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;135:1319-28.)

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

Food allergy (FA) affects up to 10% of children in the most affected countries[1] 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

From [a]the Centre for Food and Allergy Research, Murdoch Childrens Research Institute, Parkville; [b]Murdoch Childrens Research Institute, Royal Children's Hospital, Parkville; [c]the School of Paediatrics and Child Health, University of Western Australia, Perth; [d]the University of Melbourne Department of Paediatrics; [e]the Department of Allergy and Immunology, Royal Children's Hospital, Parkville; [f]In-FLAME, the International Inflammation Network, World Universities Network (WUN); [g]the School of Population and Global Health, University of Melbourne; and [h]the Institute of Inflammation and Repair, University of Manchester.

*These author 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 Nestlè Nutrition Institute, has received payment for delivering lectures from Danone Nutricia and the Nestlè Nutrition Institute, and receives royalties from UpToDate. M. L. K. Tang has received compensation for board membership from Danone Nutricia and the Nestlè Nutrition Institute; has received payment for delivering lectures from Danone Nutricia and the Nestlè 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, Nestlè, 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.

Available online February 10, 2015.

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

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

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 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 childhood[1] with the highest rate of resolution,[8] as well as peanut allergy, which is less likely to resolve[9] 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 resolve[9] 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 at www.jacionline.org). Cells were stained with fluorochrome-conjugated monoclonal or isotype control antibodies in 100-μL 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 package[14] 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 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 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.