs from both arrays in URECA and the custom array in both URECA and INSPIRE. Eight of the 10 most significant DMCs were within genes; 2 were intergenic.

We selected 3 additional regions as examples of patterns of associations (see Table E9 in this article's Online Repository at www.jacionline.org). The first includes a spike of 11 custom array DMCs in URECA, 12 in INSPIRE, and 1 high-value EPIC DMC in exons 2 and 3 of *CISH*, a member of the SOCs, suppressor of cytokines, family of negative regulators of cytokine signaling (Fig 6, A). There was 1 nearby DMC from the EPIC array downstream of *CISH*. The lead custom and EPIC URECA DMCs were eQTMs for *CISH*, with increased methylation associated with decreased gene expression and fewer allergic sensitizations. The EPIC high-value CpG at this locus was identified in previous asthma/allergy EWAS,^{30,31} but this locus has not been associated with asthma or allergic diseases in GWAS. *CISH* expression is induced by IL-13 in bronchial epithelial cells and macrophages³² and has been implicated in eosinophil physiology³³ and eosinophilic inflammation.³² Our studies are consistent with these findings and further indicate that increased expression of *CISH* is associated with increased sensitization to allergens and that the regulation of *CISH* expression may be epigenetically mediated.

A second locus included 1 custom DMC in URECA and INSPIRE, a second custom DMC in URECA, and 1 high-value EPIC DMC in an intron of *SLC22A5* (Fig 6, B). The lead DMC was a high-value EPIC CpG, which was also identified in prior EWAS in nasal epithelium.^{30,31} None of the DMCs were eQTMs for *SLC22A5*, but all 3 were in a region that physically interacted with the promoter of *IRF1*, 91 kb away, and were eQTMs for *IRF1* ($P = 2.4 \times 10^{-7}$, $\beta = 0.27$ and $P = 2.7 \times 10^{-6}$, $\beta = 0.27$). This region is at a GWAS locus for adult- and childhood-onset asthma and hay fever^{34,35} (Table E9), possibly with sex-specific effects.³⁶ Genetic variation in the *IRF1* has been associated with increased expression of proinflammatory genes and IL-13 secretion in peripheral blood cells.³⁷ Our studies further demonstrate long-range epigenetic regulation of *IRF1* expression in airway epithelial cells, with increased methylation levels associated with increased gene expression and decreased numbers of sensitizations.

At a third locus, we observed a spike of 3 custom array DMCs in URECA and 2 in INSPIRE upstream of *HDAC7* (Fig 6, C). One custom DMC (upstream of *HDAC7* and downstream of *VDR*) was an eQTM for *HDAC7*. Histone deacetylases (HDACs) have diverse functions, including regulation of inflammatory genes,³⁸ and impaired barrier function in individuals who are asthmatic was both induced by HDAC activity and reversed by inhibition of endogenous HDAC.³⁹ In this study, increased methylation in *HDAC7* was associated with increased expression of *HDAC7* and fewer sensitizations. At this extended locus, 2 EPIC DMCs were in an intergenic region 42 kb upstream of *VDR*. Vitamin D deficiency in childhood has been associated with increased risk for persistent asthma⁴⁰ and high-dose vitamin D supplementation in pregnancy reduced risk of recurrent wheeze in childhood.⁴¹ Here, increased methylation (EPIC array) near *VDR* was associated with decreased *VDR* expression and sensitization to fewer allergens. The *HDAC7-VDR* region has been identified in GWAS of childhood- and adult-onset asthma.³⁴ This locus demonstrates the complementarity of the custom and EPIC arrays in detecting differential methylation relevant to asthma and allergic diseases.

DISCUSSION

The epigenome plays a critical role in regulating gene expression in a context-specific manner, such as in specific cell types or in the presence of environmental exposures or disease states. DNA methylation, for example, is responsive to disease-promoting exposures in both *in vitro* cell models^{24,42-44} and *ex vivo* cells from individuals who are healthy and those with disease.^{1,45-48} Yet the DNA methylome has been underexplored and largely limited to the CpGs on commercial arrays. As a result, it is unknown whether the 800,000 or fewer CpGs interrogated in nearly all EWAS to date include CpGs most relevant to any specific exposure, disease, or cell type. Here, we addressed this question by designing a custom Asthma&Allergy DNA methylation array with 37,863 high-value CpGs that are not on the commercial EPIC array. Our study design allowed us to make important observations regarding features of high-value CpGs and to propose a pipeline that can be used to prioritize CpGs from among the more than 28 million CpGs in the human genome.

Although we did not use methylation levels as a selection criterion, the CpGs selected were significantly enriched for IM CpGs (β values 20%-80%). The CpGs that were DMCs and eQTM were further enriched for IM CpGs on both the EPIC and custom arrays, suggesting that this feature is a signature of high-value, variable, and likely functional CpGs in the genome. In contrast, the non-IM CpGs may be more functionally constrained and less responsive to environmental perturbations. Our results are supported by previous WGBS studies showing that IM CpGs were more likely to be tissue-specific, to vary between individuals, and to play important roles in gene regulation.^{29,49} Our study further revealed that CpGs with intermediate levels of methylation are also more likely to be associated with AS, an important clinical phenotype that reflects both an immune response to past allergen exposures²⁶ and a risk factor for the development of asthma and allergic diseases.^{21,27,28} Although the CpGs on the custom array were overall enriched for IM CpGs, AS DMCs were further enriched, with a depletion of very hypomethylated (0%-20% methylation) and very hypermethylated (80%-100%) CpGs. This important observation was further supported by EWAS results using the EPIC array. Whereas the EPIC array is enriched for hypomethylated and hypermethylated CpGs in all tissues, the AS DMCs from the EPIC array were depleted for CpGs at both extremes and enriched for IM CpGs, similar to the DMCs on the custom Asthma&Allergy array. We further showed that the 26,905 high-value EPIC CpGs were also enriched for IM CpGs and enriched among AS DMCs: 16.3% of the DMCs in the AS EWAS were among the high-value EPIC CpGs compared to 3.4% of all EPIC CpGs. These results, combined with previous studies,^{29,49} could justify filtering CpGs on the EPIC array to include just IM CpGs ($\beta = 20\%$ -80%) to enrich for functional CpGs and increase power to detect DMCs associated with exposures or disease outcomes. As proof-of-concept, including only IM EPIC CpGs and using a q value of 0.05, 2,182 CpGs were AS DMCs (0.70% of all IM CpGs), more than 3 times the proportion of DMCs among all CpGs (1,805 AS DMCs; 0.23% of all EPIC CpGs). Finally, DMCs that were eQTMs were further enriched for IM CpGs (Figs 5 and E8), supporting their role in gene regulation.

In addition to the enrichment of IM CpGs on the custom compared to the EPIC array, the different patterns of associations with AS were notable. Whereas DMCs on the EPIC array were

more sparsely distributed throughout the genome, the DMCs on the custom array often showed spikes of associations, analogous to GWAS peaks. The spatial organization among DMCs on the custom array provides confirmatory evidence for their veracity, which is generally not available for EPIC CpGs (Fig 5, D). Another distinguishing feature of the DMCs on the custom array was their enrichment in exons and depletions in intergenic regions compared to DMCs on the EPIC array. This is consistent with results of an earlier study that observed enrichments of IM CpGs in exons, where exon expression at those sites correlated with DNA methylation.²⁹

The CpGs included on the custom Asthma&Allergy array were selected based on prior evidence of associations with asthma or allergic diseases using 3 classes of evidence: within a DMR in our WGBS study, in a previous published EWAS, or at published GWAS loci. Overall, 62% of CpGs on the custom array were selected from DMRs compared to 1.6% of all CpGs on the EPIC array, yet the DMR CpGs were enriched among DMCs on both the custom (77.2%) and EPIC (3.9%) arrays. In contrast, few previous EWAS CpGs were on the custom array because we excluded all CpGs on the EPIC from the custom array, yet we still observed a modest enrichment of EWAS CpGs among custom array DMCs (1.9% overall vs 4.1% of DMCs) and a highly significant enrichment among DMCs on the EPIC array (4.9% overall vs 21.7% of DMCs). These represent replications of results from previous EWAS of asthma and allergic phenotypes in the URECA children. Surprisingly, CpGs at asthma and allergy GWAS loci were only modestly enriched among EPIC DMCs and not at all enriched among the custom DMCs, even though the EWAS in URECA and INSPIRE revealed many examples of AS DMCs at important GWAS loci. The lack of enrichment of custom array CpGs at GWAS loci among DMCs may reflect the less direct evidence for DNA methylation mechanisms at GWAS loci compared to CpGs from previous EWAS or DMR studies, the large proportion of CpGs at GWAS loci included on the custom array (44% of all CpGs), and the enrichment of AS DMCs among intergenic regions, whereas the genetic variants detected by GWAS are enriched in intergenic regions. It is also possible that the selection of these CpGs, which is based on an extreme sampling design enriched for CpGs that are near variants, associated with specific endotypes of asthma that may not reach genome-wide significance in GWAS, which typically define asthma as “self-reported” or “doctor-diagnosed.”

Our study has limitations. First, due to cost and technical limitations, the custom array included only 37,863 CpGs of the 92,024 high-value CpGs that we identified with our pipeline. Therefore, together with the 26,905 high-value EPIC CpGs, we only surveyed 54% of potential high-value CpGs in airway epithelium that are relevant to asthma and allergic diseases. The next generation (version 1.1) of the custom array includes ~8,000 more high-value CpGs, reducing this gap, although alternative array designs or other methods that allow the inclusion of the 118,929 high-value CpGs would be a powerful tool for assessing epigenetic mechanisms in these conditions. Second, we have not yet evaluated the performance of this array in blood cells, which is a much more easily sampled tissue. Those studies are currently underway. Lastly, we did not functionally validate the correlations between CpGs and gene expression or identify the “causal” CpG(s) driving the correlations, which should be considered in follow-up studies of any of the EWAS regions reported here.

In summary, our study revealed high-value CpGs in airway epithelial cells that are not interrogated on the EPIC array. The allergic sensitization EWAS revealed potential epigenetic regulatory mechanisms both for genes previously implicated in asthma or allergy pathogenesis but not identified in GWAS (eg, *CISH*,^{32,33} *NEEDAL*,^{50,51} *PRDM10*^{52,53}) as well as for genes at replicated GWAS loci (eg, *IRF1*, *CLEC16A*, *HDAC7/IVDR*, *ALOX15*, *SLC22A5*) (Table E9). These associations were robust to race/ethnicity, ascertainment, age, and geography. These results offer insight into specific mechanisms through which epithelial cells contribute to AS by highlighting their central role in mediating the effects of environmental exposures on AS and suggesting targets for functional studies of the epithelial component of allergic disease as well as potential therapeutic targets. We suggest that similar pipelines can be used to identify high-value CpGs for other disease groups or specific environmental exposures leading to the discovery of currently invisible portions of the epigenome that are relevant to human health.

We wish to thank our ECHO colleagues, the medical, nursing, and program staff, and the children and families participating in the ECHO cohorts. We also acknowledge the contribution of the following ECHO program collaborators: ECHO Components—Coordinating Center: Duke Clinical Research Institute, Durham, NC: Smith PB, Newby KL.

Clinical implications: These studies identified AS-associated differentially methylated CpGs and their target genes in airway epithelium, providing potential epigenetic mechanisms in the development of allergic diseases and suggesting novel drug targets.

REFERENCES

- Prunicki M, Cauwenberghs N, Lee J, Zhou X, Movassagh H, Noth E, et al. Air pollution exposure is linked with methylation of immunoregulatory genes, altered immune cell profiles, and increased blood pressure in children. *Sci Rep* 2021;11:4067.
- Li Y, Hamilton KJ, Wang T, Coons LA, Jefferson WN, Li R, et al. DNA methylation and transcriptome aberrations mediated by ERalpha in mouse seminal vesicles following developmental DES exposure. *Proc Natl Acad Sci U S A* 2018;115:E4189-98.
- Tobi EW, Sliker RC, Luijk R, Dekkers KF, Stein AD, Xu KM, et al. DNA methylation as a mediator of the association between prenatal adversity and risk factors for metabolic disease in adulthood. *Sci Adv* 2018;4:eaa04364.
- Li M, Zou D, Li Z, Gao R, Sang J, Zhang Y, et al. EWAS Atlas: a curated knowledgebase of epigenome-wide association studies. *Nucleic Acids Res* 2019;47:D983-8.
- Holgate ST. The sentinel role of the airway epithelium in asthma pathogenesis. *Immunol Rev* 2011;242:205-19.
- Akdis CA. The epithelial barrier hypothesis proposes a comprehensive understanding of the origins of allergic and other chronic noncommunicable diseases. *J Allergy Clin Immunol* 2022;149:41-4.
- Gern JE, Visness CM, Gergen PJ, Wood RA, Bloomberg GR, O'Connor GT, et al. The Urban Environment and Childhood Asthma (URECA) birth cohort study: design, methods, and study population. *BMC Pulm Med* 2009;9:17.
- Lemanske RF Jr. The Childhood Origins of Asthma (COAST) study. *Pediatr Allergy Immunol* 2002;13:38-43.
- Larkin EK, Gebretsadik T, Moore ML, Anderson LJ, Dupont WD, Chappell JD, et al. Objectives, design and enrollment results from the Infant Susceptibility to Pulmonary Infections and Asthma Following RSV Exposure Study (INSPIRE). *BMC Pulm Med* 2015;15:45.
- Shorey-Kendrick LE, McEvoy CT, O'Sullivan SM, Milner K, Vuylsteke B, Tepper RS, et al. Impact of vitamin C supplementation on placental DNA methylation changes related to maternal smoking: association with gene expression and respiratory outcomes. *Clin Epigenetic* 2021;13:177.
- Kechin A, Boyarskikh U, Kel A, Filipenko M. cutPrimers: a new tool for accurate cutting of primers from reads of targeted next generation sequencing. *J Comput Biol* 2017;24:1138-43.
- Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for bisulfite-seq applications. *Bioinformatics* 2011;27:1571-2.
- 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature* 2015;526:68-74.
- Amemiya HM, Kundaje A, Boyle AP. The ENCODE blacklist: identification of problematic regions of the genome. *Sci Rep* 2019;9:9354.
- ENCODE Project Consortium, Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shoresh N, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* 2020;583:699-710.
- Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. *Nature* 2015;518:317-30.
- Helling BA, Sobreira DR, Hansen GT, Sakabe NJ, Luo K, Billstrand C, et al. Altered transcriptional and chromatin responses to rhinovirus in bronchial epithelial cells from adults with asthma. *Commun Biol* 2020;3:678.
- 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.
- McKenna C. Factor analysis in high dimensional biological data with dependent observations. arXiv:200911134v1. Available at: <https://arxiv.org/abs/2009.11134>. Accessed December, 2021.
- McKenna C, Nicolae DL. Accounting for unobserved covariates with varying degree of estimability in high dimensional experimental data. *Biometrika* 2019;106:823-40.
- Altman MC, Calatroni A, Ramratnam S, Jackson DJ, Presnell S, Rosasco MG, et al. Inner City Asthma Consortium. Endotype of allergic asthma with airway obstruction in urban children. *J Allergy Clin Immunol* 2021;148:1198-209.
- Phipson B, Lee S, Majewski IJ, Alexander WS, Smyth GK. Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. *Ann Appl Stat* 2016;10:946-63.
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;43:e47.
- Thompson EE, Dang Q, Mitchell-Handley B, Rajendran K, Ram-Mohan S, Solway J, et al. Cytokine-induced molecular responses in airway smooth muscle cells inform genome-wide association studies of asthma. *Genome Med* 2020;12:64.
- Oliva M, Demandelis K, Jasnime F, Lu Y, Hahsan H, Kibriya M, et al. Genetic regulation of DNA methylation across tissues reveals thousands of molecular links to complex traits. *Nature Genetics* 2023;55:112-22.
- van Ree R, Hummelshoj L, Plantinga M, Poulsen LK, Swindle E. Allergic sensitization: host-immune factors. *Clin Transl Allergy* 2014;4:12.
- Illi S, von Mutius E, Lau S, Nickel R, Niggemann B, Sommerfeld C, et al. The pattern of atopic sensitization is associated with the development of asthma in childhood. *J Allergy Clin Immunol* 2001;108:709-14.
- Jackson DJ, Gern JE, Lemanske RF Jr. The contributions of allergic sensitization and respiratory pathogens to asthma inception. *J Allergy Clin Immunol* 2016;137:659-65, quiz 66.
- Elliott G, Hong C, Xing X, Zhou X, Li D, Coarfa C, et al. Intermediate DNA methylation is a conserved signature of genome regulation. *Nat Commun* 2015;6:6363.
- 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.
- Forno E, Wang T, Qi C, Yan Q, Xu CJ, Boutaoui N, et al. DNA methylation in nasal epithelium, atopy, and atopic asthma in children: a genome-wide study. *Lancet Respir Med* 2019;7:336-46.
- Takeshima H, Horie M, Mikami Y, Makita K, Miyashita N, Matsuzaki H, et al. CISH is a negative regulator of IL-13-induced CCL26 production in lung fibroblasts. *Allergol Int* 2019;68:101-9.
- Burnham ME, Koziol-White CJ, Esnault S, Bates ME, Evans MD, Bertics PJ, et al. Human airway eosinophils exhibit preferential reduction in STAT signaling capacity and increased CISH expression. *J Immunol* 2013;191:2900-6.
- Pividori M, Schoettler N, Nicolae DL, Ober C, Im HK. Shared and distinct genetic risk factors for childhood-onset and adult-onset asthma: genome-wide and transcriptome-wide studies. *Lancet Respir Med* 2019;7:509-22.
- Johansson A, Rask-Andersen M, Karlsson T, Ek WE. Genome-wide association analysis of 350 000 Caucasians from the UK Biobank identifies novel loci for asthma, hay fever and eczema. *Hum Mol Genet* 2019;28:4022-41.
- Myers RA, Scott NM, Gauderman WJ, Qiu W, Mathias RA, Romieu I, et al. Genome-wide interaction studies reveal sex-specific asthma risk alleles. *Hum Mol Genet* 2014;23:5251-9.
- Landgraf-Rauf K, Boeck A, Siemens D, Klucker E, Vogelsang V, Schmidt S, et al. IRF-1 SNPs influence the risk for childhood allergic asthma: a critical role for pro-inflammatory immune regulation. *Pediatr Allergy Immunol* 2018;29:34-41.

38. Bhavsar P, Ahmad T, Adcock IM. The role of histone deacetylases in asthma and allergic diseases. *J Allergy Clin Immunol* 2008;121:580-4.
39. Wawrzyniak P, Wawrzyniak M, Wanke K, Sokolowska M, Bendelja K, Ruckert B, et al. Regulation of bronchial epithelial barrier integrity by type 2 cytokines and histone deacetylases in asthmatic patients. *J Allergy Clin Immunol* 2017;139:93-103.
40. Hollams EM, Teo SM, Kusel M, Holt BJ, Holt KE, Inouye M, et al. Vitamin D over the first decade and susceptibility to childhood allergy and asthma. *J Allergy Clin Immunol* 2017;139:472-81.e9.
41. Wolsk HM, Chawes BL, Litonjua AA, Hollis BW, Waage J, Stokholm J, et al. Prenatal vitamin D supplementation reduces risk of asthma/recurrent wheeze in early childhood: a combined analysis of two randomized controlled trials. *PLoS One* 2017;12:e0186657.
42. Pacis A, Tailleux L, Morin AM, Lambourne J, MacIsaac JL, Yotova V, et al. Bacterial infection remodels the DNA methylation landscape of human dendritic cells. *Genome Res* 2015;25:1801-11.
43. Nicodemus-Johnson J, Naughton KA, Sudi J, Hogarth K, Naurekas ET, Nicolae DL, et al. Genome-wide methylation study identifies an IL-13-induced epigenetic signature in asthmatic airways. *Am J Respir Crit Care Med* 2016;193:376-85.
44. Soliai MM, Kato A, Helling BA, Stanhope CT, Norton JE, Naughton KA, et al. Multi-omics colocalization with genome-wide association studies reveals a context-specific genetic mechanism at a childhood onset asthma risk locus. *Genome Med* 2021;13:157.
45. Cardenas A, Fadadu RP, Van Der Laan L, Ward-Caviness C, Granger L, Diaz-Sanchez D, et al. Controlled human exposures to diesel exhaust: a human epigenome-wide experiment of target bronchial epithelial cells. *Environ Epigenet* 2021;7:dvab003.
46. Wan ES, Qiu W, Baccarelli A, Carey VJ, Bacherman H, Rennard SI, et al. Cigarette smoking behaviors and time since quitting are associated with differential DNA methylation across the human genome. *Hum Mol Genet* 2012;21:3073-82.
47. Bakulski KM, Dou J, Lin N, London SJ, Colacino JA. DNA methylation signature of smoking in lung cancer is enriched for exposure signatures in newborn and adult blood. *Sci Rep* 2019;9:4576.
48. Sikdar S, Joehanes R, Joubert BR, Xu CJ, Vives-Usano M, Rezwani FI, et al. Comparison of smoking-related DNA methylation between newborns from prenatal exposure and adults from personal smoking. *Epigenomics* 2019;11:1487-500.
49. Busche S, Shao X, Caron M, Kwan T, Allum F, Cheung WA, et al. Population whole-genome bisulfite sequencing across two tissues highlights the environment as the principal source of human methylome variation. *Genome Biol* 2015;16:290.
50. Modena BD, Tedrow JR, Milosevic J, Bleecker ER, Meyers DA, Wu W, et al. Gene expression in relation to exhaled nitric oxide identifies novel asthma phenotypes with unique biomolecular pathways. *Am J Respir Crit Care Med* 2014;190:1363-72.
51. Yip KH, Kolesnikoff N, Hauschild N, Biggs L, Lopez AF, Galli SJ, et al. The Nedd4-2/Ndfip1 axis is a negative regulator of IgE-mediated mast cell activation. *Nat Commun* 2016;7:13198.
52. Zhang Y, Huang Y, Chen WX, Xu ZM. Identification of key genes in allergic rhinitis by bioinformatics analysis. *J Int Med Res* 2021;49:3000605211029521.
53. Schlag K, Steinhilber D, Karas M, Sorg BL. Analysis of proximal ALOX5 promoter binding proteins by quantitative proteomics. *FEBS J* 2020;287:4481-99.