

Blood DNA methylation biomarkers predict clinical reactivity in food-sensitized infants

David Martino, PhD,^{a,b,d,f} Thanh Dang, PhD,^{a,b} Alexandra Sexton-Oates, BSc,^b Susan Prescott, MBBS, FRACP, PhD,^{a,c,f} Mimi L. K. Tang, MBBS, PhD, FRACP, FRCPA, FAAAAI,^{a,b,d,e} Shyamali Dharmage, MBBS, MSc, MD, PhD,^{a,g} Lyle Gurrin, PhD,^{a,g} Jennifer Koplin, PhD,^{a,g} Anne-Louise Ponsonby, MD, PhD,^{a,b,d} Katrina J. Allen, MBBS, FRACP, FAAAAI, PhD,^{a,b,d,f,h,*} and Richard Saffery, PhD,^{b,d,*} for the HealthNuts study investigators
Parkville, Melbourne, and Perth, Australia, and Manchester, United Kingdom

Background: The diagnosis of food allergy (FA) can be challenging because approximately half of food-sensitized patients are asymptomatic. Current diagnostic tests are excellent makers of sensitization but poor predictors of clinical reactivity. Thus oral food challenges (OFCs) are required to determine a patient's risk of reactivity.

Objective: We sought to discover genomic biomarkers of clinical FA with utility for predicting food challenge outcomes. **Methods:** Genome-wide DNA methylation (DNAm) profiling was performed on blood mononuclear cells from volunteers who had undergone objective OFCs, concurrent skin prick tests, and specific IgE tests. Fifty-eight food-sensitized patients (aged 11-15 months) were assessed, half of whom were clinically reactive. Thirteen nonallergic control subjects were also assessed.

Reproducibility was assessed in an additional 48 samples by using methylation data from an independent population of patients with clinical FA.

Results: Using a supervised learning approach, we discovered a DNAm signature of 96 CpG sites that predict clinical outcomes.

Diagnostic scores were derived from these 96 methylation sites, and cutoffs were determined in a sensitivity analysis. Methylation biomarkers outperformed allergen-specific IgE and skin prick tests for predicting OFC outcomes. FA status was correctly predicted in the replication cohort with an accuracy of 79.2%. **Conclusion:** DNAm biomarkers with clinical utility for predicting food challenge outcomes are readily detectable in blood. The development of this technology in detailed follow-up studies will yield highly innovative diagnostic assays. (J Allergy Clin Immunol 2015;■■■■:■■■■-■■■■.)

Key words: Food allergy, epigenetics, DNA methylation profiling, Infinium, methylation profiling, epigenetic epidemiology, allergy epigenetics, biomarkers, shrunk centroids

Food allergy (FA) affects up to 10% of children in the most affected countries¹ and has become a substantial public health concern. Increasing rates of FA among young children have coincided with an increase in potentially life-threatening anaphylaxis.² Laboratory testing for detection of allergen-specific IgE (sIgE) is widely used in the diagnosis of IgE-mediated FA, which offers proof of sensitization; however, further evidence of clinical manifestations on exposure is required to make a definitive diagnosis.³ This is because the majority of children with a positive skin prick test (SPT) response or sIgE test result are not allergic to that food.⁴

At present, there are no reliable biomarkers for clinical FA, and the oral food challenge (OFC) test is still required to clarify clinical reactivity. OFCs are the gold standard method to diagnose FA, but can be dangerous primarily because of the risk of inducing anaphylaxis in an allergic subject. As a result of this, OFCs are often not performed, potentially leading to overdiagnosis of FA or unnecessary avoidance programs.⁴ There is a need to improve current allergy testing practices that can help make a more specific diagnosis regarding a patient's risk of reactivity on exposure. Identifying biomarkers of clinical allergy could prove extremely useful in the absence of a clinical history or if the child has yet to be exposed to a particular food or has undergone a lengthy avoidance program.

Epigenetic biomarkers of disease risk or exposure have already generated broad interest because these factors can mediate genetic or environmental risk.⁵ In diseases underpinned by complex gene-environment interactions, epigenetics is centrally placed as a key mediator.⁶ In a previous study we reported DNA methylation (DNAm) differences associated with development of IgE-mediated FA in early life.⁷ Extending this research, in the present study we generate genome-wide DNAm profiles from volunteers recruited to the HealthNuts study undergoing

From ^athe Centre for Food and Allergy Research, Murdoch Childrens Research Institute, Parkville; ^bMurdoch Childrens Research Institute, Royal Children's Hospital, Parkville; ^cthe School of Paediatrics and Child Health, University of Western Australia, Perth; ^dthe University of Melbourne Department of Paediatrics; ^ethe Department of Allergy and Immunology, Royal Children's Hospital, Parkville; ^fIn-FLAME, the International Inflammation Network, World Universities Network (WUN); ^gthe School of Population and Global Health, University of Melbourne; and ^hthe Institute of Inflammation and Repair, University of Manchester.

*These authors contributed equally to this work.

Supported by the CASS Foundation (SM/11/3647/RMR:DSA), the DHB Foundation (CT21242), and the NHMRC (1072752).

Disclosure of potential conflict of interest: This study was funded by the CASS Foundation, DHB Foundation, and the National Health and Medical Research Council, Australia's leading expert body promoting the development and maintenance of public and individual health standards. S. Prescott has received compensation for board membership from Danone Nutricia and the Nestlé Nutrition Institute, has received payment for delivering lectures from Danone Nutricia and the Nestlé Nutrition Institute, and receives royalties from UpToDate. M. L. K. Tang has received compensation for board membership from Danone Nutricia and the Nestlé Nutrition Institute; has received payment for delivering lectures from Danone Nutricia and the Nestlé Nutrition Institute; and has received compensation for travel and other meeting-related expenses from APARPI. A.-L. Ponsonby is employed by the NHMRC, from which she has also received or has grants pending. K. J. Allen has received consultancy fees from Nutricia, Danone, Nestlé, Aspen, and Alphapharm. The rest of the authors declare that they have no relevant conflicts of interest.

Received for publication September 1, 2014; revised November 28, 2014; accepted for publication December 29, 2014.

Corresponding author: Katrina J. Allen, MBBS, FRACP, FAAAAI, PhD, Gastro and Food Allergy, Murdoch Childrens Research Institute, Flemington Rd, Parkville, Melbourne, Australia. E-mail: Katie.Allen@rch.org.au.

0091-6749/\$36.00

© 2015 American Academy of Allergy, Asthma & Immunology

<http://dx.doi.org/10.1016/j.jaci.2014.12.1933>

Abbreviations used

DNAm: DNA methylation
 FA: Food allergy
 FDR: False discovery rate
 FS: Food sensitization
 MAP: Mitogen-activated protein
 OFC: Oral food challenge
 ROC: Receiver operating characteristic
 sIgE: Allergen-specific IgE
 SPT: Skin prick test
 Treg: Regulatory T

laboratory tests for food sensitization (FS) and objective OFCs. We examined the medical utility of DNAm to yield clinically relevant biomarkers of OFC outcomes for different FA phenotypes. We chose to study egg allergy as the most common form of FA in early childhood¹ with the highest rate of resolution,⁸ as well as peanut allergy, which is less likely to resolve⁹ and more commonly associated with both anaphylaxis and mortality.¹⁰ Using machine-learning approaches that have proved success in biomarker translation in the cancer space, we report CpG sites at which DNAm levels vary according to clinical FA status. Development of this technology in detailed follow-up studies could yield highly innovative diagnostic testing.

METHODS**Sample population**

We obtained blood samples from the HealthNuts study in Melbourne, Australia, a population-based study of infant FA.¹¹ These samples were collected from 11- to 15-month-old infants who underwent SPTs to egg white, peanut, and 1 of 2 other foods (cow's milk or shrimp) and had undergone objective OFCs yielding definitive phenotyping data.¹² Although egg allergy is the most common form of FA in early childhood,¹ it has the highest rate of resolution,⁸ whereas peanut allergy is unlikely to resolve⁹ and is more commonly associated with both anaphylaxis and mortality.¹⁰ All infants in the HealthNuts study underwent SPTs to egg white, peanut, sesame, and 1 of 2 other foods (cow's milk or shrimp). Those with detectable SPT-induced wheal reactions (wheal size 1 mm or larger than that elicited by the negative control) were invited to the HealthNuts research clinic at Melbourne's Royal Children's Hospital within the next 4 to 8 weeks for a repeat SPT and a formal OFC, which were undertaken irrespective of SPT-induced wheal size and used predetermined challenge criteria.¹² Blood was collected into a sodium heparin tube (Sarstedt, Newton, NC) 1 to 2 hours after the last OFC dose (see below). Ethical approval was obtained from the Office for Children HREC (reference no. CDF/07/492), the Department of Human Services HREC (reference no. 10/07), and Royal Children's Hospital HREC (reference no. 27047).

Predefined criteria for a positive OFC result

The criterion for a positive food challenge result was at least 1 of the following signs present during the OFC: 3 or more concurrent noncontact hives (urticarial lesion) lasting for more than 5 minutes, perioral or periorbital angioedema, vomiting (excluding immediate postingestion gagging/vomiting) or evidence of anaphylaxis, as defined by the Australian Society of Clinical Allergy and Immunology (evidence of circulatory or respiratory compromise) within 2 hours of the last OFC dose.¹²

Purification of mononuclear cells and nucleic extraction

PBMCs were isolated within 2 hours of collection by means of density gradient centrifugation and cryopreserved in RPMI with 15% dimethyl

sulfoxide in FCS. DNA was collected from PBMCs by using Qiagen AllPrep Kits (Qiagen, Doncaster, Victoria, Australia), according to the manufacturer's instructions. The purity and concentration of DNA were assessed by using spectrophotometry. Flow cytometry was performed in parallel to estimate cell counts.

Flow cytometry

Quantitation of total CD4⁺ T-cell numbers and total CD4⁺ regulatory T (Treg) cell numbers was performed on all blood samples as a gold standard to assess potential heterogeneity in patients' blood cell components (see Fig E2 in this article's [Online Repository](http://www.jacionline.org) at www.jacionline.org). Cells were stained with fluorochrome-conjugated monoclonal or isotype control antibodies in 100- μ L staining volumes for 30 minutes at room temperature. For intracellular staining, cells were subsequently permeabilized, fixed, and stained with forkhead box protein 3–phycoerythrin antibody or isotype control, according to the manufacturer's instructions (BD Biosciences, San Jose, Calif). All flow cytometric data were acquired on a 10-color LSR II (BD Biosciences) and analyzed with FACSDiva version 8.2 software by using well-defined gating strategies. Compensation experiments were performed with positive and negative control beads (BD Biosciences). The same compensation settings were used for each flow cytometric analysis. Data were captured from 2×10^5 or more cells to obtain 75,000 or more viable CD4⁺ lymphocytes. The Treg cell population was characterized as CD4⁺CD25⁺ forkhead box protein 3–positive cells.

DNAm analysis

To quantitate DNAm levels, 500 ng of genomic DNA derived from patients' PBMCs was bisulfite treated with the Human Genomic Signatures MethylEasy Xceed kit (Genetic Signatures, North Ryde, New South Wales, Australia), according to the manufacturer's instructions. Successful conversion was verified in all samples by using an in-house bisulfite-specific PCR assay, as previously described.¹³ Bisulfite-treated DNA was submitted to the Australian Genome Research Facility (Parkville, Melbourne, Australia) as a single randomized batch for hybridization to Illumina Infinium HumanMethylation450 arrays (Illumina, San Diego, Calif).

Preprocessing of microarray data

Raw .iDAT files were preprocessed by using the Minfi package¹⁴ from the Bioconductor Project (<http://www.bioconductor.org>) in the R statistical environment (<http://cran.r-project.org/>, version 3.0.2). Quality assessment of control probes on the array indicated high-quality data with excellent performance of control probes in all samples. The Minfi package was used for array preprocessing with the stratified quantile normalization method. Technical bias attributable to different probe chemistries between type I and type II probes was adjusted in this procedure (see Fig E1 in this article's [Online Repository](http://www.jacionline.org) at www.jacionline.org). Cell counts were estimated empirically from DNAm data by using the EstimateCellComposition function in the Minfi package and compared with flow cytometric data collected in parallel (see Fig E2 in this article's [Online Repository](http://www.jacionline.org) at www.jacionline.org). Sex calls were obtained from probe intensities on the sex chromosomes and checked against the recorded sexes to rule out any sample mix ups during processing. Probes on the X and Y chromosomes were then removed to eliminate sex bias, as were poorly performing probes with a signal detection *P* value call of greater than .01 in 1 or more samples. Probes previously demonstrated to potentially cross-hybridize nonspecifically in the genome were also removed.¹⁵ Probes containing a single nucleotide polymorphism at the single-base extension site with a minor allele frequency of less than 0.05 were removed. Methylation percentages were estimated as β values as follows:

$$\beta = \text{Methylated} / (\text{Unmethylated} + \text{Methylated} * 100).$$

They were then used to develop the classifier model. This data set has been submitted to the Gene Expression Omnibus and is available under accession number GSE59999.

Feature selection

We used the shrunken centroids classifier method of Tibshirani et al¹⁶ to develop a classifier model based on patients' methylation profiles. The method computes a standardized centroid equivalent to a reference profile for each group (FA and FS). This is the average DNAm for each CpG in each class divided by the within-class SD for that CpG. Classification takes place by comparing the methylation profile of a new sample with each of these reference class centroids. The reference centroid closest in squared distance to an individual patient sample is the predicted class for the new sample. Posterior probabilities were estimated for each prediction call to estimate confidence in the prediction. The classifier model was built on a training set consisting of 80% of the sample population, and performance was validated in the reserved 20% that was previously unseen by the model. Ten-fold cross-validations were performed on the training samples and classification errors were estimated for a range of CpGs included in the model to determine the optimal number of CpGs in the prediction model. The minimum number of CpGs that produce no misclassification errors was determined empirically from the training data. The false discovery rate (FDR) was estimated by using the *q* value method of Storey and Tibshirani.¹⁷ The model was validated on the unseen 20% of samples and performance was recorded. Gene sets enrichment analysis was performed by using the method of Subramanian et al¹⁸ through the Broad Institute molecular signatures database (<http://www.broadinstitute.org/gsea>) with an FDR *q* value of 0.05.

Replication study

Reproducibility was assessed in a publicly available data set previously generated for the Illumina Human Methylation 450k array derived from total CD4⁺ T cells (GSE34639). This study includes genome-wide DNAm from 12 patients with FA and 12 age-matched nonsensitized control subjects.⁷ In this study a pediatric allergist had determined allergic status based on clinical outcomes at the 12-month physical assessment, case history, and allergy testing. The FA group consisted of patients with FA and IgE mostly to hen's egg (*n* = 11/12), cow's milk (*n* = 1/12), or peanut (*n* = 2/12). FA was defined by clear immediate symptoms (1-2 hours) on exposure to egg, milk, or peanut (including anaphylaxis, angioedema, or urticaria) and IgE-mediated sensitivity confirmed by a positive SPT response of 3 mm or greater than that elicited by the negative control at 12 months of age. Matched DNAm data derived from total CD4⁺ T cells were available at 2 time points, birth and 12-months, and more detailed experimental procedures can be found in the study by Martino et al.⁷

Serum IgE measures

Whole-peanut and egg white IgE levels were measured in the primary cohort by using the ImmunoCAP System FEIA (Phadia AB, Uppsala, Sweden).

RESULTS

Clinical features of the sample population

For this study, we selected subjects monosensitized to either egg or peanut to represent both the common and severe forms of FA, respectively. Genome-scale methylation analysis was performed on PBMCs from 58 food-sensitized subjects (aged 11-15 months), half of whom were clinically reactive based on OFC results. Thirteen nonsensitized nonallergic control subjects were also assessed. We divided patients into major subgroups according to their food challenge outcome, as shown in Table I.

Nonallergic control subjects had a negative SPT response with a negative OFC outcome. Patients with FS has a positive SPT response (either egg or peanut) and a negative OFC outcome (either egg or peanut). Patients with FA had a positive SPT response (either egg or peanut) and an unequivocal allergic reaction during OFCs (either egg or peanut). By using these

TABLE I. Sample groupings for HealthNuts study blood samples

Major group	Subgroup	Clinical outcome
FA	Egg (<i>n</i> = 15)	SPT >2 mm to egg, unequivocal allergic reaction during OFC
	Peanut (<i>n</i> = 14)	SPT >2 mm to peanut, unequivocal allergic reaction during OFC
FS	Egg (<i>n</i> = 14)	SPT >2 mm to egg, no allergic reaction during OFC
	Peanut (<i>n</i> = 15)	SPT >2 mm to peanut, no allergic reaction during OFC
NA	— (<i>n</i> = 13)	Negative SPT response to peanut, egg, sesame, and cow's milk; no allergic reaction during OFC to these allergens

definitions, the study cohort consisted of 71 samples in total, comprising subgroups of patients with egg sensitization (*n* = 14), egg allergy (*n* = 15), peanut sensitization (*n* = 15), or peanut allergy (*n* = 14) or nonatopic subjects (*n* = 13). With this sample size, we had approximately 80% power to detect between-group differences in methylation of 4% or greater, assuming 99% of the genome is not differentially methylated. Parallel-flow cytometric cell counts were obtained from *ex vivo* blood samples to assess for confounding effects of blood cell composition on methylation profiles (see Fig E2). The DNAm assay was carried out with the Illumina Human Methylation 450k array, as previously described by us.^{7,13}

Development of a methylation signature predictive of challenge outcomes

After data preprocessing and quality control, we first identified probes most variably methylated between our subgroups of interest (egg allergy, egg sensitization, peanut allergy, and peanut sensitization) using ANOVA, and then from these probes, we built a classifier model trained to predict food challenge outcomes. Nonallergic subjects were excluded from building the classifier model because these subjects would not be seen in the clinic and therefore were not of clinical relevance; however, they were reintroduced for the subsequent sensitivity analysis. The approach is outlined in Fig 1. In total, 4485 probes varied significantly between clinical subgroups of interest (egg allergy, egg sensitization, peanut allergy, and peanut sensitization; *P* < .01, ANOVA). We used multidimensional scaling analysis to cluster the 58 samples according to the 4485 variable probes. All samples clustered according to their FA phenotype, suggesting these probes correlate with the clinical phenotypes of interest (Fig 1, A). We then trained a nearest shrunken centroid classifier to predict the food challenge outcome group (FA or FS) to which each patient belonged. The 58 samples were partitioned into a training set consisting of 80% of the sample population for model building (*n* = 22 for the FA group and *n* = 24 for the FS group) and a validation set consisting of 7 patients with FA and 5 patients with FS. Further independent validation was assessed subsequently in a separate population. Using a 10-fold cross-validation on the training data set, we developed a classifier model based on 649 CpG sites, a threshold at which cross-validation errors were minimized at 2.2% in the training data set (Fig 1, B). The median FDR (*q* value) was calculated

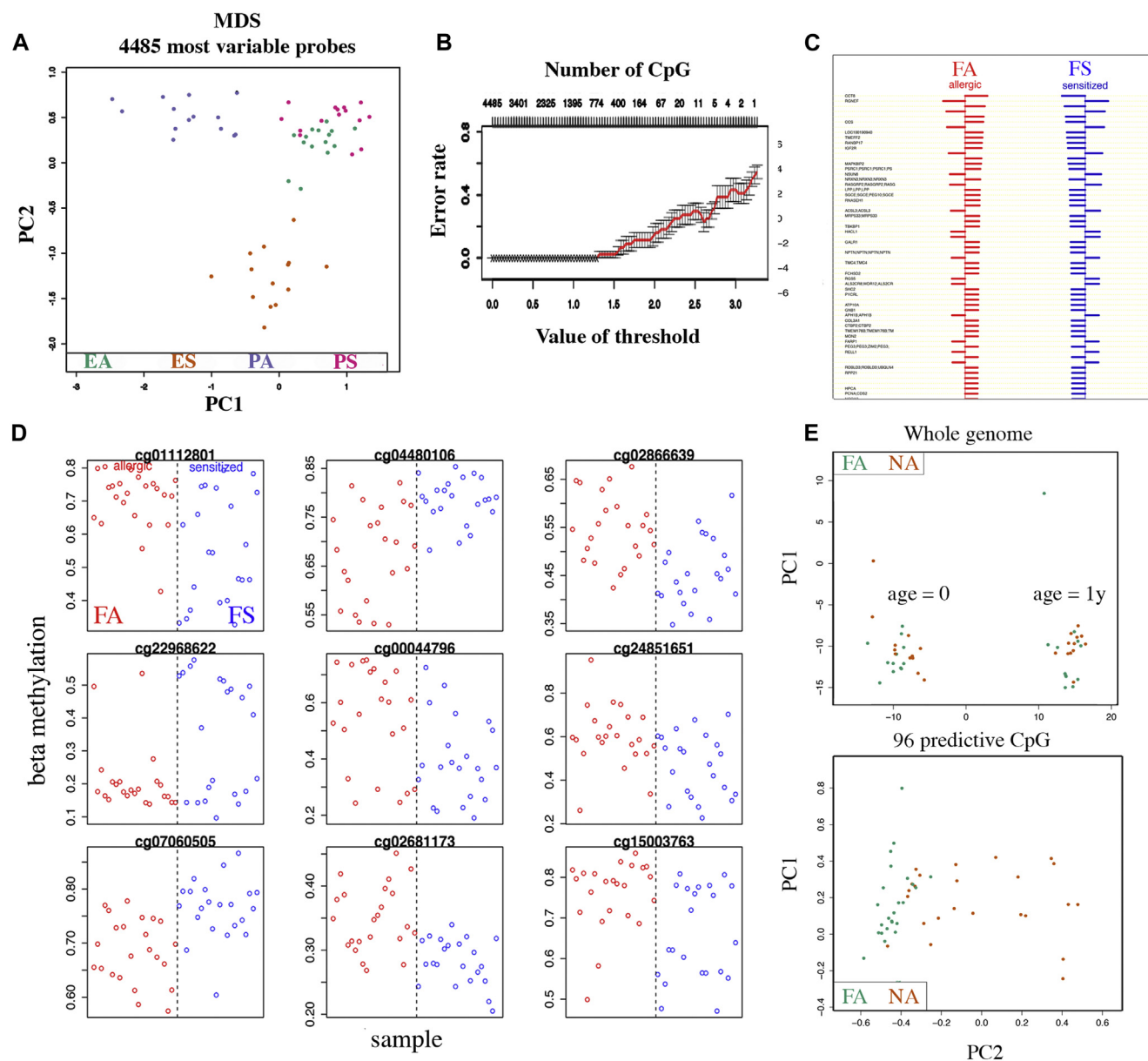


FIG 1. Development of a predictive model based on blood-derived DNAm levels. **A**, Multidimensional scaling analysis of sample relationships based on 4485 probes significantly associated ($P < .01$, ANOVA) with phenotype. EA, Egg allergic; ES, egg sensitized; PA, peanut allergic; PS, peanut sensitized. **B**, Error curves from 10-fold cross-validation experiments. The model was trained to predict food challenge outcome. Misclassification errors are shown (y-axis) for each threshold at which noisy CpGs were removed. **C**, Class centroids are represented for each food challenge outcome group. Approximately 50 centroids are shown for ease of visualization. **D**, Scatter plots of β DNAm values from the top 6 ranked CpG sites based on average cross-validation scores. The x-axis shows samples (red, FA; blue, FS), and the y-axis shows the β methylation value. **E**, Multidimensional scaling analysis of validation study. The top panel shows sample relationships based on all somatic probes, and the bottom panel shows sample relationships based on the 96 predictive CpG signatures. Samples are numbered by age (0, birth; 1, 12 months) and colored by phenotype (FA, green; NA, orange). PC, Principal component of variation.

by using 100 permutations and controlled at 0.05. This analysis identified a signature of 649 CpG sites with apparent predictive utility for challenge outcome.

Biological relevance of the methylation signature

We annotated these 649 predictive CpG sites according to genomic location to identify overlapping protein-coding genes

and performed an enrichment test for the gene list. The list was substantially enriched for immunologic genes derived from numerous cell types, including genes differentially expressed between plasma cells and memory B lymphocytes (*VEGFA*, *TG*, *TIMP2*, *PLA2G16*, and *VKORC1*; FDR q value = 7.09×10^{-4}), effector memory and central memory cells (*DTX1*, *MAP3K1*, *SLC24A6*, and *PKMYT1*; FDR q value = 8.70×10^{-4}), Treg cells and conventional T cells (*CAMKK1*, *TRPM1*, *SCL45A2*, and

TABLE II. Gene set enrichment test

Gene set name	No. of genes in set (K)	Description	No. of genes in overlap (k)	k/K	FDR q value
GSE13411_PLASMA_CELL_VS_MEMORY_BCELL_UP	200	Genes upregulated in comparison of plasma cells vs memory B cells	12	0.06	7.09E-04
GSE13306_TREG_VS_TCONV_UP	200	Genes upregulated in Treg cells vs conventional T cells	11	0.06	8.07E-04
GSE15659_CD45RA_NEG_CD4_TCELL_VS_ACTIVATED_TREG_UP	200	Genes upregulated in comparison of PTPRC [−] (Gene ID = 5788) CD4 (Gene ID = 920) T cells vs activated Treg cells	11	0.06	8.07E-04
GSE11057_EFF_MEM_VS_CENT_MEM_CD4_TCELL_DN	199	Genes downregulated in comparison of effector memory T cells vs central memory T cells from PBMCs	11	0.06	8.07E-04
GSE17974_IL4_AND_ANTI_IL12_VS_UNTREATED_6H_ACT_CD4_TCELL_DN	200	Genes downregulated in comparison of CD4 (Gene ID = 920) T cells treated with IL-4 (Gene ID = 3565) and anti-IL-12 at 6 h vs the untreated cells at 6 h	11	0.06	8.07E-04
GSE22886_NAIVE_TCELL_VS_NEUTROPHIL_UP	200	Genes upregulated in comparison of naive CD4 (Gene ID = 920) CD8 T cells vs unstimulated neutrophils	12	0.06	7.09E-04
GSE22886_UNSTIM_VS_IL2_STIM_NKCELL_DN	200	Genes downregulated in comparison of unstimulated NK cells versus those stimulated with IL-2 (Gene ID = 3558) at 16 h	12	0.06	7.09E-04
GSE13229_IMM_VS_MATURE_NKCELL_UP	200	Genes upregulated in comparison of immature vs mature NK cells	11	0.06	8.07E-04
GSE17721_12H_VS_24H_PAM3CSK4_BMDM_UP	200	Genes upregulated in comparison of dendritic cells stimulated with Pam3Csk4 (TLR1/2 agonist) at 12 h vs those stimulated with Pam3Csk4 (TLR1/2 agonist) at 24 h	11	0.06	8.07E-04
GSE17721_CTRL_VS_LPS_24H_BMDM_DN	200	Genes downregulated in comparison of control dendritic cells at 24 h vs those stimulated with LPS (TLR4 agonist) at 24 h	11	0.06	8.07E-04

Ten enriched gene sets are shown for convenience.

NK, Natural killer; PTPRC, protein tyrosine phosphatase receptor C; TLR, Toll-like receptor.

ARF4; FDR q-value = 8.70E-04), and genes induced by dendritic cells on stimulation (Table II; a full list is provided in Table E1 in this article's Online Repository at www.jacionline.org). The enrichment test was consistent with a mixed cell-type epigenetic signature, as expected from data derived from whole PBMCs.

We next ranked the 649 probes according to their absolute between-group difference for our main outcomes of interest (FA vs FS) and applied a cutoff for a minimum effect size of a 5% difference in methylation between groups, a cutoff for which our sample size has reliable detection power and also within the measurable range of EpiTYPER Mass Spectrometry, which is often used to measure DNAm.¹⁹ This reduced the list of 649 probes to 96 probes, which overlapped with 73 protein-coding genes (Table III). This gene list was significantly enriched for the mitogen-activated protein (MAP) kinase canonical pathway (q value = 1.34E-05), including the family members *MAPK8IP1*, *MAPK8IP2*, *MAP3K1*, *RPS6KA2*, *RASGRP2*, *NF1*, *ZAK*, and *FGF12*. We then tested this refined 96-CpG model to predict food challenge outcome in the validation samples. Predictions were made for each patient, with high posterior probabilities in each case (>0.9), and food challenge outcome was correctly predicted for each patient with no misclassification errors (see Table E2 in this article's Online Repository at www.jacionline.org). This analysis suggests optimal performance using just 96 CpGs with the largest between-group effect sizes.

Performance characteristics in a diagnostic setting

To explore the diagnostic utility of the 96-CpG DNAm signature, we examined the distribution of measurements from the 96 CpGs. Compared with the genome-wide β distribution, distribution of the predictive CpGs was restricted to the intermediate β methylation range (20% to 80% methylated), which has been shown to be more environmentally sensitive compared with the extremes (Fig 2, A).^{20,21} In addition, the distribution of patients with FA relative to those with FS was skewed toward higher methylation in this Australian cohort ($P = 7.322\text{E-}06$, Wilcoxon rank sum test), indicating general hypermethylation at these CpGs in clinically allergic patients. We computed patients' methylation scores by summing the β methylation ratios for all 96 CpGs for each subject in the study. The transition to a clinically reactive phenotype was associated with a significantly higher ($P < .001$, Mann-Whitney test) methylation score (Fig 2, B). We performed an area under the curve analysis to determine diagnostic cutoffs for these scores. In this analysis we included the nonallergic control group to examine the specific utility of the methylation score in a clinical setting (to predict FA vs FS) versus a population screening setting (nonallergic control subjects vs patients with FS/FA). Receiver operating characteristic (ROC) curve analysis of patients (FA group) versus control subjects (FS group) produced an area under the curve of 0.9774 ($P < .0001$; 95% CI, 0.9482-1.007), providing excellent

TABLE III. Gene annotations for 96 CpG signatures predictive of clinical allergy

Entrez Gene ID	Gene symbol	Gene description
2181	<i>ACSL3</i>	Acyl-CoA synthetase long-chain family member 3
57143	<i>ADCK1</i>	aarF domain containing kinase 1
83464	<i>APH1B</i>	Anterior pharynx defective 1 homolog B (C. elegans)
378	<i>ARF4</i>	ADP-ribosylation factor 4
57194	<i>ATP10A</i>	ATPase, class V, type 10A
4287	<i>ATXN3</i>	Ataxin 3
9577	<i>BRE</i>	Brain and reproductive organ-expressed (TNFRSF1A modulator)
27101	<i>CACYBP</i>	Calcyclin binding protein
23261	<i>CAMTA1</i>	Calmodulin binding transcription activator 1
900	<i>CCNG1</i>	Cyclin G1
9973	<i>CCS</i>	Copper chaperone for superoxide dismutase
10694	<i>CCT8</i>	Chaperonin containing TCP1, subunit 8 (theta)
55536	<i>CDCA7L</i>	Cell division cycle associated 7-like
53841	<i>CDHR5</i>	Cadherin-related family member 5
1281	<i>COL3A1</i>	Collagen, type III, alpha 1
1488	<i>CTBP2</i>	C-terminal binding protein 2
10455	<i>ECI2</i>	Enoyl-CoA delta isomerase 2
10160	<i>FARP1</i>	FERM, RhoGEF (ARHGEF) and pleckstrin domain protein 1 (chondrocyte-derived)
2194	<i>FASN</i>	Fatty acid synthase
9873	<i>FCHSD2</i>	FCH and double SH3 domains 2
2257	<i>FGF12</i>	Fibroblast growth factor 12
51303	<i>FKBP11</i>	FK506 binding protein 11, 19 kDa
83786	<i>FRMD8</i>	FERM domain containing 8
2782	<i>GNBI</i>	Guanine nucleotide binding protein (G protein), beta polypeptide 1
9759	<i>HDAC4</i>	Histone deacetylase 4
3108	<i>HLA-DMA</i>	MHC class II, DM alpha
3119	<i>HLA-DQB1</i>	MHC class II, DQ beta 1
3208	<i>HPCA</i>	Hippocalcin
57110	<i>HRASLS</i>	HRAS-like suppressor
8870	<i>IER3</i>	Immediate early response 3
3482	<i>IGF2R</i>	Insulin-like growth factor 2 receptor
93185	<i>IGSF8</i>	Immunoglobulin superfamily, member 8
149837	<i>LOC149837</i>	Uncharacterized LOC149837
4026	<i>LPP</i>	LIM domain containing preferred translocation partner in lipoma
57497	<i>LRFN2</i>	Leucine rich repeat and fibronectin type III domain containing 2
4050	<i>LTB</i>	Lymphotoxin beta (TNF superfamily, member 3)
84441	<i>MAML2</i>	Mastermind-like 2 (Drosophila)
4214	<i>MAP3K1</i>	Mitogen-activated protein kinase kinase kinase 1
9479	<i>MAPK8IP1</i>	Mitogen-activated protein kinase 8 interacting protein 1
23542	<i>MAPK8IP2</i>	Mitogen-activated protein kinase 8 interacting protein 2
84879	<i>MFSD2A</i>	Major facilitator superfamily domain containing 2A
84789	<i>MGC2889</i>	Uncharacterized protein MGC2889
51650	<i>MRPS33</i>	Mitochondrial ribosomal protein S33
4763	<i>NF1</i>	Neurofibromin 1
59277	<i>NTN4</i>	Netrin 4
4974	<i>OMG</i>	Oligodendrocyte myelin glycoprotein
56114	<i>PCDHGA1</i>	Protocadherin gamma subfamily A, 1
8863	<i>PER3</i>	Period homolog 3 (Drosophila)

(Continued)

TABLE III. (Continued)

Entrez Gene ID	Gene symbol	Gene description
5527	<i>PPP2R5C</i>	Protein phosphatase 2, regulatory subunit B', gamma
84722	<i>PSRC1</i>	Proline/serine-rich coiled-coil 1
9771	<i>RAPGEF5</i>	Rap guanine nucleotide exchange factor (GEF) 5
10235	<i>RASGRP2</i>	RAS guanyl releasing protein 2 (calcium and DAG-regulated)
282996	<i>RBM20</i>	RNA binding motif protein 20
768211	<i>RELL1</i>	RELT-like 1
64283	<i>RGNEF</i>	190 kDa guanine nucleotide exchange factor
8490	<i>RGS5</i>	Regulator of G-protein signaling 5
246243	<i>RNASEH1</i>	Ribonuclease H1
79897	<i>RPP21</i>	Ribonuclease P/MRP 21kDa subunit
6196	<i>RPS6KA2</i>	Ribosomal protein S6 kinase, 90kDa, polypeptide 2
6262	<i>RYR2</i>	Ryanodine receptor 2 (cardiac)
9644	<i>SH3PXD2A</i>	SH3 and PX domains 2A
25759	<i>SHC2</i>	SHC (Src homology 2 domain containing) transforming protein 2
25942	<i>SIN3A</i>	SIN3 homolog A, transcription regulator (yeast)
9962	<i>SLC23A2</i>	Solute carrier family 23 (nucleobase transporters), member 2
9755	<i>TBKBP1</i>	TBK1 binding protein 1
147798	<i>TMC4</i>	Transmembrane channel-like 4
55365	<i>TMEM176A</i>	Transmembrane protein 176A
28959	<i>TMEM176B</i>	Transmembrane protein 176B
7422	<i>VEGFA</i>	Vascular endothelial growth factor A
51776	<i>ZAK</i>	Sterile alpha motif and leucine zipper containing kinase AZK
23619	<i>ZIM2</i>	Zinc finger, imprinted 2
339500	<i>ZNF678</i>	Zinc finger protein 678

Gene annotations of 96 CpGs for which the FDR q value is less than 0.05 and the effect size is greater than 5% are shown.

test accuracy. A methylation score of greater than 47.49 produced 96.55% specificity (95% CI, 82.24% to 99.91%) and 89.66% sensitivity (95% CI, 72.65% to 97.81%) and a likelihood ratio of 26 (Table IV). Performance was marginally reduced when using the DNAm signature to diagnose FA versus lack of allergy (area = 0.8408; $P = .005$; 95% CI, 0.6854% to 0.9963%), and the scores were least predictive (although still significant) when comparing patients with FS versus nonallergic control subjects (area = 0.6976; $P = .043$; 95% CI, 0.5126% to 0.8826%), demonstrating that the utility of the DNAm signature performs best when addressing the research question (ie, is most useful in a clinical setting; Fig 2, C).

To determine whether the methylation signature would constitute an improvement over existing diagnostic tests, we performed a comparative ROC curve analysis against concurrently measured sIgE levels and SPT-induced wheal sizes. Methylation scores were superior predictors of clinical allergy status than sIgE measures for both egg- and peanut-specific FA phenotypes. For patients with egg sensitization, egg white IgE levels predicted allergy with an area under the curve of 0.6333 ($P = .2310$; 95% CI, 0.4204-0.8462), whereas methylation scores among egg-sensitized patients predicted allergy with an area under the curve of 0.9857 ($P < .0001$; 95% CI, 0.9546-1.017). Methylation scores from peanut-sensitized patients predicted allergy with an ROC curve of 0.9571 ($P < .0001$; 95% CI,

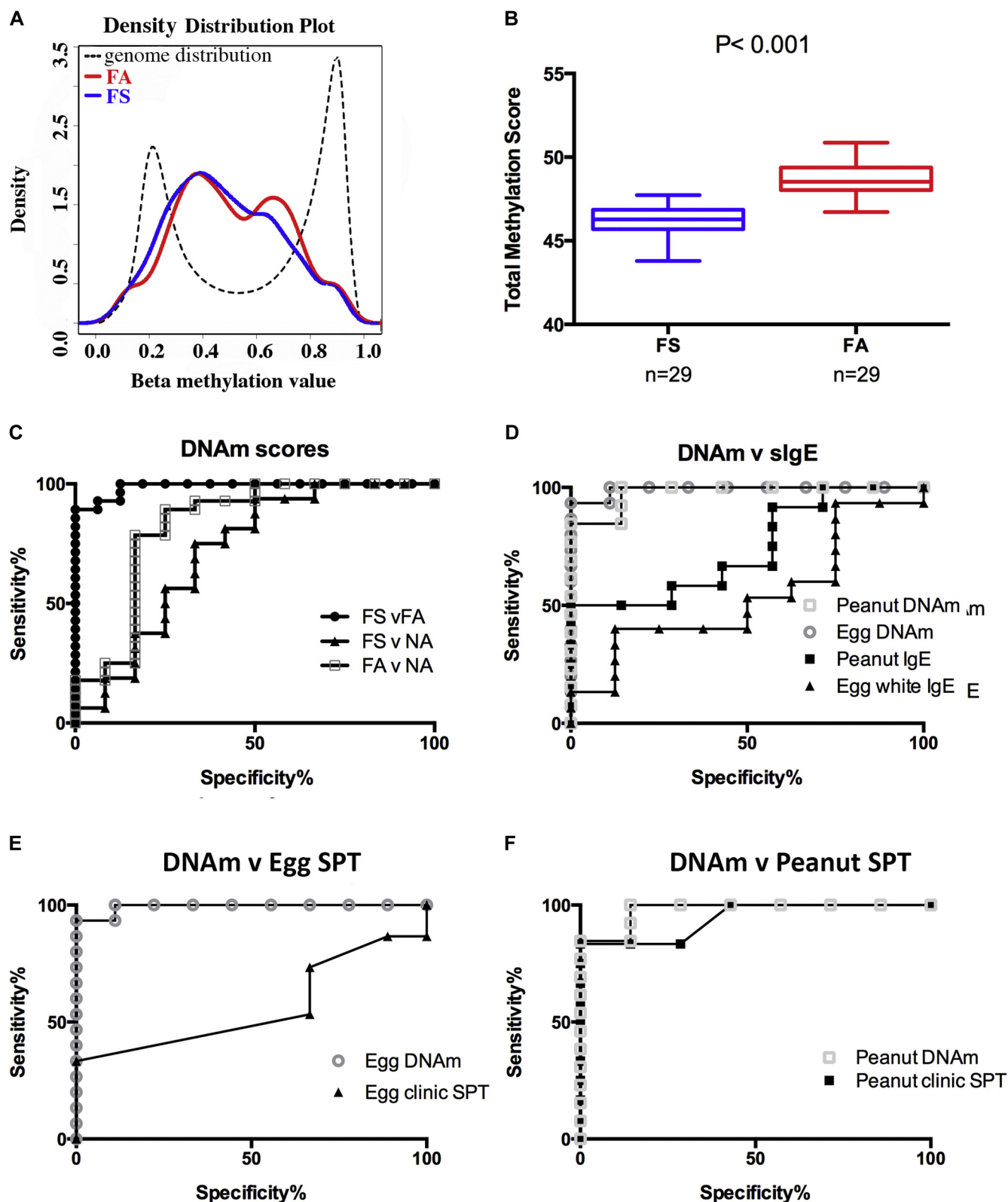


FIG 2. Sensitivity and specificity analysis of DNAm patient scores. **A**, Distribution of methylation ratios for the 96 CpG signatures stratified by phenotype. **B**, Box plots of total DNAm scores showing medians and ranges (statistical analysis using the Man-Whitney test). **C**, ROC curve analysis of DNAm scores for predicting clinical allergy between groups. **D**, Performance comparisons of DNAm scores against serum IgE measures. **E**, Performance comparisons of DNAm scores against egg wheal size among egg-sensitized patients. **F**, Performance comparisons of DNAm scores against peanut SPT wheal sizes among peanut-sensitized patients.

TABLE IV. Sensitivity and specificity analysis of methylation scores from DNAm signature

Cutoff	Sensitivity (%)	95% CI	Specificity (%)	95% CI	Likelihood ratio
>46.71	100	88.06% to 100.0%	72.41	52.76% to 87.27%	3.625
>46.86	96.55	82.24% to 99.91%	75.86	56.46% to 89.70%	4
>47.03	93.1	77.23% to 99.15%	79.31	60.28% to 92.01%	4.5
>47.12	93.1	77.23% to 99.15%	86.21	68.34% to 96.11%	6.75
>47.34	89.66	72.65% to 97.81%	89.66	72.65% to 97.81%	8.667
>47.49	89.66	72.65% to 97.81%	96.55	82.24% to 99.91%	26
>47.66	82.76	64.23% to 94.15%	96.55	82.24% to 99.91%	24
>47.75	82.76	64.23% to 94.15%	100	88.06% to 100.0%	

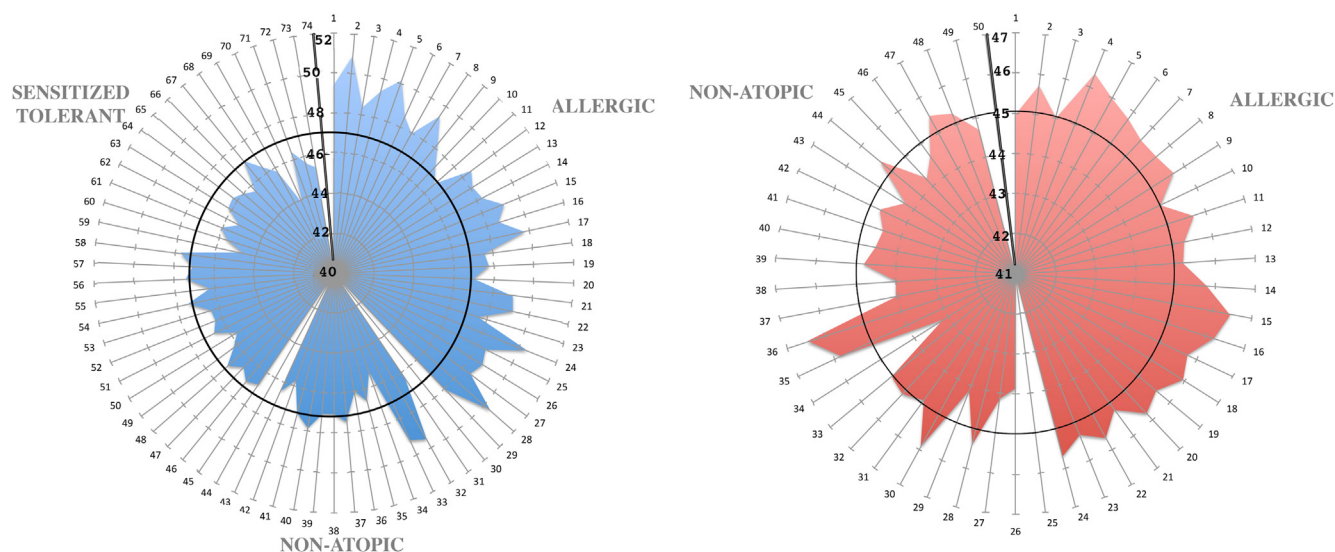


FIG 3. Starburst plot of patient methylation scores for the discovery (left) and replication (right) cohorts. Concentric numerals denote sample numbers, and patients' scores are shown on the vertical axis in boldface. Methylation scores derived from CD4⁺ T cells (red) were consistently lower than scores derived from total PBMCs (blue); however, differences between phenotype classes were conserved in each cohort. Solid circles are visual guides for diagnostic cutoffs determined by means of sensitivity analysis. Note that y-axis scales differ in the 2 plots.

0.8897-1.025) compared with a peanut IgE test ROC curve of 0.8722 ($P = .00108$; 95% CI, 0.7422-1.022; Fig 2, D).

We have also compared performance characteristics of the methylation test against SPTs for prediction of clinical allergy, and the methylation test was superior to both egg and peanut SPT-induced wheal size (Fig 2, E and F). For patients with egg sensitization, egg SPT wheal size predicted allergy with an area under the curve of 0.56 ($P = .6123$; 95% CI, 0.3308-0.7951), whereas methylation scores among egg-sensitized patients predicted allergy with an area under the curve of 0.9405 ($P < .0018$; 95% CI, 0.8386-1.042). Methylation scores from peanut-sensitized patients predicted allergy with an ROC curve of 0.9571 ($P < .0001$; 95% CI, 0.8897-1.025) compared with the peanut IgE test ROC curve of 0.9780 ($P = .0005$; 95% CI, 0.9228-1.033). These results also held true when the analysis was restricted to patients with a wheal size of 3 mm or greater (data not shown).

Assessment of reproducibility in an independent cohort

We next assessed independent reproducibility in a publicly available DNAm data set obtained from a West Australian cohort

of patients with FA and nonallergic control subjects previously generated by our laboratory.⁷ This data set comprised 48 samples in total, 12 FA and 12 nonallergic control samples, assayed at 2 time points (birth and 12 months). FA status was determined when infants were 12 months old based on prior evidence of clinical reactivity in case histories and objective SPT responses. DNAm data were collected at birth and 12 months from total CD4⁺ T cells, as published previously.⁷ We applied multidimensional scaling analysis to assess sample relationships based on our predictive methylation signature. Multidimensional scaling analysis of all probes or the top 1000 most variable probes did not discriminate the FA group from the nonallergic control group in this data set, suggesting the principal components of variation are driven by unknown effects. However, when this analysis was restricted to the 96 predictive CpG methylation signatures, the first principal component of variation between groups was due to phenotype (Fig 1, E). We generated methylation scores for each sample in this cohort using the 96 predictive CpGs and examined the distribution of scores between the West Australian and Melbourne cohorts. It was clear that methylation scores derived from total CD4⁺ T cells were consistently lower than scores derived from PBMCs; however, differences in methylation

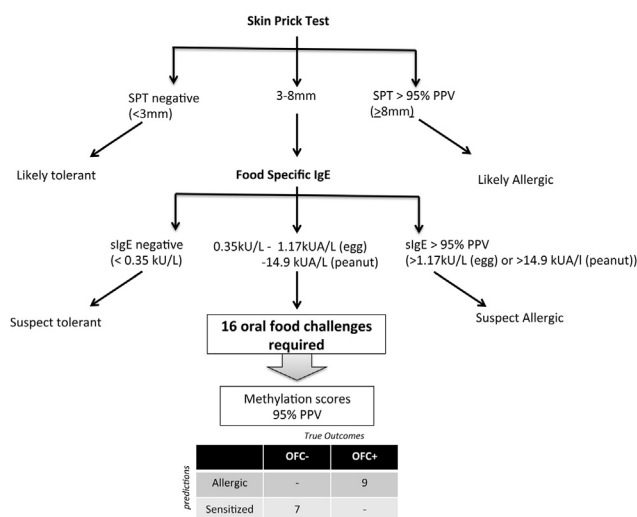


FIG 4. Diagnosing FA with a combination of IgE and methylation testing. The number of challenges required was determined by using 95% positive predictive values (PPV) for skin tests (>2 mm and <4 mm for egg and >2 mm and <8 mm for peanut) and sIgE tests (>0.35 kU_A/L and <1.17 kU_A/L for egg and >0.35 kU_A/L and <15 kU_A/L for peanut). Methylation predictions for challenge outcome were based on a cutoff of greater than 47.5 (95% negative predictive value). Error rates are shown.

scores between phenotypes were still conserved (Fig 3). This difference in methylation profiles because of cell type meant that the diagnostic cutoffs determined for total PBMCs would need to be recalibrated on total CD4⁺ T cells to perform predictions. Performing ROC curve analysis on T cell–derived methylation scores produced an area under the curve of 0.8368 ($P < .001$; 95% CI, 0.7139–0.9598). A sensitivity analysis suggested a score of greater than 45.52 resulted in a test sensitivity of 87.5% (95% CI, 67.64% to 97.34%) and a specificity of 70.83% (95% CI, 48.91% to 87.38%), with a likelihood ratio of 3.0. The difference in diagnostic cutoffs between the PBMC-derived and T cell–derived patients' scores is shown in Fig 2.

We performed an unsupervised classification of samples in the T-cell data set by assigning patients with a methylation score of greater than 45.52 to the nonallergic group and compared these assignments with the true phenotype. In total, 10 of 48 samples were misclassified, resulting in a true error rate of 20.8%. Furthermore, we show that the model can be retrained on the T-cell data and still yield accurate predictions.

We partitioned the T-cell data into a 20:80 split, reserving the bulk of samples for predictions. The model was retrained on 10 samples ($n = 5$ for the FA group and $n = 5$ for the nonallergic control group) by using all 96 predictive CpGs, and predictions were successful for 34 of the 38 remaining samples ($n = 19$ for the FA group and $n = 19$ for the nonallergic control group). Posterior probabilities were high (>0.9) for correctly classified samples but ranged from 0.4 to 0.6 for misclassified samples. Although this replication occurred in a supervised fashion, it demonstrates the flexibility of the model.

Potential diagnostic utility of combining SPTs, specific IgE tests, and DNAm tests

Finally, we explored how DNAm biomarkers might be combined with existing tests to aid clinical diagnosis. Fig 4 shows the number of OFCs in HealthNuts study subjects that would need

be undertaken to test for clinical reactivity based on a situation in which the current 95% positive predictive value thresholds for SPTs and sIgE tests were used in combination to define a threshold for such testing.^{22,23} Based on these criteria, an OFC would be recommended in 16 children to unequivocally determine their clinical reactivity. The 95% positive predictive value methylation score for these 16 children correctly predicts their challenge outcome (Fig 4), demonstrating utility in these subjects that would otherwise be difficult to classify by using a combination of SPT and sIgE results.

DISCUSSION

Biomarkers of FA are urgently needed to develop more powerful diagnostic tools that identify *bona fide* FA against a background of FS. Here we have developed a multivariate classifier model with utility for predicting food challenge outcomes. Different subtypes of FA exist, and we have shown that pattern-finding computational methods perform extremely well when prior knowledge of these subtypes are available. The strength of this approach is rooted in the definitive food challenge outcome data, which are best used with supervised classification methods²⁴ that can be used to build models to diagnose FA in future patients. By selecting monosensitized patients, we knew *a priori* which patients belong to which subgroup. This is important because it facilitated selection of clinically significant CpGs before model building. Our model was then built on the assumption that different subgroups of FA are likely to share overlapping pathogenic pathways involved in atopy and clinical allergy. Conceptually, this is not unreasonable, and we have demonstrated that common methylation pathways can be used to make accurate predictions of allergy status for different FA phenotypes. It is noteworthy that the sole canonical pathway enriched in the list of genes associated with CpGs most predictive of food challenge outcome is the MAP kinase pathway. This is the second FA cohort in which we have reported an association between clinical FA and DNAm in MAP kinase signaling genes; the former was discovered in the West Australian cohort.⁷ The strength of this association at present is unclear because different MAP kinase genes were implicated in each cohort, although collectively, the data suggest a shared pathway. The upstream regulators of this pathway are currently unknown but warrant further investigation.

The enrichment tests for the predictive CpG model built on the Melbourne cohort are consistent with a mixed cell–type signature, as would be expected from total mononuclear cells. This represents proof of principle that peripheral blood potentially offers a noninvasive, easily accessible, safe, and widely applicable diagnostic approach. The translational significance of this is further demonstrated by the diagnostic utility of composite methylation scores, which represent a substantial improvement over current diagnostic tests. Replication of the findings presented here will be important to validate the clinical effectiveness of DNAm diagnostics. A direct comparison of the approach described here with emerging component-resolved diagnostic tests, such as Ara h 2, will be important to identify an approach with the most clinical utility. This will also allow an investigation of whether a combined approach, using both DNAm and other measures, is of even greater clinical utility.

There are a growing number of multivariate diagnostic assays (“gene signatures”) marketed as diagnostic/prognostic tests.²⁵

The reproducible accuracy of patient classification demonstrated in the replication cohort was particularly encouraging. In total, 79.2% of samples were correctly diagnosed in an unsupervised manner by using the 96-CpG predictive model, despite these methylation data being derived from a different cell type, and in the absence of the “sensitized but tolerant” group, which is expected to produce the most accurate classification (ROC: FS vs FA = 0.9774 compared with ROC: FA vs nonallergic control = 0.8404). Therefore we anticipate better accuracy will be obtained when the test is applied in the intended clinical setting. Collectively, our data demonstrate reproducibility despite challenges of different cell types and population.

Although our approach is highly novel, with tremendous potential for development into a clinically relevant diagnostic test, it has several limitations that require substantial additional investigation. For example, it is not currently clear how applicable our methylation signature might be as a prognostic marker of FA resolution (acquired tolerance). Furthermore, the signature reported here is likely to require refinement after further investigation in larger sample sizes.

Other caveats include the cross-sectional analysis and the restricted scope of FA phenotypes investigated here. The data presented here pertain to age-matched, single-sensitized patients, and further testing is required in larger data sets of different aged children with multiple FSs. Furthermore, larger cohorts of single-sensitized subjects will be better suited to address any mechanistic differences between egg and peanut FA.

Biomarkers of clinical allergy that supplement existing patient management strategies will help avoid unnecessary food challenges in food-sensitized subjects. This could reduce waiting times in the immediate term and substantially alleviate the current burden on the public health system, at least in part. We believe further development of the approaches demonstrated here can lead to powerful new tools for diagnosing and treating allergy and other immune-related diseases.

We thank all the parents and children who participated in the study. Additional HealthNuts study investigators include Melanie Matheson, Melissa Wake, Anne-Adrian Lowe, David Hill, and Rachel Peters.

Clinical implications: Through validation in 2 separate study populations, this article provides proof of principle that DNAm biomarkers can be used for novel diagnostic testing in patients with FA.

REFERENCES

- Osborne NJ, Koplin JJ, Martin PE, Gurrin LC, Lowe AJ, Matheson MC, et al. Prevalence of challenge-proven IgE-mediated food allergy using population-based sampling and predetermined challenge criteria in infants. *J Allergy Clin Immunol* 2011;127:668-76, e1-2.
- Prescott S, Allen KJ. Food allergy: riding the second wave of the allergy epidemic. *Pediatr Allergy Immunol* 2011;22:155-60.
- Ito K. Diagnosis of food allergies: the impact of oral food challenge testing. *Asia Pac Allergy* 2013;3:59-69.
- Santos AF, Lack G. Food allergy and anaphylaxis in pediatrics: update 2010-2012. *Pediatr Allergy Immunol* 2012;23:698-706.
- Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. *Nat Rev Genet* 2011;12:529-41.
- Martino D, Kesper DA, Amarasekera M, Harb H, Renz H, Prescott S. Epigenetics in immune development and in allergic and autoimmune diseases. *J Reprod Immunol* 2014;104-105:43-8.
- Martino D, Joo JE, Sexton-Oates A, Dang T, Allen K, Saffery R, et al. Epigenome-wide association study reveals longitudinally stable DNA methylation differences in CD4+ T cells from children with IgE-mediated food allergy. *Epigenetics* 2014;9:998-1006.
- Peters RL, Dharmage SC, Gurrin LC, Koplin JJ, Ponsonby AL, Lowe AJ, et al. The natural history and clinical predictors of egg allergy in the first 2 years of life: a prospective, population-based cohort study. *J Allergy Clin Immunol* 2014;133:485-91.
- Ho MH, Wong WH, Heine RG, Hosking CS, Hill DJ, Allen KJ. Early clinical predictors of remission of peanut allergy in children. *J Allergy Clin Immunol* 2008;121:731-6.
- Pumphrey RS, Gowland MH. Further fatal allergic reactions to food in the United Kingdom, 1999-2006. *J Allergy Clin Immunol* 2007;119:1018-9.
- Osborne NJ, Koplin JJ, Martin PE, Gurrin LC, Thiele L, Tang ML, et al. The HealthNuts population-based study of paediatric food allergy: validity, safety and acceptability. *Clin Exp Allergy* 2010;40:1516-22.
- Koplin JJ, Tang ML, Martin PE, Osborne NJ, Lowe AJ, Ponsonby AL, et al. Predetermined challenge eligibility and cessation criteria for oral food challenges in the HealthNuts population-based study of infants. *J Allergy Clin Immunol* 2012;129:1145-7.
- Martino D, Loke YJ, Gordon L, Ollikainen M, Cruickshank MN, Saffery R, et al. Longitudinal, genome-scale analysis of DNA methylation in twins from birth to 18 months of age reveals rapid epigenetic change in early life and pair-specific effects of discordance. *Genome Biol* 2013;14:R42.
- 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.
- 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.
- Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci U S A* 2002;99:6567-72.
- Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* 2003;100:9440-5.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545-50.
- Coolen MW, Statham AL, Gardiner-Garden M, Clark SJ. Genomic profiling of CpG methylation and allelic specificity using quantitative high-throughput mass spectrometry: critical evaluation and improvements. *Nucleic Acids Res* 2007;35:e119.
- Novakovic B, Yuen RK, Gordon L, Penaherrera MS, Sharkey A, Moffett A, et al. Evidence for widespread changes in promoter methylation profile in human placenta in response to increasing gestational age and environmental/stochastic factors. *BMC Genomics* 2011;12:529.
- Oh JE, Chambwe N, Klein S, Gal J, Andrews S, Gleason G, et al. Differential gene body methylation and reduced expression of cell adhesion and neurotransmitter receptor genes in adverse maternal environment. *Transl Psychiatry* 2013;3:e218.
- Peters RL, Allen KJ, Dharmage SC, Tang ML, Koplin JJ, Ponsonby AL, et al. Skin prick test responses and allergen-specific IgE levels as predictors of peanut, egg, and sesame allergy in infants. *J Allergy Clin Immunol* 2013;132:874-80.
- Dang TD, Tang M, Choo S, Licciardi PV, Koplin JJ, Martin PE, et al. Increasing the accuracy of peanut allergy diagnosis by using Ara h 2. *J Allergy Clin Immunol* 2012;129:1056-63.
- Bair E, Tibshirani R. Semi-supervised methods to predict patient survival from gene expression data. *PLoS Biol* 2004;2:E108.
- McCarthy JJ, McLeod HL, Ginsburg GS. Genomic medicine: a decade of successes, challenges, and opportunities. *Sci Transl Med* 2013;5:189sr4.

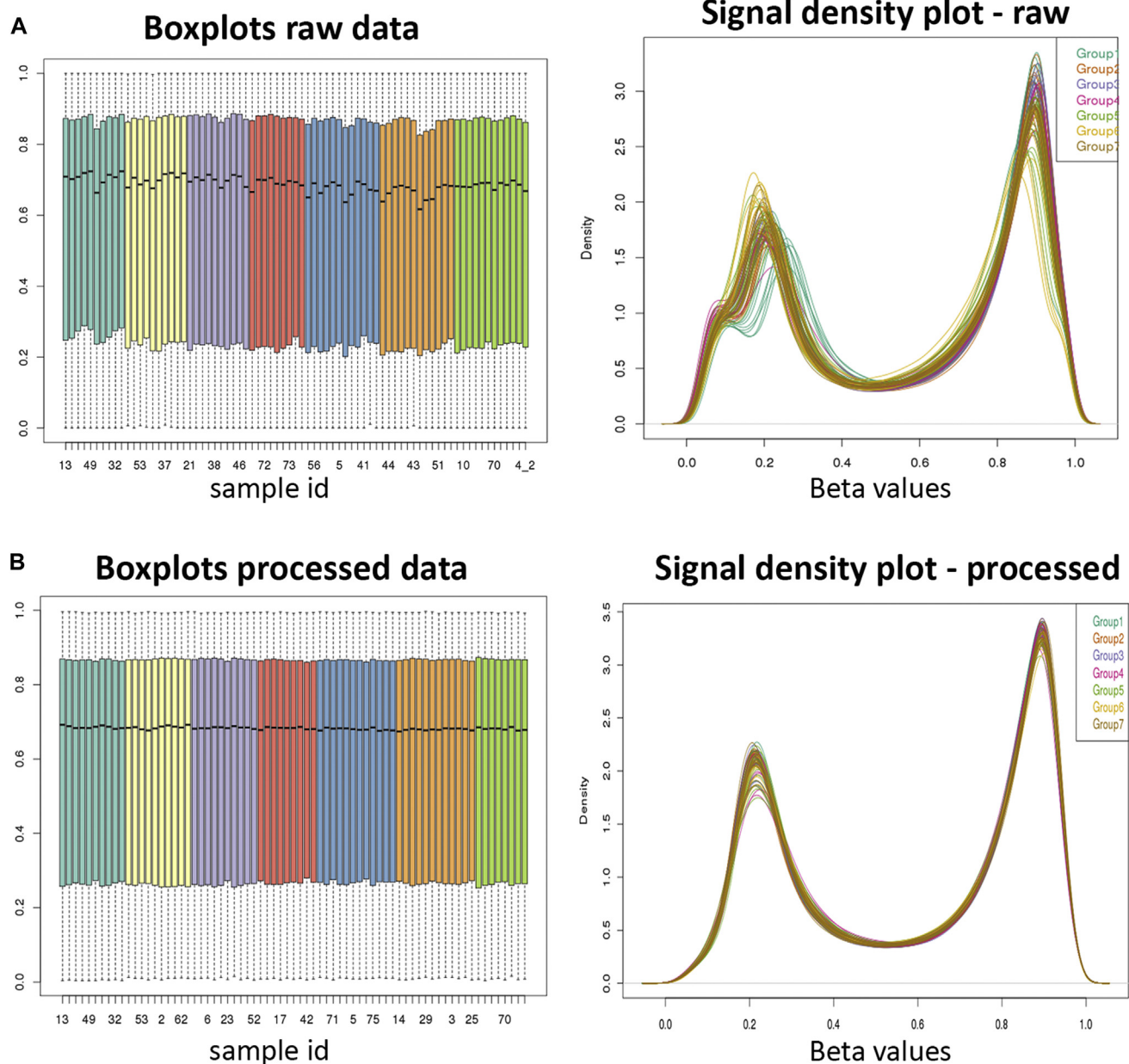


FIG E1. Quality control results for methylation data. **A**, Box plots and density distribution of raw unprocessed β values. **B**, Processed data after correcting for probe chemistry, probe level quality control, and normalization.

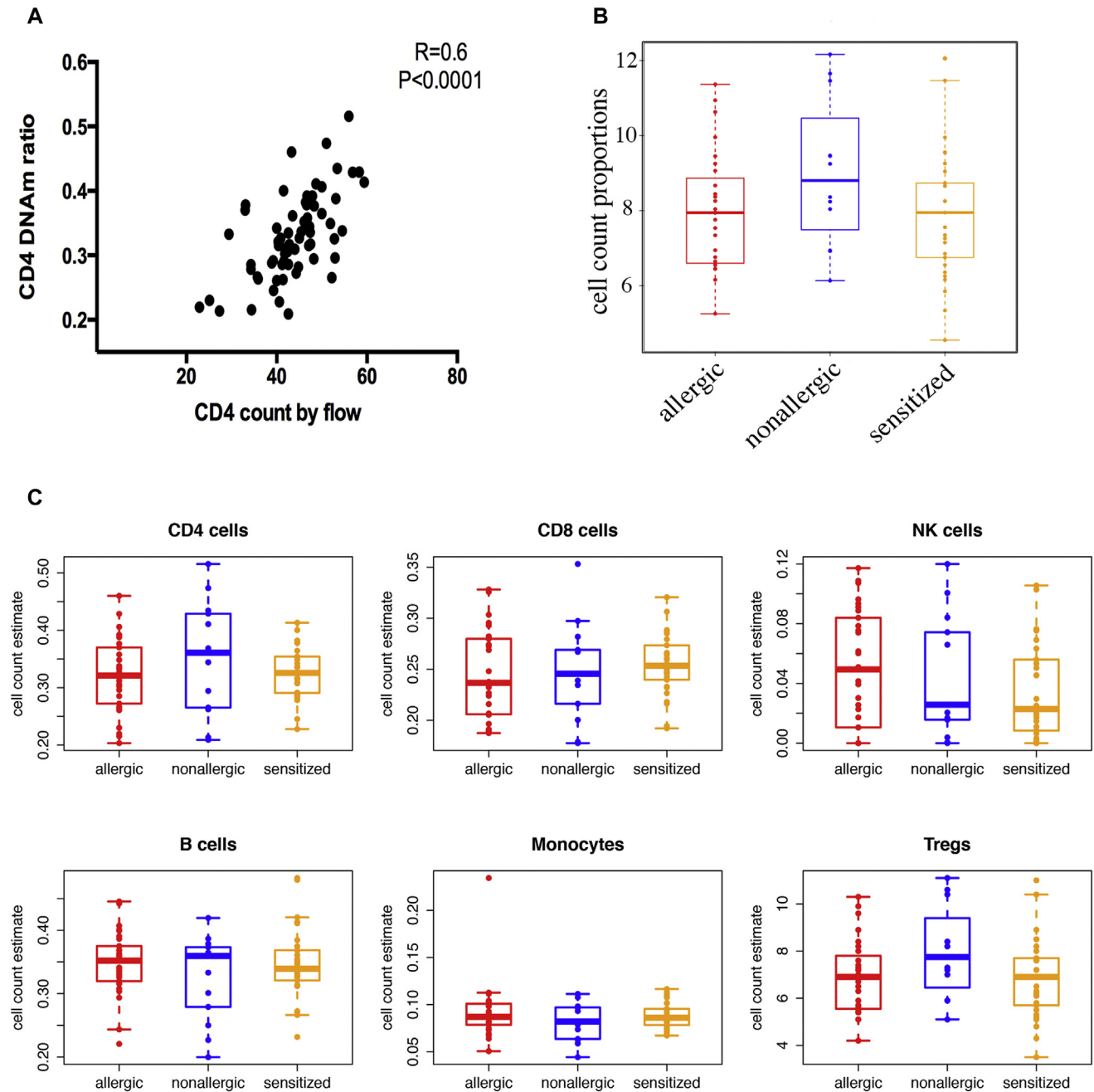


FIG E2. Assessment of confounding because of cellular heterogeneity of PBMC samples. **A**, Correlation between flow cytometry-measured total CD4⁺ counts and *in silico*-estimated CD4⁺ counts using cell type-predictive CpGs determined by using the method of Houseman. Correlations between gold standard and data-driven methods were highly comparable. **B**, Between-group comparisons of cell composition in PBMC fractions. By using *in silico* estimates for blood cell subsets, there was no evidence of differential methylation between primary outcome groups because of cell type proportion. **C**, Cell counts for each subtype are shown. Statistical analysis was done by using the Mann-Whitney *U* test.

TABLE E1. Gene set enrichment results

Gene symbol	Gene description	Gene set name	Gene set description
<i>VEGFA</i>	Vascular endothelial growth factor A	GSE13411_PLASMA_CELL_VS_MEMORY_BCELL_UP	Genes upregulated in comparison of plasma cells vs memory B cells
<i>TG</i>	Thyroglobulin	GSE13411_PLASMA_CELL_VS_MEMORY_BCELL_UP	Genes upregulated in comparison of plasma cells vs memory B cells
<i>TIMP2</i>	TIMP metalloproteinase inhibitor 2	GSE13411_PLASMA_CELL_VS_MEMORY_BCELL_UP	Genes upregulated in comparison of plasma cells vs memory B cells
<i>PLA2G16</i>	Phospholipase A2, group XVI	GSE13411_PLASMA_CELL_VS_MEMORY_BCELL_UP	Genes upregulated in comparison of plasma cells vs memory B cells
<i>COL3A1</i>	Collagen, type III, alpha 1	GSE13411_PLASMA_CELL_VS_MEMORY_BCELL_UP	Genes upregulated in comparison of plasma cells vs memory B cells
<i>VKORC1</i>	Vitamin K epoxide reductase complex, subunit 1	GSE13411_PLASMA_CELL_VS_MEMORY_BCELL_UP	Genes upregulated in comparison of plasma cells vs memory B cells
<i>LMF1</i>	Lipase maturation factor 1	GSE13411_PLASMA_CELL_VS_MEMORY_BCELL_UP	Genes upregulated in comparison of plasma cells vs memory B cells
<i>ELL2</i>	Elongation factor, RNA polymerase II, 2	GSE13411_PLASMA_CELL_VS_MEMORY_BCELL_UP	Genes upregulated in comparison of plasma cells vs memory B cells
<i>PSRC1</i>	Proline/serine-rich coiled-coil 1	GSE13411_PLASMA_CELL_VS_MEMORY_BCELL_UP	Genes upregulated in comparison of plasma cells vs memory B cells
<i>HIST1H2BI</i>	Histone cluster 1, H2bi	GSE13411_PLASMA_CELL_VS_MEMORY_BCELL_UP	Genes upregulated in comparison of plasma cells vs memory B cells
<i>HIST1H2AI</i>	Histone cluster 1, H2ai	GSE13411_PLASMA_CELL_VS_MEMORY_BCELL_UP	Genes upregulated in comparison of plasma cells vs memory B cells
<i>TRIM29</i>	Tripartite motif containing 29	GSE13411_PLASMA_CELL_VS_MEMORY_BCELL_UP	Genes upregulated in comparison of plasma cells vs memory B cells
<i>HLA-DMA</i>	MHC, class II, DM alpha	GSE17721_CPG_VS_GARDIQUIMOD_16H_BMDM_UP	Genes upregulated in comparison of DCs stimulated with CpG DNA (TLR9 agonist) at 16 h vs DCs stimulated with Gardiquimod (TLR7 agonist) at 16 h
<i>SMARCE1</i>	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily c, member 1	GSE17721_CPG_VS_GARDIQUIMOD_16H_BMDM_UP	Genes upregulated in comparison of DCs stimulated with CpG DNA (TLR9 agonist) at 16 h vs DCs stimulated with Gardiquimod (TLR7 agonist) at 16 h
<i>CD209</i>	CD209 molecule	GSE17721_CPG_VS_GARDIQUIMOD_16H_BMDM_UP	Genes upregulated in comparison of DCs stimulated with CpG DNA (TLR9 agonist) at 16 h vs DCs stimulated with Gardiquimod (TLR7 agonist) at 16 h
<i>PER3</i>	Period homolog 3 (Drosophila)	GSE17721_CPG_VS_GARDIQUIMOD_16H_BMDM_UP	Genes upregulated in comparison of DCs stimulated with CpG DNA (TLR9 agonist) at 16 h vs DCs stimulated with Gardiquimod (TLR7 agonist) at 16 h
<i>HIST3H2A</i>	Histone cluster 3, H2a	GSE17721_CPG_VS_GARDIQUIMOD_16H_BMDM_UP	Genes upregulated in comparison of DCs stimulated with CpG DNA (TLR9 agonist) at 16 h vs DCs stimulated with Gardiquimod (TLR7 agonist) at 16 h
<i>ANGPT1</i>	Angiopoietin 1	GSE17721_CPG_VS_GARDIQUIMOD_16H_BMDM_UP	Genes upregulated in comparison of DCs stimulated with CpG DNA (TLR9 agonist) at 16 h vs DCs stimulated with Gardiquimod (TLR7 agonist) at 16 h
<i>DHX16</i>	DEAH (Asp-Glu-Ala-His) box polypeptide 16	GSE17721_CPG_VS_GARDIQUIMOD_16H_BMDM_UP	Genes upregulated in comparison of DCs stimulated with CpG DNA (TLR9 agonist) at 16 h vs DCs stimulated with Gardiquimod (TLR7 agonist) at 16 h
<i>FAM50B</i>	Family with sequence similarity 50, member B	GSE17721_CPG_VS_GARDIQUIMOD_16H_BMDM_UP	Genes upregulated in comparison of DCs stimulated with CpG DNA (TLR9 agonist) at 16 h vs DCs stimulated with Gardiquimod (TLR7 agonist) at 16 h
<i>DCLK1</i>	Doublecortin-like kinase 1	GSE17721_CPG_VS_GARDIQUIMOD_16H_BMDM_UP	Genes upregulated in comparison of DCs stimulated with CpG DNA (TLR9 agonist) at 16 h vs DCs stimulated with Gardiquimod (TLR7 agonist) at 16 h
<i>ATP11A</i>	ATPase, class VI, type 11A	GSE17721_CPG_VS_GARDIQUIMOD_16H_BMDM_UP	Genes upregulated in comparison of DCs stimulated with CpG DNA (TLR9 agonist) at 16 h vs DCs stimulated with Gardiquimod (TLR7 agonist) at 16 h

(Continued)

TABLE E1. (Continued)

Gene symbol	Gene description	Gene set name	Gene set description
<i>UBAP1</i>	Ubiquitin associated protein 1	GSE17721_CPG_VS_GARDIQUIMOD_16H_BMDM_UP	Genes upregulated in comparison of DCs stimulated with CpG DNA (TLR9 agonist) at 16 h vs DCs stimulated with Gardiquimod (TLR7 agonist) at 16 h
<i>CARD14</i>	Caspase recruitment domain family, member 14	GSE17721_CPG_VS_GARDIQUIMOD_16H_BMDM_UP	Genes upregulated in comparison of DCs stimulated with CpG DNA (TLR9 agonist) at 16 h vs DCs stimulated with Gardiquimod (TLR7 agonist) at 16 h
<i>PLEKHA1</i>	Pleckstrin homology domain containing, family A (phosphoinositide binding specific) member 1	GSE17721_CTRL_VS_GARDIQUIMOD_4H_BMDM_UP	Genes upregulated in comparison of control DCs at 4 h vs those stimulated with Gardiquimod (TLR7 agonist) at 4 h
<i>MTFMT</i>	Mitochondrial methionyl-tRNA formyltransferase	GSE17721_CTRL_VS_GARDIQUIMOD_4H_BMDM_UP	Genes upregulated in comparison of control DCs at 4 h vs those stimulated with Gardiquimod (TLR7 agonist) at 4 h
<i>PANX1</i>	Pannexin 1	GSE17721_CTRL_VS_GARDIQUIMOD_4H_BMDM_UP	Genes upregulated in comparison of control DCs at 4 h vs those stimulated with Gardiquimod (TLR7 agonist) at 4 h
<i>FAM193B</i>	Family with sequence similarity 193, member B	GSE17721_CTRL_VS_GARDIQUIMOD_4H_BMDM_UP	Genes upregulated in comparison of control DCs at 4 h vs those stimulated with Gardiquimod (TLR7 agonist) at 4 h
<i>TRIM39</i>	Tripartite motif containing 39	GSE17721_CTRL_VS_GARDIQUIMOD_4H_BMDM_UP	Genes upregulated in comparison of control DCs at 4 h vs those stimulated with Gardiquimod (TLR7 agonist) at 4 h
<i>MAD1L1</i>	MAD1 mitotic arrest deficient-like 1 (yeast)	GSE17721_CTRL_VS_GARDIQUIMOD_4H_BMDM_UP	Genes upregulated in comparison of control DCs at 4 h vs those stimulated with Gardiquimod (TLR7 agonist) at 4 h
<i>C9orf21</i>	Chromosome 9 open reading frame 21	GSE17721_CTRL_VS_GARDIQUIMOD_4H_BMDM_UP	Genes upregulated in comparison of control DCs at 4 h vs those stimulated with Gardiquimod (TLR7 agonist) at 4 h
<i>EMILIN1</i>	Elastin microfibril interfacer 1	GSE17721_CTRL_VS_GARDIQUIMOD_4H_BMDM_UP	Genes upregulated in comparison of control DCs at 4 h vs those stimulated with Gardiquimod (TLR7 agonist) at 4 h
<i>MYO7A</i>	Myosin VIIA	GSE17721_CTRL_VS_GARDIQUIMOD_4H_BMDM_UP	Genes upregulated in comparison of control DCs at 4 h vs those stimulated with Gardiquimod (TLR7 agonist) at 4 h
<i>TREH</i>	Trehalase (brush-border membrane glycoprotein)	GSE17721_CTRL_VS_GARDIQUIMOD_4H_BMDM_UP	Genes upregulated in comparison of control DCs at 4 h vs those stimulated with Gardiquimod (TLR7 agonist) at 4 h
<i>AP2S1</i>	Adaptor-related protein complex 2, sigma 1 subunit	GSE17721_CTRL_VS_GARDIQUIMOD_4H_BMDM_UP	Genes upregulated in comparison of control DCs at 4 h vs those stimulated with Gardiquimod (TLR7 agonist) at 4 h
<i>SARS</i>	Seryl-tRNA synthetase	GSE17721_CTRL_VS_PAM3CSK4_4H_BMDM_UP	Genes upregulated in comparison of control DCs at 4 h vs those stimulated with Pam3Csk4 (TLR1/2 agonist) at 4 h
<i>OPA1</i>	Optic atrophy 1 (autosomal dominant)	GSE17721_CTRL_VS_PAM3CSK4_4H_BMDM_UP	Genes upregulated in comparison of control DCs at 4 h vs those stimulated with Pam3Csk4 (TLR1/2 agonist) at 4 h
<i>CACYBP</i>	Calcyclin binding protein	GSE17721_CTRL_VS_PAM3CSK4_4H_BMDM_UP	Genes upregulated in comparison of control DCs at 4 h vs those stimulated with Pam3Csk4 (TLR1/2 agonist) at 4 h
<i>MAST1</i>	Microtubule associated serine/threonine kinase 1	GSE17721_CTRL_VS_PAM3CSK4_4H_BMDM_UP	Genes upregulated in comparison of control DCs at 4 h vs those stimulated with Pam3Csk4 (TLR1/2 agonist) at 4 h
<i>FKBP11</i>	FK506 binding protein 11, 19 kDa	GSE17721_CTRL_VS_PAM3CSK4_4H_BMDM_UP	Genes upregulated in comparison of control DCs at 4 h vs those stimulated with Pam3Csk4 (TLR1/2 agonist) at 4 h
<i>KHK</i>	Ketohexokinase (fructokinase)	GSE17721_CTRL_VS_PAM3CSK4_4H_BMDM_UP	Genes upregulated in comparison of control DCs at 4 h vs those stimulated with Pam3Csk4 (TLR1/2 agonist) at 4 h
<i>CCNG1</i>	Cyclin G1	GSE17721_CTRL_VS_PAM3CSK4_4H_BMDM_UP	Genes upregulated in comparison of control DCs at 4 h vs those stimulated with Pam3Csk4 (TLR1/2 agonist) at 4 h

(Continued)

TABLE E1. (Continued)

Gene symbol	Gene description	Gene set name	Gene set description
<i>FAM149B1</i>	Family with sequence similarity 149, member B1	GSE17721_CTRL_VS_PAM3CSK4_4H_BMDM_UP	Genes upregulated in comparison of control DCs at 4 h vs those stimulated with Pam3Csk4 (TLR1/2 agonist) at 4 h
<i>KIAA1967</i>	KIAA1967	GSE17721_CTRL_VS_PAM3CSK4_4H_BMDM_UP	Genes upregulated in comparison of control DCs at 4 h vs those stimulated with Pam3Csk4 (TLR1/2 agonist) at 4 h
<i>SLC23A2</i>	Solute carrier family 23 (nucleobase transporters), member 2	GSE17721_POLYIC_VS_PAM3CSK4_1H_BMDM_DN	Genes downregulated in comparison of DCs stimulated with poly(I:C) (TLR3 agonist) at 1 h vs DCs stimulated with Pam3Csk4 (TLR1/2 agonist) at 1 h
<i>RPS6KA2</i>	Ribosomal protein S6 kinase, 90kDa, polypeptide 2	GSE17721_POLYIC_VS_PAM3CSK4_1H_BMDM_DN	Genes downregulated in comparison of DCs stimulated with poly(I:C) (TLR3 agonist) at 1 h vs DCs stimulated with Pam3Csk4 (TLR1/2 agonist) at 1 h
<i>RELL1</i>	RELT-like 1	GSE17721_POLYIC_VS_PAM3CSK4_1H_BMDM_DN	Genes downregulated in comparison of DCs stimulated with poly(I:C) (TLR3 agonist) at 1 h vs DCs stimulated with Pam3Csk4 (TLR1/2 agonist) at 1 h
<i>RYR1</i>	Ryanodine receptor 1 (skeletal)	GSE17721_POLYIC_VS_PAM3CSK4_1H_BMDM_DN	Genes downregulated in comparison of DCs stimulated with poly(I:C) (TLR3 agonist) at 1 h vs DCs stimulated with Pam3Csk4 (TLR1/2 agonist) at 1 h
<i>FHOD3</i>	Formin homology 2 domain containing 3	GSE17721_POLYIC_VS_PAM3CSK4_1H_BMDM_DN	Genes downregulated in comparison of DCs stimulated with poly(I:C) (TLR3 agonist) at 1 h vs DCs stimulated with Pam3Csk4 (TLR1/2 agonist) at 1 h
<i>PRIM2</i>	Primase, DNA, polypeptide 2 (58kDa)	GSE17721_POLYIC_VS_PAM3CSK4_1H_BMDM_DN	Genes downregulated in comparison of DCs stimulated with poly(I:C) (TLR3 agonist) at 1 h vs DCs stimulated with Pam3Csk4 (TLR1/2 agonist) at 1 h
<i>RAP1B</i>	RAP1B, member of RAS oncogene family	GSE17721_POLYIC_VS_PAM3CSK4_1H_BMDM_DN	Genes downregulated in comparison of DCs stimulated with poly(I:C) (TLR3 agonist) at 1 h vs DCs stimulated with Pam3Csk4 (TLR1/2 agonist) at 1 h
<i>USP13</i>	Ubiquitin specific peptidase 13 (isopeptidase T-3)	GSE17721_POLYIC_VS_PAM3CSK4_1H_BMDM_DN	Genes downregulated in comparison of DCs stimulated with poly(I:C) (TLR3 agonist) at 1 h vs DCs stimulated with Pam3Csk4 (TLR1/2 agonist) at 1 h
<i>TFF1</i>	Trefoil factor 1	GSE17721_POLYIC_VS_PAM3CSK4_1H_BMDM_DN	Genes downregulated in comparison of DCs stimulated with poly(I:C) (TLR3 agonist) at 1 h vs DCs stimulated with Pam3Csk4 (TLR1/2 agonist) at 1 h
<i>RTBDN</i>	Retbindin	GSE17721_POLYIC_VS_PAM3CSK4_1H_BMDM_DN	Genes downregulated in comparison of DCs stimulated with poly(I:C) (TLR3 agonist) at 1 h vs DCs stimulated with Pam3Csk4 (TLR1/2 agonist) at 1 h
<i>INTS12</i>	Integrator complex subunit 12	GSE17721_POLYIC_VS_PAM3CSK4_1H_BMDM_DN	Genes downregulated in comparison of DCs stimulated with poly(I:C) (TLR3 agonist) at 1 h vs DCs stimulated with Pam3Csk4 (TLR1/2 agonist) at 1 h
<i>DDX39A</i>	DEAD (Asp-Glu-Ala-Asp) box polypeptide 39A	GSE22886_NAIVE_TCELL_VS_NEUTROPHIL_UP	Genes upregulated in comparison of naive CD4 (Gene ID = 920) CD8 T cells vs unstimulated neutrophils
<i>CCT8</i>	Chaperonin containing TCP1, subunit 8 (theta)	GSE22886_NAIVE_TCELL_VS_NEUTROPHIL_UP	Genes upregulated in comparison of naive CD4 (Gene ID = 920) CD8 T cells vs unstimulated neutrophils
<i>CD7</i>	CD7 molecule	GSE22886_NAIVE_TCELL_VS_NEUTROPHIL_UP	Genes upregulated in comparison of naive CD4 (Gene ID = 920) CD8 T cells vs unstimulated neutrophils

(Continued)

TABLE E1. (Continued)

Gene symbol	Gene description	Gene set name	Gene set description
<i>NCL</i>	Nucleolin	GSE22886_NAIVE_TCELL_VS_NEUTROPHIL_UP	Genes upregulated in comparison of naive CD4 (Gene ID = 920) CD8 T cells vs unstimulated neutrophils
<i>C7orf23</i>	Chromosome 7 open reading frame 23	GSE22886_NAIVE_TCELL_VS_NEUTROPHIL_UP	Genes upregulated in comparison of naive CD4 (Gene ID = 920) CD8 T cells vs unstimulated neutrophils
<i>BCKDHA</i>	Branched chain keto acid dehydrogenase E1, alpha polypeptide	GSE22886_NAIVE_TCELL_VS_NEUTROPHIL_UP	Genes upregulated in comparison of naive CD4 (Gene ID = 920) CD8 T cells vs unstimulated neutrophils
<i>RCN2</i>	Reticulocalbin 2, EF-hand calcium binding domain	GSE22886_NAIVE_TCELL_VS_NEUTROPHIL_UP	Genes upregulated in comparison of naive CD4 (Gene ID = 920) CD8 T cells vs unstimulated neutrophils
<i>EIF2AK1</i>	Eukaryotic translation initiation factor 2-alpha kinase 1	GSE22886_NAIVE_TCELL_VS_NEUTROPHIL_UP	Genes upregulated in comparison of naive CD4 (Gene ID = 920) CD8 T cells vs unstimulated neutrophils
<i>GMPS</i>	Guanine monophosphate synthetase	GSE22886_NAIVE_TCELL_VS_NEUTROPHIL_UP	Genes upregulated in comparison of naive CD4 (Gene ID = 920) CD8 T cells vs unstimulated neutrophils
<i>ATP5D</i>	ATP synthase, H+ transporting, mitochondrial F1 complex, delta subunit	GSE22886_NAIVE_TCELL_VS_NEUTROPHIL_UP	Genes upregulated in comparison of naive CD4 (Gene ID = 920) CD8 T cells vs unstimulated neutrophils
<i>ATP1A1</i>	ATPase, Na+/K+ transporting, alpha 1 polypeptide	GSE22886_UNSTIM_VS_IL2_STIM_NKCELL_DN	Genes downregulated in comparison of unstimulated NK cells vs those stimulated with IL-2 (Gene ID = 3558) at 16 h
<i>LTB</i>	Lymphotoxin beta (TNF superfamily, member 3)	GSE22886_UNSTIM_VS_IL2_STIM_NKCELL_DN	Genes downregulated in comparison of unstimulated NK cells vs those stimulated with IL-2 (Gene ID = 3558) at 16 h
<i>GEM</i>	GTP binding protein overexpressed in skeletal muscle	GSE22886_UNSTIM_VS_IL2_STIM_NKCELL_DN	Genes downregulated in comparison of unstimulated NK cells vs those stimulated with IL-2 (Gene ID = 3558) at 16 h
<i>WDR12</i>	WD repeat domain 12	GSE22886_UNSTIM_VS_IL2_STIM_NKCELL_DN	Genes downregulated in comparison of unstimulated NK cells vs those stimulated with IL-2 (Gene ID = 3558) at 16 h
<i>HNRNPC</i>	Heterogeneous nuclear ribonucleoprotein C (C1/C2)	GSE22886_UNSTIM_VS_IL2_STIM_NKCELL_DN	Genes downregulated in comparison of unstimulated NK cells vs those stimulated with IL-2 (Gene ID = 3558) at 16 h
<i>PTPN7</i>	Protein tyrosine phosphatase, non-receptor type 7	GSE22886_UNSTIM_VS_IL2_STIM_NKCELL_DN	Genes downregulated in comparison of unstimulated NK cells vs those stimulated with IL-2 (Gene ID = 3558) at 16 h
<i>FASN</i>	Fatty acid synthase	GSE22886_UNSTIM_VS_IL2_STIM_NKCELL_DN	Genes downregulated in comparison of unstimulated NK cells vs those stimulated with IL-2 (Gene ID = 3558) at 16 h
<i>ACACA</i>	Acetyl-CoA carboxylase alpha	GSE22886_UNSTIM_VS_IL2_STIM_NKCELL_DN	Genes downregulated in comparison of unstimulated NK cells vs those stimulated with IL-2 (Gene ID = 3558) at 16 h
<i>PAICS</i>	Phosphoribosylaminoimidazole carboxylase, phosphoribosylaminoimidazole succinocarboxamide synthetase	GSE22886_UNSTIM_VS_IL2_STIM_NKCELL_DN	Genes downregulated in comparison of unstimulated NK cells vs those stimulated with IL-2 (Gene ID = 3558) at 16 h
<i>NAA16</i>	N(alpha)-acetyltransferase 16, NatA auxiliary subunit	GSE26669_CD4_VS_CD8_TCELL_IN_MLR_COSTIM_BLOCK_UP	Genes upregulated in comparison of untreated CD4 (Gene ID = 920) T cells vs CD8 T cells treated with leukocyte costimulatory blockade antibodies
<i>CERS4</i>	Ceramide synthase 4	GSE26669_CD4_VS_CD8_TCELL_IN_MLR_COSTIM_BLOCK_UP	Genes upregulated in comparison of untreated CD4 (Gene ID = 920) T cells vs CD8 T cells treated with leukocyte costimulatory blockade antibodies
<i>BRE</i>	Brain and reproductive organ-expressed (TNFRSF1A modulator)	GSE26669_CD4_VS_CD8_TCELL_IN_MLR_COSTIM_BLOCK_UP	Genes upregulated in comparison of untreated CD4 (Gene ID = 920) T cells vs CD8 T cells treated with leukocyte costimulatory blockade antibodies

(Continued)

TABLE E1. (Continued)

Gene symbol	Gene description	Gene set name	Gene set description
<i>STT3B</i>	STT3, subunit of the oligosaccharyltransferase complex, homolog B (<i>S. cerevisiae</i>)	GSE26669_CD4_VS_CD8_TCELL_IN_MLR_COSTIM_BLOCK_UP	Genes upregulated in comparison of untreated CD4 (Gene ID = 920) T cells vs CD8 T cells treated with leukocyte costimulatory blockade antibodies
<i>NTN4</i>	Netrin 4	GSE26669_CD4_VS_CD8_TCELL_IN_MLR_COSTIM_BLOCK_UP	Genes upregulated in comparison of untreated CD4 (Gene ID = 920) T cells vs CD8 T cells treated with leukocyte costimulatory blockade antibodies
<i>ATRNL1</i>	Attractin-like 1	GSE26669_CD4_VS_CD8_TCELL_IN_MLR_COSTIM_BLOCK_UP	Genes upregulated in comparison of untreated CD4 (Gene ID = 920) T cells vs CD8 T cells treated with leukocyte costimulatory blockade antibodies
<i>GATA3</i>	GATA binding protein 3	GSE26669_CD4_VS_CD8_TCELL_IN_MLR_COSTIM_BLOCK_UP	Genes upregulated in comparison of untreated CD4 (Gene ID = 920) T cells vs CD8 T cells treated with leukocyte costimulatory blockade antibodies
<i>RASGRP2</i>	RAS guanyl releasing protein 2 (calcium and DAG-regulated)	GSE26669_CD4_VS_CD8_TCELL_IN_MLR_COSTIM_BLOCK_UP	Genes upregulated in comparison of untreated CD4 (Gene ID = 920) T cells vs CD8 T cells treated with leukocyte costimulatory blockade antibodies
<i>CSNK1G3</i>	Casein kinase 1, gamma 3	GSE26669_CD4_VS_CD8_TCELL_IN_MLR_COSTIM_BLOCK_UP	Genes upregulated in comparison of untreated CD4 (Gene ID = 920) T cells vs CD8 T cells treated with leukocyte costimulatory blockade antibodies
<i>ABI2</i>	Abl-interactor 2	GSE26669_CD4_VS_CD8_TCELL_IN_MLR_COSTIM_BLOCK_UP	Genes upregulated in comparison of untreated CD4 (Gene ID = 920) T cells vs CD8 T cells treated with leukocyte costimulatory blockade antibodies
<i>ACSL3</i>	Acyl-CoA synthetase long-chain family member 3	GSE26669_CD4_VS_CD8_TCELL_IN_MLR_COSTIM_BLOCK_UP	Genes upregulated in comparison of untreated CD4 (Gene ID = 920) T cells vs CD8 T cells treated with leukocyte costimulatory blockade antibodies
<i>BRSK2</i>	BR serine/threonine kinase 2	GSE26669_CD4_VS_CD8_TCELL_IN_MLR_COSTIM_BLOCK_UP	Genes upregulated in comparison of untreated CD4 (Gene ID = 920) T cells vs CD8 T cells treated with leukocyte costimulatory blockade antibodies
<i>DTX1</i>	Deltex homolog 1 (<i>Drosophila</i>)	GSE11057_EFF_MEM_VS_CENT_MEM_CD4_TCELL_DN	Genes downregulated in comparison of effector memory T cells vs central memory T cells from PBMCs
<i>SLC24A6</i>	Solute carrier family 24 (sodium/potassium/calcium exchanger), member 6	GSE11057_EFF_MEM_VS_CENT_MEM_CD4_TCELL_DN	Genes downregulated in comparison of effector memory T cells vs central memory T cells from PBMCs
<i>PKMYT1</i>	Protein kinase, membrane associated tyrosine/threonine 1	GSE11057_EFF_MEM_VS_CENT_MEM_CD4_TCELL_DN	Genes downregulated in comparison of effector memory T cells vs central memory T cells from PBMCs
<i>MAP3K1</i>	Mitogen-activated protein kinase kinase kinase 1	GSE11057_EFF_MEM_VS_CENT_MEM_CD4_TCELL_DN	Genes downregulated in comparison of effector memory T cells vs central memory T cells from PBMCs
<i>ALS2CL</i>	ALS2 C-terminal like	GSE11057_EFF_MEM_VS_CENT_MEM_CD4_TCELL_DN	Genes downregulated in comparison of effector memory T cells vs central memory T cells from PBMCs
<i>MAML2</i>	Mastermind-like 2 (<i>Drosophila</i>)	GSE11057_EFF_MEM_VS_CENT_MEM_CD4_TCELL_DN	Genes downregulated in comparison of effector memory T cells vs central memory T cells from PBMCs
<i>USP34</i>	Ubiquitin specific peptidase 34	GSE11057_EFF_MEM_VS_CENT_MEM_CD4_TCELL_DN	Genes downregulated in comparison of effector memory T cells vs central memory T cells from PBMCs
<i>SLC39A10</i>	Solute carrier family 39 (zinc transporter), member 10	GSE11057_EFF_MEM_VS_CENT_MEM_CD4_TCELL_DN	Genes downregulated in comparison of effector memory T cells vs central memory T cells from PBMCs

(Continued)

TABLE E1. (Continued)

Gene symbol	Gene description	Gene set name	Gene set description
<i>LOC338799</i>	Uncharacterized LOC338799	GSE11057_EFF_MEM_VS_CENT_MEM_CD4_TCELL_DN	Genes downregulated in comparison of effector memory T cells vs central memory T cells from PBMCs
<i>ILF3</i>	Interleukin enhancer binding factor 3, 90kDa	GSE12845_IGD_POS_BLOOD_VS_NAIVE_TONSIL_BCELL_DN	Genes downregulated in comparison of IgD ⁺ peripheral blood B cells vs IgD ⁻ naive tonsil B cells
<i>BIK</i>	BCL2-interacting killer (apoptosis-inducing)	GSE12845_IGD_POS_BLOOD_VS_NAIVE_TONSIL_BCELL_DN	Genes downregulated in comparison of IgD ⁺ peripheral blood B cells vs IgD ⁻ naive tonsil B cells
<i>EGR2</i>	Early growth response 2	GSE12845_IGD_POS_BLOOD_VS_NAIVE_TONSIL_BCELL_DN	Genes downregulated in comparison of IgD ⁺ peripheral blood B cells vs IgD ⁻ naive tonsil B cells
<i>SPAG4</i>	Sperm associated antigen 4	GSE12845_IGD_POS_BLOOD_VS_NAIVE_TONSIL_BCELL_DN	Genes downregulated in comparison of IgD ⁺ peripheral blood B cells vs IgD ⁻ naive tonsil B cells
<i>LPP</i>	LIM domain containing preferred translocation partner in lipoma	GSE12845_IGD_POS_BLOOD_VS_NAIVE_TONSIL_BCELL_DN	Genes downregulated in comparison of IgD ⁺ peripheral blood B cells vs IgD ⁻ naive tonsil B cells
<i>TAF1D</i>	TATA box binding protein (TBP)-associated factor, RNA polymerase I, D, 41kDa	GSE12845_IGD_POS_BLOOD_VS_NAIVE_TONSIL_BCELL_DN	Genes downregulated in comparison of IgD ⁺ peripheral blood B cells vs IgD ⁻ naive tonsil B cells
<i>USP8</i>	Ubiquitin specific peptidase 8	GSE12845_IGD_POS_BLOOD_VS_NAIVE_TONSIL_BCELL_DN	Genes downregulated in comparison of IgD ⁺ peripheral blood B cells vs IgD ⁻ naive tonsil B cells
<i>TMEM176B</i>	Transmembrane protein 176B	GSE13229_IMM_VS_MATURE_NKCELL_UP	Genes upregulated in comparison of immature NK cells vs mature NK cells
<i>RNF213</i>	Ring finger protein 213	GSE13229_IMM_VS_MATURE_NKCELL_UP	Genes upregulated in comparison of immature NK cells vs mature NK cells
<i>RPN1</i>	Ribophorin I	GSE13229_IMM_VS_MATURE_NKCELL_UP	Genes upregulated in comparison of immature NK cells vs mature NK cells
<i>CTBP2</i>	C-terminal binding protein 2	GSE13229_IMM_VS_MATURE_NKCELL_UP	Genes upregulated in comparison of immature NK cells vs mature NK cells
<i>DAAM2</i>	Disheveled associated activator of morphogenesis 2	GSE13229_IMM_VS_MATURE_NKCELL_UP	Genes upregulated in comparison of immature NK cells vs mature NK cells
<i>IL20RA</i>	Interleukin 20 receptor, alpha	GSE13229_IMM_VS_MATURE_NKCELL_UP	Genes upregulated in comparison of immature NK cells vs mature NK cells
<i>NAT14</i>	N-acetyltransferase 14 (GCN5-related, putative)	GSE13229_IMM_VS_MATURE_NKCELL_UP	Genes upregulated in comparison of immature NK cells vs mature NK cells
<i>SH2D5</i>	SH2 domain containing 5	GSE13229_IMM_VS_MATURE_NKCELL_UP	Genes upregulated in comparison of immature NK cells vs mature NK cells
<i>ZC3HAV1L</i>	Zinc finger CCCH-type, antiviral 1-like	GSE13229_IMM_VS_MATURE_NKCELL_UP	Genes upregulated in comparison of immature NK cells vs mature NK cells
<i>ARF4</i>	ADP-ribosylation factor 4	GSE13306_TREG_VS_TCONV_UP	Genes upregulated in Treg cells vs conventional T cells
<i>ARHGEF25</i>	Rho guanine nucleotide exchange factor (GEF) 25	GSE13306_TREG_VS_TCONV_UP	Genes upregulated in Treg cells vs conventional T cells
<i>CAMKK1</i>	Calcium/calmodulin-dependent protein kinase kinase 1, alpha	GSE13306_TREG_VS_TCONV_UP	Genes upregulated in Treg cells vs conventional T cells
<i>ACOT7</i>	Acyl-CoA thioesterase 7	GSE13306_TREG_VS_TCONV_UP	Genes upregulated in Treg cells vs conventional T cells
<i>SH3PXD2A</i>	SH3 and PX domains 2A	GSE13306_TREG_VS_TCONV_UP	Genes upregulated in Treg cells vs conventional T cells
<i>TRMT12</i>	tRNA methyltransferase 12 homolog (S. cerevisiae)	GSE13306_TREG_VS_TCONV_UP	Genes upregulated in Treg cells vs conventional T cells
<i>TRPM1</i>	Transient receptor potential cation channel, subfamily M, member 1	GSE13306_TREG_VS_TCONV_UP	Genes upregulated in Treg cells vs conventional T cells

(Continued)

TABLE E1. (Continued)

Gene symbol	Gene description	Gene set name	Gene set description
<i>RGNEF</i>	190 kDa guanine nucleotide exchange factor	GSE13306_TREG_VS_TCONV_UP	Genes upregulated in Treg cells vs conventional T cells
<i>SLC45A2</i>	Solute carrier family 45, member 2	GSE13306_TREG_VS_TCONV_UP	Genes upregulated in Treg cells vs conventional T cells
<i>ARHGAP35</i>	Rho GTPase activating protein 35	GSE15659_CD45RA_NEG_CD4_TCELL_VS_ACTIVATED_TREG_UP	Genes upregulated in comparison of PTPRC [−] (Gene ID = 5788) CD4 (Gene ID = 920) T cells vs activated Treg cells
<i>EXOC2</i>	Exocyst complex component 2	GSE15659_CD45RA_NEG_CD4_TCELL_VS_ACTIVATED_TREG_UP	Genes upregulated in comparison of PTPRC [−] (Gene ID = 5788) CD4 (Gene ID = 920) T cells vs activated Treg cells
<i>C12orf23</i>	Chromosome 12 open reading frame 23	GSE15659_CD45RA_NEG_CD4_TCELL_VS_ACTIVATED_TREG_UP	Genes upregulated in comparison of PTPRC [−] (Gene ID = 5788) CD4 (Gene ID = 920) T cells vs activated Treg cells
<i>FBXW9</i>	F-box and WD repeat domain containing 9	GSE15659_CD45RA_NEG_CD4_TCELL_VS_ACTIVATED_TREG_UP	Genes upregulated in comparison of PTPRC [−] (Gene ID = 5788) CD4 (Gene ID = 920) T cells vs activated Treg cells
<i>GFRA1</i>	GNDF family receptor alpha 1	GSE15659_CD45RA_NEG_CD4_TCELL_VS_ACTIVATED_TREG_UP	Genes upregulated in comparison of PTPRC [−] (Gene ID = 5788) CD4 (Gene ID = 920) T cells vs activated Treg cells
<i>FNDC8</i>	Fibronectin type III domain containing 8	GSE15659_CD45RA_NEG_CD4_TCELL_VS_ACTIVATED_TREG_UP	Genes upregulated in comparison of PTPRC [−] (Gene ID = 5788) CD4 (Gene ID = 920) T cells vs activated Treg cells
<i>C1QTNF7</i>	C1q and tumor necrosis factor related protein 7	GSE15659_CD45RA_NEG_CD4_TCELL_VS_ACTIVATED_TREG_UP	Genes upregulated in comparison of PTPRC [−] (Gene ID = 5788) CD4 (Gene ID = 920) T cells vs activated Treg cells
<i>CDHR5</i>	Cadherin-related family member 5	GSE15659_CD45RA_NEG_CD4_TCELL_VS_ACTIVATED_TREG_UP	Genes upregulated in comparison of PTPRC [−] (Gene ID = 5788) CD4 (Gene ID = 920) T cells vs activated Treg cells
<i>CCS</i>	Copper chaperone for superoxide dismutase	GSE15659_CD45RA_NEG_CD4_TCELL_VS_ACTIVATED_TREG_UP	Genes upregulated in comparison of PTPRC [−] (Gene ID = 5788) CD4 (Gene ID = 920) T cells vs activated Treg cells
<i>ATOH7</i>	Atonal homolog 7 (Drosophila)	GSE15659_CD45RA_NEG_CD4_TCELL_VS_ACTIVATED_TREG_UP	Genes upregulated in comparison of PTPRC [−] (Gene ID = 5788) CD4 (Gene ID = 920) T cells vs activated Treg cells
<i>FRMD8</i>	FERM domain containing 8	GSE17721_0.5H_VS_12H_POLYIC_BMDM_DN	Genes downregulated in comparison of DCs stimulated with poly(I:C) (TLR3 agonist) at 0.5 h vs those stimulated at 4 h
<i>DUSP26</i>	Dual specificity phosphatase 26 (putative)	GSE17721_0.5H_VS_12H_POLYIC_BMDM_DN	Genes downregulated in comparison of DCs stimulated with poly(I:C) (TLR3 agonist) at 0.5 h vs those stimulated at 4 h
<i>RNF34</i>	Ring finger protein 34	GSE17721_0.5H_VS_12H_POLYIC_BMDM_DN	Genes downregulated in comparison of DCs stimulated with poly(I:C) (TLR3 agonist) at 0.5 h vs those stimulated at 4 h
<i>KCNMA1</i>	Potassium large conductance calcium-activated channel, subfamily M, alpha member 1	GSE17721_0.5H_VS_12H_POLYIC_BMDM_DN	Genes downregulated in comparison of DCs stimulated with poly(I:C) (TLR3 agonist) at 0.5 h vs those stimulated at 4 h
<i>DOCK6</i>	Dedicator of cytokinesis 6	GSE17721_0.5H_VS_12H_POLYIC_BMDM_DN	Genes downregulated in comparison of DCs stimulated with poly(I:C) (TLR3 agonist) at 0.5 h vs those stimulated at 4 h
<i>PLK2</i>	Polo-like kinase 2	GSE17721_12H_VS_24H_PAM3CSK4_BMDM_UP	Genes upregulated in comparison of DCs stimulated with Pam3Csk4 (TLR1/2 agonist) at 12 h vs those stimulated with Pam3Csk4 (TLR1/2 agonist) at 24 h
<i>WRNIP1</i>	Werner helicase interacting protein 1	GSE17721_12H_VS_24H_PAM3CSK4_BMDM_UP	Genes upregulated in comparison of DCs stimulated with Pam3Csk4 (TLR1/2 agonist) at 12 h vs those stimulated with Pam3Csk4 (TLR1/2 agonist) at 24 h
<i>ATP10A</i>	ATPase, class V, type 10A	GSE17721_12H_VS_24H_PAM3CSK4_BMDM_UP	Genes upregulated in comparison of DCs stimulated with Pam3Csk4 (TLR1/2 agonist) at 12 h vs those stimulated with Pam3Csk4 (TLR1/2 agonist) at 24 h

(Continued)

TABLE E1. (Continued)

Gene symbol	Gene description	Gene set name	Gene set description
<i>SLC3A2</i>	Solute carrier family 3 (activators of dibasic and neutral amino acid transport), member 2	GSE17721_12H_VS_24H_PAM3CSK4_BMDM_UP	Genes upregulated in comparison of DCs stimulated with Pam3Csk4 (TLR1/2 agonist) at 12 h vs those stimulated with Pam3Csk4 (TLR1/2 agonist) at 24 h
<i>EPB41L2</i>	Erythrocyte membrane protein band 4.1-like 2	GSE17721_12H_VS_24H_PAM3CSK4_BMDM_UP	Genes upregulated in comparison of DCs stimulated with Pam3Csk4 (TLR1/2 agonist) at 12 h vs those stimulated with Pam3Csk4 (TLR1/2 agonist) at 24 h
<i>CCNJ</i>	Cyclin J	GSE17721_12H_VS_24H_PAM3CSK4_BMDM_UP	Genes upregulated in comparison of DCs stimulated with Pam3Csk4 (TLR1/2 agonist) at 12 h vs those stimulated with Pam3Csk4 (TLR1/2 agonist) at 24 h
<i>ARID1A</i>	AT rich interactive domain 1A (SWI-like)	GSE17721_CTRL_VS_CPG_2H_BMDM_UP	Genes upregulated in comparison of control DCs at 2 h vs those stimulated with CpG DNA (TLR9 agonist) at 2 h
<i>RGS19</i>	Regulator of G-protein signaling 19	GSE17721_CTRL_VS_CPG_2H_BMDM_UP	Genes upregulated in comparison of control DCs at 2 h vs those stimulated with CpG DNA (TLR9 agonist) at 2 h
<i>TF</i>	Transferrin	GSE17721_CTRL_VS_CPG_2H_BMDM_UP	Genes upregulated in comparison of control DCs at 2 h vs those stimulated with CpG DNA (TLR9 agonist) at 2 h
<i>SAP30L</i>	SAP30-like	GSE17721_CTRL_VS_CPG_2H_BMDM_UP	Genes upregulated in comparison of control DCs at 2 h vs those stimulated with CpG DNA (TLR9 agonist) at 2 h
<i>WDR33</i>	WD repeat domain 33	GSE17721_CTRL_VS_CPG_2H_BMDM_UP	Genes upregulated in comparison of control DCs at 2 h vs those stimulated with CpG DNA (TLR9 agonist) at 2 h
<i>FAM53A</i>	Family with sequence similarity 53, member A	GSE17721_CTRL_VS_CPG_2H_BMDM_UP	Genes upregulated in comparison of control DCs at 2 h vs those stimulated with CpG DNA (TLR9 agonist) at 2 h
<i>IER3</i>	Immediate early response 3	GSE17721_CTRL_VS_LPS_24H_BMDM_DN	Genes downregulated in comparison of control DCs at 24 h vs those stimulated with LPS (TLR4 agonist) at 24 h
<i>COX16</i>	COX16 cytochrome c oxidase assembly homolog (<i>S. cerevisiae</i>)	GSE17721_CTRL_VS_LPS_24H_BMDM_DN	Genes downregulated in comparison of control DCs at 24 h vs those stimulated with LPS (TLR4 agonist) at 24 h
<i>GNB1</i>	Guanine nucleotide binding protein (G protein), beta polypeptide 1	GSE17721_CTRL_VS_LPS_24H_BMDM_DN	Genes downregulated in comparison of control DCs at 24 h vs those stimulated with LPS (TLR4 agonist) at 24 h
<i>RILPL2</i>	Rab interacting lysosomal protein-like 2	GSE17721_LPS_VS_CPG_6H_BMDM_UP	Genes upregulated in comparison of DCs stimulated with LPS (TLR4 agonist) at 6 h vs DCs stimulated with CpG DNA (TLR9 agonist) at 6 h
<i>PRKD2</i>	Protein kinase D2	GSE17721_LPS_VS_CPG_6H_BMDM_UP	Genes upregulated in comparison of DCs stimulated with LPS (TLR4 agonist) at 6 h vs DCs stimulated with CpG DNA (TLR9 agonist) at 6 h
<i>TRIM21</i>	Tripartite motif containing 21	GSE17721_LPS_VS_CPG_6H_BMDM_UP	Genes upregulated in comparison of DCs stimulated with LPS (TLR4 agonist) at 6 h vs DCs stimulated with CpG DNA (TLR9 agonist) at 6 h
<i>IQSEC1</i>	IQ motif and Sec7 domain 1	GSE17721_LPS_VS_CPG_6H_BMDM_UP	Genes upregulated in comparison of DCs stimulated with LPS (TLR4 agonist) at 6 h vs DCs stimulated with CpG DNA (TLR9 agonist) at 6 h
<i>MAFK</i>	v-maf musculoaponeurotic fibrosarcoma oncogene homolog K (avian)	GSE17721_LPS_VS_CPG_6H_BMDM_UP	Genes upregulated in comparison of DCs stimulated with LPS (TLR4 agonist) at 6 h vs DCs stimulated with CpG DNA (TLR9 agonist) at 6 h

(Continued)

TABLE E1. (Continued)

Gene symbol	Gene description	Gene set name	Gene set description
<i>TADA2A</i>	Transcriptional adaptor 2A	GSE17974_IL4_AND_ANTI_IL12_VS_UNTREATED_6H_ACT_CD4_TCELL_DN	Genes downregulated in comparison of CD4 (Gene ID = 920) T cells treated with IL-4 (Gene ID = 3565) and anti-IL-12 at 6 h vs untreated cells at 6 h
<i>TOX2</i>	TOX high mobility group box family member 2	GSE17974_IL4_AND_ANTI_IL12_VS_UNTREATED_6H_ACT_CD4_TCELL_DN	Genes downregulated in comparison of CD4 (Gene ID = 920) T cells treated with IL-4 (Gene ID = 3565) and anti-IL-12 at 6 h vs untreated cells at 6 h
<i>C2CD3</i>	C2 calcium-dependent domain containing 3	GSE17974_IL4_AND_ANTI_IL12_VS_UNTREATED_6H_ACT_CD4_TCELL_DN	Genes downregulated in comparison of CD4 (Gene ID = 920) T cells treated with IL-4 (Gene ID = 3565) and anti-IL-12 at 6 h vs untreated cells at 6 h
<i>NEDD4</i>	Neural precursor cell expressed, developmentally down-regulated 4	GSE17974_IL4_AND_ANTI_IL12_VS_UNTREATED_6H_ACT_CD4_TCELL_DN	Genes downregulated in comparison of CD4 (Gene ID = 920) T cells treated with IL-4 (Gene ID = 3565) and anti-IL-12 at 6 h vs untreated cells at 6 h
<i>MAP3K13</i>	Mitogen-activated protein kinase kinase kinase 13	GSE17974_IL4_AND_ANTI_IL12_VS_UNTREATED_6H_ACT_CD4_TCELL_DN	Genes downregulated in comparison of CD4 (Gene ID = 920) T cells treated with IL-4 (Gene ID = 3565) and anti-IL-12 at 6 h vs untreated cells at 6 h
<i>PDGFA</i>	Platelet-derived growth factor alpha polypeptide	GSE17974_IL4_AND_ANTI_IL12_VS_UNTREATED_6H_ACT_CD4_TCELL_DN	Genes downregulated in comparison of CD4 (Gene ID = 920) T cells treated with IL-4 (Gene ID = 3565) and anti-IL-12 at 6 h vs untreated cells at 6 h
<i>COL6A3</i>	Collagen, type VI, alpha 3	GSE17974_IL4_AND_ANTI_IL12_VS_UNTREATED_6H_ACT_CD4_TCELL_DN	Genes downregulated in comparison of CD4 (Gene ID = 920) T cells treated with IL-4 (Gene ID = 3565) and anti-IL-12 at 6 h vs untreated cells at 6 h
<i>KIAA1199</i>	KIAA1199	GSE17974_IL4_AND_ANTI_IL12_VS_UNTREATED_6H_ACT_CD4_TCELL_DN	Genes downregulated in comparison of CD4 (Gene ID = 920) T cells treated with IL-4 (Gene ID = 3565) and anti-IL-12 at 6 h vs untreated cells at 6 h
<i>LRCH1</i>	Leucine-rich repeats and calponin homology (CH) domain containing 1	GSE17974_IL4_AND_ANTI_IL12_VS_UNTREATED_6H_ACT_CD4_TCELL_DN	Genes downregulated in comparison of CD4 (Gene ID = 920) T cells treated with IL-4 (Gene ID = 3565) and anti-IL-12 at 6 h vs untreated cells at 6 h
<i>TMEM154</i>	Transmembrane protein 154	GSE20366_EX_VIVO_VS_DEC205_CONVERSION_NAIVE_CD4_TCELL_UP	Genes upregulated in comparison of TconvLP vs DEC-Pept CD25 ⁺ (see Table E1 for details)
<i>ZBTB24</i>	Zinc finger and BTB domain containing 24	GSE20366_EX_VIVO_VS_DEC205_CONVERSION_NAIVE_CD4_TCELL_UP	Genes upregulated in comparison of TconvLP vs DEC-Pept CD25 ⁺ (see Table E1 for details)
<i>CHST11</i>	Carbohydrate (chondroitin 4) sulfotransferase 11	GSE20366_EX_VIVO_VS_DEC205_CONVERSION_NAIVE_CD4_TCELL_UP	Genes upregulated in comparison of TconvLP vs DEC-Pept CD25 ⁺ (see Table E1 for details)
<i>TMEM176A</i>	Transmembrane protein 176A	GSE20366_EX_VIVO_VS_DEC205_CONVERSION_NAIVE_CD4_TCELL_UP	Genes upregulated in comparison of TconvLP vs DEC-Pept CD25 ⁺ (see Table E1 for details)

DC, Dendritic cell; LPS, lipopolysaccharide; NK, natural killer; *poly(I:C)*, polyinosinic-polycytidylic acid; *PTPRC*, protein tyrosine phosphatase receptor C; *Tconv*, conventional T; *TLR*, Toll-like receptor.

TABLE E2. Results from prediction models based on 80:20 split of training and test data

	True		Error rate
	Allergic	Sensitized	
Training set (HealthNuts discovery cohort)			
Predicted			
Allergic	22	1	0.043
Sensitized	0	22	0.000
Overall error rate = 0.022			
Validation set (HealthNuts discovery cohort)			
Predicted			
Allergic	6	0	0.000
Sensitized	0	7	0.000
Overall error rate = 0.000			