

## Targeted DNA methylation profiling reveals epigenetic signatures in peanut allergy

Xiaoying Zhou, ... , Vanitha Sampath, Kari C. Nadeau

*JCI Insight*. 2021. <https://doi.org/10.1172/jci.insight.143058>.

Research

In-Press Preview

Genetics

Immunology

DNA methylation (DNAm) has been shown to play a role in mediating food allergy, however, the mechanism by which it does so is poorly understood. In this study, we used targeted NextGen bisulfite sequencing to evaluate DNAm levels in 125 targeted highly informative genomic regions containing 602 CpG sites on 70 immune-related genes to understand whether DNAm can differentiate peanut allergy (PA) vs non-allergy (NA). We found PA-associated DNAm signatures associated with 12 genes (7 novel to food allergy, 3 associated with Th1/Th2, and 2 associated with innate immunity) as well as DNAm signature combinations with superior diagnostic potential compared to serum peanut specific-IgE for PA vs. NA. Further, we found that following peanut protein stimulation, peripheral blood mononuclear cell (PBMCs) from PA participants showed increased production of cognate cytokines compared to NA participants. The varying responses between PA and NA participants may be associated with the interaction between the modification of DNAm and the interference of environment. Using Euclidean distance analysis, we found that the distances of methylation profile comprising 12 DNAm signatures between PA and NA pairs in monozygotic (MZ) twins were smaller than that in randomly paired genetically unrelated individuals, suggesting that PA related DNAm signatures may be associated with genetic factors.

Find the latest version:

<https://jci.me/143058/pdf>



1   **Targeted DNA methylation profiling reveals epigenetic signatures in peanut allergy**

2  
3   Xiaoying Zhou<sup>a</sup>, Xiaorui Han<sup>a</sup>, Shu-Chen Lyu, Bryan Bunning, Laurie Kost, Iris Chang, Shu  
4   Cao, Vanitha Sampath, Kari C. Nadeau\*

5  
6   Sean N. Parker Center for Allergy & Asthma Research at Stanford University and Division of  
7   Pulmonary, Allergy, and Critical Care Medicine, Stanford University, Stanford, CA, 94305,  
8   USA

9  
10   <sup>a</sup>These authors contributed equally to this work as first authors.

11  
12   \*Corresponding author

13   Kari C. Nadeau, MD, PhD, Director, Sean N. Parker Center for Allergy and Asthma Research at  
14   Stanford University, 240 Pasteur Dr. BMI Rm.1755, Palo Alto, CA 94304, USA. E-mail:  
15   knadeau@stanford.edu.

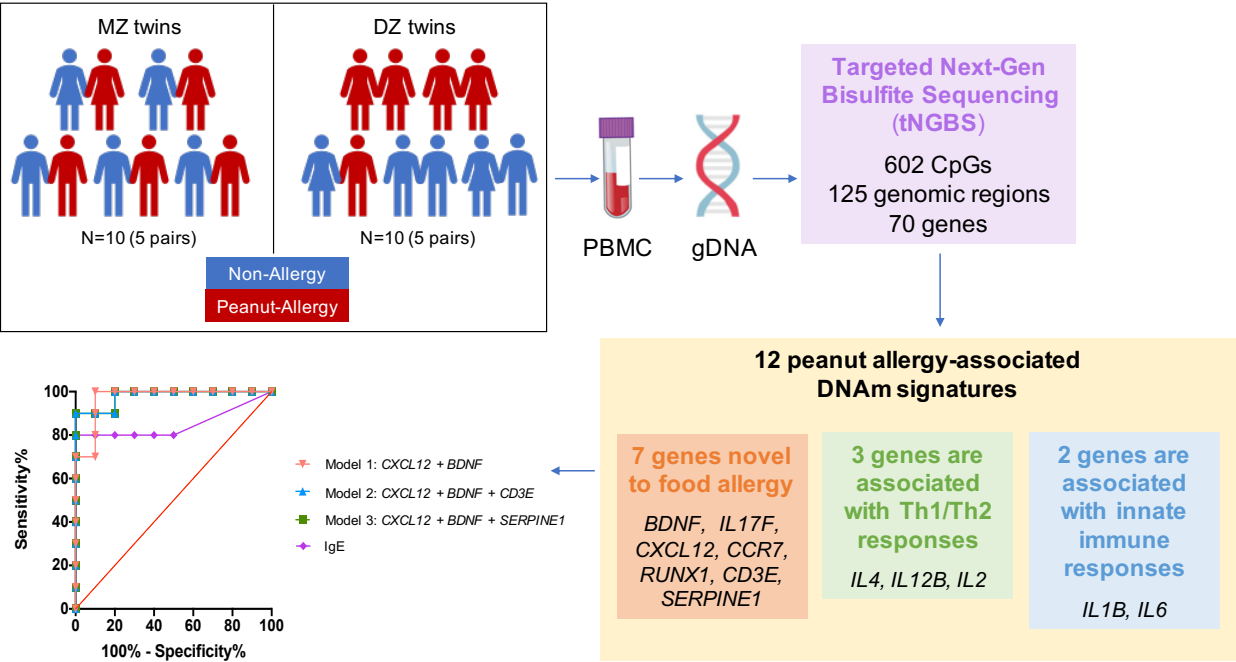
16  
17   **Conflict of interest:** Dr. Nadeau reports grants from National Institute of Allergy and Infectious  
18   Diseases (NIAID), Food Allergy Research & Education (FARE), and End Allergies Together  
19   (EAT); Grant awardee at NIAID, National Heart, Lung, and Blood Institute (NHLBI), and  
20   National Institute of Environmental Health Sciences (NIEHS); Director of FARE and World  
21   Allergy Organization (WAO) Center of Excellence at Stanford; Advisor at Cour Pharma; Co-  
22   founder of Before Brands, Alladapt, and Iggenix; National Scientific Committee member at  
23   Immune Tolerance Network (ITN) and National Institutes of Health (NIH); Research

24 Sponsorship from Nestle; Consultant and Advisory Board Member at Before Brands, Alladapt,  
25 Iggenix, NHLBI, and Probio; Data and Safety Monitoring Board member at NHLBI; and US  
26 patents for basophil testing, multifood immunotherapy and prevention, monoclonal antibody  
27 from plasmoblasts, and device for diagnostics. All other authors have declared that no conflict of  
28 interest exists  
29

## Abstract

DNA methylation (DNAm) has been shown to play a role in mediating food allergy, however, the mechanism by which it does so is poorly understood. In this study, we used targeted NextGen bisulfite sequencing to evaluate DNAm levels in 125 targeted highly informative genomic regions containing 602 CpG sites on 70 immune-related genes to understand whether DNAm can differentiate peanut allergy (PA) vs non-allergy (NA). We found PA-associated DNAm signatures associated with 12 genes (7 potentially novel to food allergy, 3 associated with Th1/Th2, and 2 associated with innate immunity) as well as DNAm signature combinations with superior diagnostic potential compared to serum peanut specific-IgE for PA vs. NA. Further, we found that following peanut protein stimulation, peripheral blood mononuclear cell (PBMCs) from PA participants showed increased production of cognate cytokines compared to NA participants. The varying responses between PA and NA participants may be associated with the interaction between the modification of DNAm and the interference of environment. Using Euclidean distance analysis, we found that the distances of methylation profile comprising 12 DNAm signatures between PA and NA pairs in monozygotic (MZ) twins were smaller than that in randomly paired genetically unrelated individuals, suggesting that PA related DNAm signatures may be associated with genetic factors.

47    **Graphic Abstract**



48

49

## Introduction

Food allergies affect up to 7.6% and 10.8% of children and adults in the United States, respectively (1, 2). Because of the life-threatening potential of anaphylaxis associated with IgE mediated food allergies, food allergy has become a growing clinical and public health problem (1).

Food allergy is a complex immune disease influenced by an interplay of genetic variants, environmental exposures, gene-environment interactions, and epigenetic modifications (3). Epigenetic modifications have been shown as one of the mechanisms used to adapt to environmental exposures and in mediating gene-environment interactions (4). Epigenetic factors, specifically DNA methylation (DNAm), play an important role in the development of food allergy (3, 5). A study by Canani et al. found higher DNAm levels associated with Th1 related genes (*IL10* and *IFNG*) and lower DNAm levels associated with Th2 related genes (*IL4* and *IL5*) in those with cow's milk allergy compared to those tolerant to cow's milk (6). Another study found that peanut oral immunotherapy decreased DNAm of the *FOXP3* gene, a gene associated with tolerogenic responses (7). The study by Martino et al. (8) showed 96 CpG sites that predicted clinical reactivity to food challenge in food-sensitized infants. These sites overlapped with 73 protein-coding genes significantly enriched with the mitogen-activated protein (MAP) kinase canonical pathway. The same group (9) applied genome-wide DNAm to delineate epigenetic modifications in naïve T cells activated by bead-bound anti-CD3/anti-CD28 in egg-allergic participants and nonatopic controls and found a distinct DNAm profile for genes involved in metabolic and immunological regulation of egg allergy.

These studies are encouraging and DNAm signatures offer diagnostic and therapeutic potential for food allergy. However, DNAm signatures for food allergy differ between studies and further research, including exploratory research to identify other potentially novel DNAm signatures, is needed. In this study, we performed targeted NextGen bisulfite sequencing (tNGBS) to evaluate DNAm levels in 125 highly informative genomic regions containing 602 CpG sites for 70 immune-related genes.

We performed tNGBS to evaluate DNAm levels on peripheral blood mononuclear cells (PBMCs) from 10 PA and 10 NA participants, aged 5 to 10 years old. Among these participants, 5 pairs of participants were (PA vs NA) monozygotic (MZ) twin siblings. The other 10 participants included 5 pairs of dizygotic (DZ) twin siblings (2 pairs PA, 2 pairs NA, 1 pair of PA vs NA. The DZ twin siblings were used as randomly paired genetically unrelated individuals. MZ twins discordant for PA are rare in the population and the inclusion of these twin samples improves the statistical power by reducing the amount of genetic and/or environmental variability. We observed that the average CpG methylation levels within the targeted genomic regions for 12 genes showed significant differences between PA and NA participants. PA participants also showed decreased DNAm levels at each of 5 CpG sites in the targeted SERPINE1 region compared to NA participants. Our results indicate PA-associated DNAm signatures at targeted genomic regions are associated with 12 genes, of which 7 are potentially novel to food allergy (brain-derived neurotrophic factor (*BDNF*), *IL17F*, *CXCL12*, *CCR7*, runt-related transcription factor 1 (*RUNX1*), CD3 $\epsilon$  and serine protease inhibitor E1 (*SERPINE1*); 3 are associated with Th1/Th2 responses (*IL4*, *IL12B* and *IL2*) and 2 are associated with innate immune responses (*IL1B* and *IL6*). Using Luminex assay, we found increased secretion of

96 cytokines IL-4, IL-12B, IL-1B, IL-6, CXCL12, and BDNF and secreted protein SERPINE1  
97 following allergen-specific stimulation in PA compared to NA participants. Using stepwise  
98 regression analysis and ROC curve analysis, three combinations of the DNAm signatures from  
99 the initial 12 DNAm signatures were selected with top ranked Akaike Information Criterion  
100 (AIC) and area under the curve (AUC). ROC comparison analysis was performed to compare the  
101 diagnostic performance of these three combinations of DNAm signatures against the existing  
102 diagnostic test-serum peanut-specific IgE. Our results indicate that the combinations of DNAm  
103 signatures from the initial 12 DNAm signatures had superior diagnostic potential compared to  
104 serum peanut-specific IgE for discriminating PA vs. NA. We also evaluated the similarity of 12  
105 DNAm signatures between PA and NA participants in MZ twin sibling pairs and randomly  
106 paired genetically unrelated individuals using Euclidean distance analysis. We found smaller  
107 distances between PA and NA participants in MZ twins compared to randomly paired genetically  
108 unrelated individuals, suggesting that the 12 PA -associated DNAm signatures may be associated  
109 with genetic factors. Altogether, our results demonstrate 12 food allergy-associated DNAm  
110 signatures and differences in protein secretion in responses to allergen-specific stimulation  
111 between PA and NA participants. In addition, our results indicate the diagnostic potential of  
112 DNAm signature combinations for PA and indicate genetic influences on PA-associated DNAm  
113 signatures.

## Results

*DNAm signatures were identified in PA.* To achieve a set of food allergy related DNAm signatures, we selected 125 highly informative genomic regions containing 602 CpG sites for 70 immune-related genes based on a comprehensive literature review, DNAm results from our previous study (10), and the genes important immune system function as per existing immunology panel from EpigenDx Inc (**Supplementary Table 1 and Supplementary Table 2**). tNGBS was performed on PBMCs from 10 NA and 10 PA participants (**Table 1**). First, we compared the average CpG methylation levels within individual targeted genomic regions using the Wilcoxon rank sum test. Significant differences were observed in the average methylation levels in the targeted 12 genomic regions for 12 genes (*IL4*, *IL12B*, *IL2*, *IL17F*, *IL1B*, *IL6*, *CXCL12*, *BDNF*, *CCR7*, *CD3E*, *RUNX1*, and *SERPINE1*) between NA and PA participants (**Figure 1A-B**) (\* $p < 0.05$ , **Supplementary Table 3**). Next, we compared DNAm levels at each CpG site in the above 12 genomic regions for 12 genes between PA and NA participants (**Supplementary Table 3**). Of note, the DNAm levels at each of 5 CpG sites in the targeted genomic region (chr7:101126423-101126457) for the gene *SERPINE1* were significant decreased in PA, compared to NA participants (FDR adjusted \* $p$ -value  $< 0.1$ ) (**Figure 2**).

The principal component analysis (PCA) of the above 12 DNAm signatures showed the first principal components accounted for 94.68% of the dataset variation, and separated PA from NA participants (**Figure 1C**). To further confirm that the centroid and dispersion of two distinct clusters were different between PA and NA samples, we carried out a permutational multivariate analysis of variance (PERMANOVA) for PCA. This test examines the contribution of variables

to the separation of the data in multiple dimensional spaces and showed a significant p-value between two groups (\*\*p<0.001).

We then compared the composition of major immune cell populations between NA and PA participants who are genetical unrelated (demographic characteristics shown in Supplementary Table 4) using flow cytometry and found no differences in major immune cell types including CD3, CD4, CD8, B cells, NK cells, monocytes and dendritic cells (DCs) between NA and PA participants (Supplementary Figure 1). Since there were no statistical differences in the cell phenotypes between groups, it is unlikely that the statistically significant differences in methylation profiles between the groups can be attributed in a major way to cell heterogeneity.

*The secretion levels of the proteins encoded by the genes overlapping DNAm signatures were increased from PBMCs in PA compared to NA participants after peanut stimulation.* Of the above 12 genes showing significant differences in DNAm levels between PA and NA participants, there were 8 cytokine genes (*IL4*, *IL12B*, *IL2*, *IL17F*, *IL1B*, *IL6*, *CXCL12* and *BDNF*) and 1 gene encoding secreted protein SERPINE1. To determine if these pre-existing differential DNAm signatures were associated with expression of their cognate proteins, the PBMCs from NA and PA individuals were incubated either with or without peanut protein. After a 3-day incubation, the secretion levels of these cytokines and SERPINE1 from PBMCs in supernatants were measured using a Luminex based assay. We found that compared to NA participants, peanut protein stimulated PBMCs from PA participants showed increased production of 7 cytokines (IL-4, IL-12 $\beta$ , IL-2, IL-1 $\beta$ , IL-6, CXCL12, and BDNF) and SERPINE1 protein (**Figure 3**). These results suggest that the differences in protein secretions in

response to specific food allergens between PA and NA may be associated with DNAm modifications and environmental interactions.

*The performance of combinations of DNAm signatures for discrimination of PA vs. NA participants was superior to serum peanut specific IgE.* Next, we performed ROC analysis to calculate the AUC of 12 DNAm signatures. The AUC for 12 individual ROC curves varied from 0.77 to 0.845 for PA vs. NA. Diagnostic sensitivity for individual DNAm signatures at 90 % specificity ranged from 10 to 70 % for PA vs. NA (**Figure 4A**). On the basis of the assumption that each DNAm signature can be considered as a diagnostic test, we attempted to optimize DNAm signature combinations from the initial 12 DNAm signatures. We applied a stepwise (step-up) regression analysis and selected 3 models that have top ranked AIC and AUC [model 1: + 82.518 (*CXCL12*) - 2.5 (*BDNF*) +107.287; model 2: +134.199 (*CXCL12*) - 2.230 (*BDNF*) + 12.416 (*CD3E*) -10.798; model 3: +73.62 (*CXCL12*) - 1.543 (*BDNF*) -1.267 (*SERPINE1*) +85.039] (**Table 2**). To determine whether these 3 models with the combination of DNAm signatures could exceed the diagnostic potential of existing diagnostic tests, we compared the above ROC curves against peanut-specific IgE levels in serum (**Figure 4B**). The 3 models with the combination of the DNAm signatures had the AUC of 0.97 for model 1(*CXCL12*+*BDNF*), 0.98 for model 2 (*CXCL12*+*BDNF*+*CD3E* ) and 0.98 for model 3 (*CXCL12*+*BDNF*+*SERPINE1*), respectively, compared to 0.85 for serum peanut-specific IgE (**Figure 4B and Table 2**). The sensitivities at 90 % specificity for discrimination of PA vs. NA for 3 models with the combination of the selected DNAm signatures were 70% (model 1), 90% (model 2) and 90% (model 3), respectively. These results suggest that DNAm signatures can be combined to produce highly clinically sensitive and specific DNAm panels, and the above 3 combinations of

DNAm signatures have superior diagnostic potential compared to peanut-specific IgE in serum (Figure 4B and Table 2).

*PA-associated DNAm signatures were genetically influenced.* Among the 10 PA and 10 NA participants, 5 pairs were PA discordant MZ twin siblings aged 5 to 10 years old (Table 1). Since MZ twin pairs are widely regarded as genetically “identical”, and young twins usually share similar environmental backgrounds, the discordant PA in MZ twin pairs is presumed to result from different epigenetic mechanisms. The epigenetic drift during the lifetime of MZ twin pairs has been suggested to arise from differing environmental histories (11, 12). In contrast, the contributions of both environmental in utero and underlying genetic factors to epigenetic profile have been suggested in neonatal epigenome for MZ twins with discordant phenotypes (13). To estimate relative contributions of the genetic factors to the epigenetic profile comprising 12 DNAm signatures, we performed a Euclidean distance-based analysis to compare the similarity of food allergy-associated epigenetic profile in the MZ twins and the randomly paired genetically unrelated individuals. The median of distances between PA and NA pairs in the MZ twins decreased by 48.28% compared to that in the randomly paired genetically unrelated individuals, although the results did not reach statistical significance ( $p=0.14$ ) (Figure 5A-B). The smaller distance between PA and NA participants in MZ twins suggests that PA related DNAm signatures may be associated with genetic factors. Further studies are needed to test if this is significant. We also observed the smaller distance between MZ twins and genetically unrelated individuals in PA participants than in NA participants, which is consistent with our results of PCA showing two distinct clusters from PA and NA participants, and suggests epigenetic

205 similarities for the above 12 DNAm signatures in PA, but not NA, regardless of whether the PA  
206 participants were MZ twin siblings or genetically unrelated individuals (**Figure 5A and 5C**).

207

## Discussion

This study presents potentially novel findings on the loci differentially methylated for food allergy. The PA-associated DNAm signatures at the targeted genomic regions for 12 genes were identified by comparing the DNAm levels in 125 targeted genomic regions containing 602 CpG sites for 70 immune-related genes between PA and NA participants. These DNAm signatures for 12 genes include the genes associated with Th1/Th2 differentiation (*IL4*, *IL12B* and *IL2*), innate immunity (*IL1B*, *IL6*), and those involved with immune regulation, but not specifically for food allergy, such as *BDNF*, *IL17F*, *CXCL12*, *CCR7*, *RUNX1*, *CD3E* and *SERPINE1*. Incubating PBMCs from PA participants with peanut protein resulted in the increased secretion of IL-4, IL-12 $\beta$ , IL-2, IL-1 $\beta$ , IL-6, CXCL12, BDNF and SERPINE1, compared to NA participants, which may suggest that the differences in PBMC responses on stimulation with specific food allergens between PA and NA groups are associated with DNAm changes and environmental interactions. In addition, our data show that three combinations of DNAm signatures from the 12 PA-associated DNAm signatures have superior diagnostic performance against serum peanut-specific IgE for discriminating PA vs. NA. Our results also demonstrate that PA-associated 12 DNAm signatures are influenced by genetic factors.

It has long been understood that IgE mediated food allergy results from a Th2 immune response of the adaptive immune system to protein antigens associated with specific foods (14). Therefore, the skewing of naïve CD4<sup>+</sup> T cell differentiation into Th1 or Th2 effector cells, driven by the cytokine environment, is critical to the development of food allergy. Cytokine IL-12 is mainly produced by phagocytic cells (monocytes, macrophages, neutrophils and dendritic cells) and has been shown to drive naïve T cells to differentiate into Th1 cells. IL-4 (produced by Th2

cells) is a major cytokine driving the differentiation of naive T cells into a Th2 subset (15). In addition, studies have suggested that IL-2 also has a role in facilitating Th2 differentiation (16). Low doses of IL-2 induced regulatory T cell (Treg) expansion provides protection against clinical manifestation of food allergy by Treg dependent modification of Th1/Th2 balance (17). Consistent with previous studies showing the epigenetic regulation of Th1/Th2 differentiation (18), compared to NA participants, PA participants showed a decrease in DNAm levels at the targeted genomic regions of *IL4* and *IL2*, and an increase in DNAm levels at the targeted genomic region of *IL12B*.

The pro-inflammatory cytokines IL-1 $\beta$  and IL-6, mainly produced by different innate immune cells, were significantly increased in food allergic compared to NA participants when stimulated by lipopolysaccharides (LPS) (19) or by specific allergens as shown in Figure 2. These observations suggest that different innate immune response occur in food allergic vs non-allergic participants. A study demonstrated that exposure to LPS (tolerance immunity) or bacterial  $\beta$ -glucan (trained immunity) induces epigenetic changes in monocytes. These re-programmed epigenetic landscapes of innate immune cells determine the capacity for developing a “memory” in response to exogenous exposure (20). In our study, we also observed significant decreases in DNAm levels for the gene *IL1B* and *IL6* in PA compared to NA individuals, suggesting a link between epigenetic regulation of the innate immune system and food allergy.

It is becoming increasingly clear that immune cells do not act alone and that cross talk and reciprocal regulation between neural and immune systems are essential in the pathophysiology of allergic diseases including allergic asthma, atopic dermatitis, and food allergies (21, 22). Both immune and neural cells detect and respond to environmental threats and harmful stimuli

including allergens. Proinflammatory mediators, such as cytokines or chemokines, mediate allergic responses and also directly activate sensory neurons that regulate itch, sneezing, bronchoconstriction, and alterations in gastrointestinal motility (23). However, the mediators between neuronal and immune cells and their role in mediating allergic responses still remain unclear. BDNF is a member of the neurotrophin family, which are known to be related to canonical nerve growth factor and neurogenic inflammation (24, 25). Recently, neurotrophins were also found to be produced continuously during allergic inflammation (26). Increased expression of BDNF has been observed in severe asthma, bronchial hyperresponsiveness and inflammation (26), but to our knowledge, it has not been observed in food allergy. Our results, show that DNAm levels of *BDNF* are significantly decreased in PA compared to NA participants, suggesting that epigenetic modifications of the BDNF gene is associated with food allergy. In addition, the combinations of DNAm signatures of *BDNF* and *CXCL12* had the most superior diagnostic potential among all other combinations selected compared to serum peanut-specific IgE for discriminating PA vs. NA.

Of note, the average DNAm level in the targeted *SERPINE1* genomic region (chr7:101126423-101126457) and the DNAm levels at each of 5 CpG sites in this targeted region were significantly decreased in PA compared to NA participants. In mice, following intranasal OVA challenge, wild type mice skewed to a Th2 immune response while *SERPINE1*<sup>-/-</sup> mice skewed from a Th2 to a Th1 immune response (27). In house dust mite (HDM) allergic asthma patients, SNP (rs1799768) in the *SERPINE1* gene was associated with bronchial reactivity to histamine and IgE response (28). Interestingly, the gene *SERPINE1* 5'-upstream promoter region (chr7:101126423-101126457, shown in the Figure 1C) contains the allergic disease-associated SNP (rs1799768, chr7: 101126426) (28). It has been proposed that epigenetic modification may

mediate the effect of genetic variants on the development of food allergy (3). For example, the DNAm of HLA-DQB1 and HLA-DRB1 genes were implicated in the mediation of the association between SNPs in (HLA)-DQ and DR regions and PA (29). Consistent with this possibility, the observation on the genomic location of allergy-associated *SERPINE1* SNP suggests that the DNAm in the above targeted regions for *SERPINE1* gene might act as a mediator of the association between genetic variation and allergic disease.

IL-17F and IL-17A are related homodimeric proteins of the IL-17 family and are produced by Th17 cells. Both 17F and IL-17A cytokines have been implicated in allergic inflammation (21) and inflammation resulting from mucosal immunity or autoimmunity (30). It has been suggested that IL-17A plays a role in the regulation of food allergy and is a potential biomarker of tolerance to food allergens (31); however, the role of IL-17F in food allergy is not well understood. Our results, suggest that epigenetic modifications of the *IL17F* gene is associated with food allergy.

Increased levels of the chemokine CXCL12 has been found in the bronchoalveolar lavage of patients with asthma (32). Overexpression of the chemokine receptor CCR7 in dendritic cells (DC) is thought to mediate DC lymph node migration and promote the development of allergic responses (33). Genome-wide association studies (GWAS) showed that the SNPs in *RUNX1* are associated with airway responsiveness in asthmatic children (34) and the *RUNX1* transcription factor has been identified as a molecular link in TGF $\beta$  induced FOXP3 expression in inducible Treg (35). It has been proposed that the signals transduced by CD3 $\epsilon$  contribute quantitatively to TCR signaling, and that CD3 $\epsilon$  signals suggest a potential role in the generation and/or survival of

mature T cells(36). Here, our results, show the association between DNAm levels with food allergy for *CXCL12*, *CCR7*, *RUNXI* and *CD3E*.

DNAm is most commonly associated with downregulation of gene expression especially when the hypermethylation is in the promotor region of a gene. In our study, *IL4*, *IL2*, *IL1B*, *IL6*, *BDNF* and *SERPINE1* showed lower methylation level and higher protein level in the PA compared to NA participants. However, the PA patients have increased DNAm at 5-upstream of *IL12B* and *CXCL12*, which is associated with increased IL-12 $\beta$  and CXCL12 protein expression compared to NA participants. It should be noted that, although gene silencing by promoter hypermethylation seems to be the most likely mode of action, there is growing evidence of a more complex view on the effect of DNAm in various contexts (37-41). In particular, for genes that become methylated, the associated expression level can be unaffected or even upregulated in some cases (40), suggesting a more diverse mechanism of epigenetic regulation. Such additional complexity could have important implications for understanding allergic disease but has not been studied at a genome-wide scale.

Among the 10 PA and 10 NA participants in this study, 7 are allergic to tree nuts, 4 have asthma and 6 have atopic dermatitis. To examine whether similar DNAm signatures are presented in other allergic disease, we applied the Wilcoxon rank sum test to compare the average CpG methylation levels within individual targeted genomic regions using our original data (total 125 targeted genomic regions covering 602 CpG sites) between 6 atopic dermatitis vs 14 non-dermatitis participants, 4 asthma vs 16 non-asthma participants, and 7 tree nut allergic vs 13 non-tree nut allergic participants. The differential methylated genomic regions related to atopic

dermatitis, asthma or tree nut allergy are presented in Supplementary Figures 2-4. Among the 10 differentially methylated genomic regions associated with atopic dermatitis (Supplementary Figure 2), none of them overlap with the 12 PA-associated differentially methylated genomic region and all of the 10 genes have been reported to be associated with atopic dermatitis in previous studies (42-51). Of note, the differential DNAm in the genomic region of the filaggrin gene (*FLG*), which plays an important role in the pathogenesis of atopic dermatitis and allergic disease (42, 52), is identified in our comparison analysis between non-atopic dermatitis and atopic dermatitis participants. Among 4 differentially methylated genomic regions associated with asthma (Supplementary Figure 3), 2 of them (*IL12B*, *IL2*) overlap with the 12 PA associated differentially methylated genomic regions and all of the 4 genes have been reported to be associated with asthma in previous studies (53-57). Among 9 differentially methylated genomic regions associated with tree nut allergy (Supplementary Figure 4), only one (*CXCL12*) overlaps with the 12 PA associated differential methylated genomic regions and 6 of the 9 genes (*IL10*, *IKZF2*, *CCL5*, *IL21*, *CCR9*, *IL-33*) have been reported to be associated with food allergy in previous studies (58-63). All of the above results support that the 12 differentially methylated genomic regions are to some extent specific to PA.

Oral food challenges (OFCs) are the gold standard for diagnosis of food allergy, however OFCs are associated with risk of allergic reaction and therefore need to be performed in clinics with trained staff, thus limiting widespread use (64). The study by Martino and colleagues (8) showed DNAm signatures at 96 CpG sites can predict food challenge outcomes by comparing the differences of DNAm between food allergic and food sensitization groups. Their results showed that the 73 genes overlapped by these 96 CpG sites were enriched in the sole MAP kinase

canonical pathway but were not involved in the well-known food allergy-associated Th1/Th2 pathway. In addition, these clinically relevant biomarkers have not been widely used in clinic diagnosis. We also compared the diagnostic performance for these 12 DNAm signatures for PA against peanut specific IgE in serum. We found that diagnostic sensitivity for peanut-specific IgE at 90% specificity is 80%, which is superior to the sensitivity of individual DNAm signatures at 90 % specificity, which ranged from 10 to 70 % for PA vs. NA, respectively. To select the optimal combination of DNAm signatures from the initial 12 DNAm signatures, we further performed stepwise regression analysis and ROC curve analysis, and our pilot exploratory results indicate that the combinations of DNAm signatures selected from the initial 12 DNAm signatures had superior diagnostic potential compared to serum peanut-specific IgE for discriminating PA vs. NA.

The birth rate for MZ twins is about 0.3% of the world population, and a previous study showed that among 14 pairs of MZ twins, only 5 of them were discordant for peanut allergy (65). The young allergy-discordant MZ twin participants used in this study have nearly perfect controls of covariates such as age, sex, genetic and even environmental factors, which increase the rigor and reproducibility of our epigenetic association studies. It increases the power estimation over ordinary case-control designs by minimizing confounding genetic and environmental factors (66, 67). In addition, to avoid reporting findings caused by natural variability within twin pairs, we compared our data with a study that investigated genome-wide DNAm variability in adolescent MZ twins followed since birth (68). This study showed that probes with the highest within-pair differences in DNAm were enriched in gene ontologies related to development and cell growth. The PA associated 12 differential methylated gene regions identified in our study do not overlap

with the hypervariable genes across MZ twins in previous studies (68). The lack of overlap between the two studies, in gene variability within twin pair differences in DNAm, reduces the likelihood that our findings were due to general within-pair variability and increases our confidence that these data represent true methylation difference associated with PA. The development of these combinations of DNAm signatures for diagnosis of PA need further verification in detailed follow-up studies with large sample size.

Overall, our results reveal that the food allergy associated DNAm signatures suggest epigenetic modifications could discern PA vs NA individuals, and that these DNAm signatures could be potentially used for diagnostics or future medical research in not only PA but also food allergy. Several DNAm signatures highlighted in the current study have been associated with PA, suggesting that these genes may have an important role in food allergy with potential as biomarkers and for targeted therapy. Further validation with a larger cohort and further functional studies are warranted to enable further understanding of the molecular mechanisms underpinning food allergy. The potentially novel PA associated DNAm genes, such as BDNF (neurotrophin) and SERPINE1 (serine protease inhibitor) suggest an additional direction of research for deciphering the molecular basis of PA. Furthermore, a high similarity between PA and NA participants were observed in MZ twins compared to randomly paired genetically unrelated individuals, indicating that the PA-associated 12 DNAm signatures were genetically influenced.

## **Methods**

### **Study Participants.**

PA discordant MZ twin siblings, non-twin PA pediatric participants, and non-twin NA pediatric participants were recruited at the Sean N. Parker Center for Allergy and Asthma Research at Stanford University. Patient demographics, food allergy history, atopic history and peanut-specific IgE are summarized in Table 1. PA was confirmed by a food challenge by a certified allergy specialist. Blood specimens were drawn before food challenges were performed and no participant was taking any medications (i.e. steroids, valproic acid, folic acid, DNA intercalating agents, methotrexate or DNA methyltransferase inhibitors) that could have affected the epigenetics.

### **Collection and processing of blood specimens**

Peripheral blood mononuclear cells (PBMCs) and plasma were isolated from blood samples by density gradient centrifugation over Ficoll-Paque, cryopreserved PBMCs in 10% dimethyl sulfoxide in fetal calf serum and stored in liquid nitrogen. Plasma was stored at -80 °C.

### **Targeted next-generation bisulfite sequencing (tNGBS)**

tNGBS was performed on PBMCs from 10 PA and 10 NA participants by EpigenDx. Total 125 targeted genomic regions containing 602 CpG sites for 70 genes were analyzed using tNGBS. Targeted bisulfite deep sequencing PCR products were purified using QIAquick PCR purification kit (QIAGEN). Ion Torrent deep sequencing libraries were constructed from bisulfite-converted DNA using the KAPA Library Preparation Kit (Kapa Biosystems), quantified

using the QIAxcel Advanced System (QIAGEN), templated using the Ion PGM Template OT2 200 kit (Thermo Fisher) and sequenced using the Ion PGM™ Sequencing Hi-Q™ OT2 Kit with the Ion 314™ Chip Kit v2 on an Ion PGM™ System (Thermo Fisher Scientific, MA), which generated non-directional ~200 nt length reads, 1500–7500 reads per library in the fastq format. FASTQ files from the Ion PGM™ System were filtered and aligned to the human genome assembly hg38 using Bismark Bisulfite Mapper v0.15.0 (Babraham Bioinformatics, UK) with the Bowtie 2 alignment algorithm. Methylation levels (% Methylation) were calculated in Bismark by dividing the number of cytosine converted (C versus T) reads by the number of total reads, considering all CpG sites covered by a minimum of 30 total reads. The DNAm level at each of 602 CpG sites were reported by EpigenDx.

#### **In vitro stimulation**

The peanut proteins added to cell culture were derived from peanut flour used for double-blind placebo-controlled food challenges (DBPCFCs) in the clinic. The peanut flour was dissolved in PBS and sterilized by filtration. The peanut protein concentration was determined by BCA Protein Assay (Pierce, Rockford, IL). The endotoxin level of peanut protein was assessed by fluorescence-based rFC assay (Indoor Biotechnologies) and the endotoxin level of peanut protein exhibited in cell culture was 0.05 EU/ml.

After overnight resting of thawed PBMCs, cells were cultured in complete RPMI medium (RPMI1640 medium, glutamine, 5% human serum, 1% penicillin/streptomycin), and in the presence or absence of peanut protein at a final concentration of 100 µg/ml for 3 days. For each condition of this experiment, PBMCs were cultured at  $5 \times 10^5$  cells per 200ul for 3 days, after which the supernatants were harvested and stored at -80°C.

435

## 436 **Cytokine assays**

437 The secretion levels of cytokines or chemokines from PBMCs in supernatants were measured  
438 using 62-multiplex assay on the Luminex 200 IS system (Affymetrix) performed by Stanford  
439 Human Immune Monitoring Center (HIMC). All samples were tested in duplicate wells. Data  
440 were analyzed using MasterPlex software (Hitachi Software Engineering America Ltd., MiraiBio  
441 Group), and the average of two median fluorescence intensity (MFI) values for each sample for  
442 each analyte were reported by Stanford HIMC. Then, the ratios were calculated by dividing the  
443 average MFI of each analyte for each sample by the average MFI of each analyte for complete  
444 RPMI medium control. These ratios were used to present the secretion level of each cytokines or  
445 chemokines from PBMCs for each sample.

446

## 447 **Flow Cytometry**

448 PBMCs from 10 NA and 10 PA participants were stained with the lineage markers listed in  
449 Supplementary Table 5 to examine the proportion of major immune cell subtypes in total  
450 PBMCs. The PBMCs (around  $1 \times 10^6$  cells per sample) were incubated in 100ul volume with  
451 Human TruStain FcX™ (Biolegend), cell surface antibodies (in Supplementary Table 5) and  
452 viability dyes (Aqua, ThermoFisher) for 30min on ice, and followed by washing two times with  
453 FACS buffer (PBS with 0.25% BSA and 1mM EDTA). All samples were run on LSRII flow  
454 cytometer and analyzed using FlowJo Version 10.6.0 software.

455

## 456 **Statistics**

The analysis was conducted using the statistical programming language R (version R 3.6.2). The differences in average methylation across all CpGs in each targeted genomic region and the differences in DNAm level at each CpG site in individual targeted genomic region, were compared between PA and NA participants using Wilcoxon rank sum test (non-parametric unpaired test, two sided). A  $p$  value less than 0.05 was considered significant. Using the Benjamini-Hochberg procedure with a false discovery rate of 0.1, the assay-wise multiple testing correction was further performed across each of the CpG sites in individual targeted genomic region (or individual independent assay). An adjusted- $p$  value less than 0.1 was considered significant.

The mixed-effects logistic regression correcting for the dependent participants, such as MZ twin samples was conducted using the “glmer” function in the “lme4” R package (version 1.1-23). The DNAm signatures used for regression and creating the ROC curve are the average methylation level across the CpG sites within a targeted genomic region, which was detected by tNGBS by one pair of primers. The step-up regression was performed to select the combination of variables in terms of the AIC value and AUC value. Three models with the combinations of two or three methylation signatures have the both top 3 lowest AIC value and top 3 highest AUC value. Equivalent ROC curves were obtained from mixed-effects logistic regression. ROC analysis was performed using Prism 8 (GraphPad).

Principal component analysis (PCA) was then conducted to visualize the DNAm signatures between NA and PA groups using pcaMethods package in R. Permutational multivariate analysis of variance (PERMANOVA) was applied to examine the contribution of variables to the

480 separation of the data in multiple dimensional space using “adonis2” function in the “vegan”  
481 package in R. The distance function in R software was employed for Euclidean distances  
482 analysis. All dot plots overlaid with boxplots or line connections were compiled with ggplot2  
483 package in R.

484

#### 485 **Study approval**

486 The study was approved by the Institutional Review Board of Stanford University and registered  
487 at Clinicaltrials.gov (NCT01613885). All participants or their caregivers provided written  
488 informed consent.

489

**Author contributions**

K.C.N conceived the study. X.Z., S.C.L. performed experiments and collected data. B.B. and L.K. contributed to recruitment of study participants. I.C. contributed to processed samples for this study. X.Z., X.H., S.C.L and K.C.N. contributed to experimental design, data analysis and data interpretation. S.C. assisted with statistical analyses. X.Z., X.H., V.S. and K.C.N. wrote the paper. X.Z. and X.H contributed equally to this research as co-first author. The order of appearance of the co-first authors was based on the timeline of contributions to the work.

**Conflict of interest:** Dr. Nadeau reports grants from National Institute of Allergy and Infectious Diseases (NIAID), Food Allergy Research & Education (FARE), and End Allergies Together (EAT); Grant awardee at NIAID, National Heart, Lung, and Blood Institute (NHLBI), and National Institute of Environmental Health Sciences (NIEHS); Director of FARE and World Allergy Organization (WAO) Center of Excellence at Stanford; Advisor at Cour Pharma; Co-founder of Before Brands, Alladapt, and Iggenix; National Scientific Committee member at Immune Tolerance Network (ITN) and National Institutes of Health (NIH); Research Sponsorship from Nestle; Consultant and Advisory Board Member at Before Brands, Alladapt, Iggenix, NHLBI, and Probio; Data and Safety Monitoring Board member at NHLBI; and US patents for basophil testing, multifood immunotherapy and prevention, monoclonal antibody from plasmoblasts, and device for diagnostics. All other authors declare no conflict of interest.

511    **Acknowledgments**

512    We acknowledge Holden T Maecker for his advice on the analysis of Luminex data.

513    This work was supported by grants from NIAID (R01AI140134), NHLBI (R01HL118612), the

514    Sean N. Parker Center for Allergy and Asthma Research at Stanford University and The Stanford

515    Institute for Immunity, Transplantation and Infection (ITI) Seed Grant.

516

517 **References:**

- 518 1. Gupta RS, Warren CM, Smith BM, Jiang J, Blumenstock JA, Davis MM, et al. Prevalence and  
519 Severity of Food Allergies Among US Adults. *JAMA Netw Open*. 2019;2(1):e185630.
- 520 2. Gupta RS, Warren CM, Smith BM, Blumenstock JA, Jiang J, Davis MM, et al. The Public Health  
521 Impact of Parent-Reported Childhood Food Allergies in the United States. *Pediatrics*.  
522 2018;142(6).
- 523 3. Bar-El Dadon S, and Reifen R. *Principles of Nutrigenetics and Nutrigenomics*. Elsevier; 2020:403-  
524 9.
- 525 4. Hong X, and Wang X. Epigenetics and development of food allergy (FA) in early childhood. *Curr*  
526 *Allergy Asthma Rep*. 2014;14(9):460.
- 527 5. Kabesch M, and Tost J. Recent findings in the genetics and epigenetics of asthma and allergy.  
528 *Semin Immunopathol*. 2020;42(1):43-60.
- 529 6. Berni Canani R, Paparo L, Nocerino R, Cosenza L, Pezzella V, Di Costanzo M, et al. Differences in  
530 DNA methylation profile of Th1 and Th2 cytokine genes are associated with tolerance  
531 acquisition in children with IgE-mediated cow's milk allergy. *Clin Epigenetics*. 2015;7:38.
- 532 7. Syed A, Garcia MA, Lyu SC, Bucayu R, Kohli A, Ishida S, et al. Peanut oral immunotherapy results  
533 in increased antigen-induced regulatory T-cell function and hypomethylation of forkhead box  
534 protein 3 (FOXP3). *J Allergy Clin Immunol*. 2014;133(2):500-10.
- 535 8. Martino D, Dang T, Sexton-Oates A, Prescott S, Tang ML, Dharmage S, et al. Blood DNA  
536 methylation biomarkers predict clinical reactivity in food-sensitized infants. *J Allergy Clin*  
537 *Immunol*. 2015;135(5):1319-28 e1-12.
- 538 9. Martino D, Neeland M, Dang T, Cobb J, Ellis J, Barnett A, et al. Epigenetic dysregulation of naive  
539 CD4+ T-cell activation genes in childhood food allergy. *Nat Commun*. 2018;9(1):3308.
- 540 10. Prunicki M, Stell L, Dinakarbandian D, de Planell-Saguer M, Lucas RW, Hammond SK, et al.  
541 Exposure to NO<sub>2</sub>, CO, and PM<sub>2.5</sub> is linked to regional DNA methylation differences in asthma.  
542 *Clin Epigenetics*. 2018;10:2.
- 543 11. Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, et al. Epigenetic differences arise  
544 during the lifetime of monozygotic twins. *Proc Natl Acad Sci U S A*. 2005;102(30):10604-9.
- 545 12. Cheung P, Vallania F, Warsinske HC, Donato M, Schaffert S, Chang SE, et al. Single-Cell  
546 Chromatin Modification Profiling Reveals Increased Epigenetic Variations with Aging. *Cell*.  
547 2018;173(6):1385-97 e14.
- 548 13. Ollikainen M, Smith KR, Joo EJ, Ng HK, Andronikos R, Novakovic B, et al. DNA methylation  
549 analysis of multiple tissues from newborn twins reveals both genetic and intrauterine  
550 components to variation in the human neonatal epigenome. *Hum Mol Genet*. 2010;19(21):4176-  
551 88.
- 552 14. Yu W, Freeland DMH, and Nadeau KC. Food allergy: immune mechanisms, diagnosis and  
553 immunotherapy. *Nat Rev Immunol*. 2016;16(12):751-65.
- 554 15. Dong C, and Flavell RA. Cell fate decision: T-helper 1 and 2 subsets in immune responses.  
555 *Arthritis Res*. 2000;2(3):179-88.
- 556 16. Le Gros G, Ben-Sasson SZ, Seder R, Finkelman FD, and Paul WE. Generation of interleukin 4 (IL-  
557 4)-producing cells in vivo and in vitro: IL-2 and IL-4 are required for in vitro generation of IL-4-  
558 producing cells. *J Immunol*. 2008;181(5):2943-51.
- 559 17. Bonnet B, Vigneron J, Levacher B, Vazquez T, Pitoiset F, Brimaud F, et al. Low-Dose IL-2 Induces  
560 Regulatory T Cell-Mediated Control of Experimental Food Allergy. *J Immunol*. 2016;197(1):188-  
561 98.
- 562 18. Hong X, and Wang X. Early life precursors, epigenetics, and the development of food allergy.  
563 *Semin Immunopathol*. 2012;34(5):655-69.

19. Neeland MR, Koplin JJ, Dang TD, Dharmage SC, Tang ML, Prescott SL, et al. Early life innate immune signatures of persistent food allergy. *J Allergy Clin Immunol*. 2018;142(3):857-64 e3.
20. Novakovic B, Habibi E, Wang SY, Arts RJW, Davar R, Megchelenbrink W, et al. beta-Glucan Reverses the Epigenetic State of LPS-Induced Immunological Tolerance. *Cell*. 2016;167(5):1354-68 e14.
21. Wang YH, and Wills-Karp M. The potential role of interleukin-17 in severe asthma. *Curr Allergy Asthma Rep*. 2011;11(5):388-94.
22. Veres TZ, Rochlitzer S, and Braun A. The role of neuro-immune cross-talk in the regulation of inflammation and remodelling in asthma. *Pharmacol Ther*. 2009;122(2):203-14.
23. Nassenstein C, Krasteva-Christ G, and Renz H. New aspects of neuroinflammation and neuroimmune crosstalk in the airways. *J Allergy Clin Immunol*. 2018;142(5):1415-22.
24. Lindsay RM, and Harmar AJ. Nerve growth factor regulates expression of neuropeptide genes in adult sensory neurons. *Nature*. 1989;337(6205):362-4.
25. Nadel JA. Neutral endopeptidase modulates neurogenic inflammation. *Eur Respir J*. 1991;4(6):745-54.
26. Nockher WA, and Renz H. Neurotrophins and asthma: novel insight into neuroimmune interaction. *J Allergy Clin Immunol*. 2006;117(1):67-71.
27. Sejima T, Madoiwa S, Mimuro J, Sugo T, Okada K, Ueshima S, et al. Protection of plasminogen activator inhibitor-1-deficient mice from nasal allergy. *J Immunol*. 2005;174(12):8135-43.
28. Pampuch A, Kowal K, Bodzenta-Lukaszyk A, Di Castelnuovo A, Chyczewski L, Donati MB, et al. The -675 4G/5G plasminogen activator inhibitor-1 promoter polymorphism in house dust mite-sensitive allergic asthma patients. *Allergy*. 2006;61(2):234-8.
29. Hong X, Hao K, Ladd-Acosta C, Hansen KD, Tsai HJ, Liu X, et al. Genome-wide association study identifies peanut allergy-specific loci and evidence of epigenetic mediation in US children. *Nat Commun*. 2015;6:6304.
30. Dubin PJ, and Kolls JK. Interleukin-17A and interleukin-17F: a tale of two cytokines. *Immunity*. 2009;30(1):9-11.
31. Dhuban KB, d'Hennezel E, Ben-Shoshan M, McCusker C, Clarke A, Fiset P, et al. Altered T helper 17 responses in children with food allergy. *Int Arch Allergy Immunol*. 2013;162(4):318-22.
32. Chen H, Xu X, Teng J, Cheng S, Bunjhoo H, Cao Y, et al. CXCR4 inhibitor attenuates allergen-induced lung inflammation by down-regulating MMP-9 and ERK1/2. *Int J Clin Exp Pathol*. 2015;8(6):6700-7.
33. Gill MA. The role of dendritic cells in asthma. *J Allergy Clin Immunol*. 2012;129(4):889-901.
34. Haley KJ, Lasky-Su J, Manoli SE, Smith LA, Shahsafaei A, Weiss ST, et al. RUNX transcription factors: association with pediatric asthma and modulated by maternal smoking. *Am J Physiol Lung Cell Mol Physiol*. 2011;301(5):L693-701.
35. Klunker S, Chong MM, Mantel PY, Palomares O, Bassin C, Ziegler M, et al. Transcription factors RUNX1 and RUNX3 in the induction and suppressive function of Foxp3+ inducible regulatory T cells. *J Exp Med*. 2009;206(12):2701-15.
36. Sommers CL, Dejarnette JB, Huang K, Lee J, El-Khoury D, Shores EW, et al. Function of CD3 epsilon-mediated signals in T cell development. *J Exp Med*. 2000;192(6):913-19.
37. Marzese DM, and Hoon DS. Emerging technologies for studying DNA methylation for the molecular diagnosis of cancer. *Expert Rev Mol Diagn*. 2015;15(5):647-64.
38. Medvedeva YA, Khamis AM, Kulakovskiy IV, Ba-Alawi W, Bhuyan MS, Kawaji H, et al. Effects of cytosine methylation on transcription factor binding sites. *BMC Genomics*. 2014;15:119.
39. Yang M, and Park JY. DNA methylation in promoter region as biomarkers in prostate cancer. *Methods Mol Biol*. 2012;863:67-109.

40. Wan J, Oliver VF, Wang G, Zhu H, Zack DJ, Merbs SL, et al. Characterization of tissue-specific differential DNA methylation suggests distinct modes of positive and negative gene expression regulation. *BMC Genomics*. 2015;16:49.
41. Rauluseviciute I, Drablos F, and Rye MB. DNA hypermethylation associated with upregulated gene expression in prostate cancer demonstrates the diversity of epigenetic regulation. *BMC Med Genomics*. 2020;13(1):6.
42. O'Regan GM, Sandilands A, McLean WHI, and Irvine AD. Filaggrin in atopic dermatitis. *J Allergy Clin Immunol*. 2008;122(4):689-93.
43. Navarini AA, French LE, and Hofbauer GF. Interrupting IL-6-receptor signaling improves atopic dermatitis but associates with bacterial superinfection. *J Allergy Clin Immunol*. 2011;128(5):1128-30.
44. Suarez-Farinas M, Ungar B, Correa da Rosa J, Ewald DA, Rozenblit M, Gonzalez J, et al. RNA sequencing atopic dermatitis transcriptome profiling provides insights into novel disease mechanisms with potential therapeutic implications. *J Allergy Clin Immunol*. 2015;135(5):1218-27.
45. Roesner LM, Floess S, Witte T, Olek S, Huehn J, and Werfel T. Foxp3(+) regulatory T cells are expanded in severe atopic dermatitis patients. *Allergy*. 2015;70(12):1656-60.
46. Kaburagi Y, Shimada Y, Nagaoka T, Hasegawa M, Takehara K, and Sato S. Enhanced production of CC-chemokines (RANTES, MCP-1, MIP-1alpha, MIP-1beta, and eotaxin) in patients with atopic dermatitis. *Arch Dermatol Res*. 2001;293(7):350-5.
47. Yamamoto N, Sugiura H, Tanaka K, and Uehara M. Heterogeneity of interleukin 5 genetic background in atopic dermatitis patients: significant difference between those with blood eosinophilia and normal eosinophil levels. *J Dermatol Sci*. 2003;33(2):121-6.
48. Namkung JH, Lee JE, Kim E, Park GT, Yang HS, Jang HY, et al. An association between IL-9 and IL-9 receptor gene polymorphisms and atopic dermatitis in a Korean population. *J Dermatol Sci*. 2011;62(1):16-21.
49. Paradis TJ, Cole SH, Nelson RT, and Gladue RP. Essential role of CCR6 in directing activated T cells to the skin during contact hypersensitivity. *J Invest Dermatol*. 2008;128(3):628-33.
50. Homey B, Steinhoff M, Ruzicka T, and Leung DY. Cytokines and chemokines orchestrate atopic skin inflammation. *J Allergy Clin Immunol*. 2006;118(1):178-89.
51. Barnes KC. An update on the genetics of atopic dermatitis: scratching the surface in 2009. *J Allergy Clin Immunol*. 2010;125(1):16-29 e1-11; quiz 30-1.
52. O'Regan GM, and Irvine AD. The role of filaggrin loss-of-function mutations in atopic dermatitis. *Curr Opin Allergy Clin Immunol*. 2008;8(5):406-10.
53. Boonpiyathad S, Pornsuriyasak P, Buranapraditkun S, and Klaewsongkram J. Interleukin-2 levels in exhaled breath condensates, asthma severity, and asthma control in nonallergic asthma. *Allergy Asthma Proc*. 2013;34(5):e35-41.
54. Murakami Y, Ishii T, Nunokawa H, Kurata K, Narita T, and Yamashita N. TLR9-IL-2 axis exacerbates allergic asthma by preventing IL-17A hyperproduction. *Sci Rep*. 2020;10(1):18110.
55. Randolph AG, Lange C, Silverman EK, Lazarus R, Silverman ES, Raby B, et al. The IL12B gene is associated with asthma. *Am J Hum Genet*. 2004;75(4):709-15.
56. Manni ML, and Alcorn JF. The enigmatic role of IL-22 in asthma. *Expert Rev Respir Med*. 2016;10(6):619-23.
57. Hwang S, Son SW, Kim SC, Kim YJ, Jeong H, and Lee D. A protein interaction network associated with asthma. *J Theor Biol*. 2008;252(4):722-31.
58. Kim E, Lember M, Fallata GM, Rowe JC, Martin TL, Satoskar AR, et al. Intestinal Epithelial Cells Regulate Gut Eotaxin Responses and Severity of Allergy. *Front Immunol*. 2018;9:1692.

59. Kishida T, Hiromura Y, Shin-Ya M, Asada H, Kuriyama H, Sugai M, et al. IL-21 induces inhibitor of differentiation 2 and leads to complete abrogation of anaphylaxis in mice. *J Immunol.* 2007;179(12):8554-61.
60. Pathak M, Padghan P, Halder N, Shilpi, Kulkarni N, Sonar SA, et al. CCR9 signaling in dendritic cells drives the differentiation of Foxp3(+) Tregs and suppresses the allergic IgE response in the gut. *Eur J Immunol.* 2020;50(3):404-17.
61. Galand C, Leyva-Castillo JM, Yoon J, Han A, Lee MS, McKenzie ANJ, et al. IL-33 promotes food anaphylaxis in epicutaneously sensitized mice by targeting mast cells. *J Allergy Clin Immunol.* 2016;138(5):1356-66.
62. HayGlass KT, Lotoski LC, Xie J, Chooniedass R, Liem J, Becker AB, et al. Distinct Roles of IL-10 in Regulation of Peanut Allergy in Humans. *J Allergy Clin Immun.* 2009;123(2):S27-S.
63. Sugita K, Hanakawa S, Honda T, Kondoh G, Miyachi Y, Kabashima K, et al. Generation of Helios reporter mice and an evaluation of the suppressive capacity of Helios(+) regulatory T cells in vitro. *Exp Dermatol.* 2015;24(7):554-6.
64. Santos AF, and Lack G. Food allergy and anaphylaxis in pediatrics: update 2010-2012. *Pediatr Allergy Immunol.* 2012;23(8):698-706.
65. Sicherer SH, Furlong TJ, Maes HH, Desnick RJ, Sampson HA, and Gelb BD. Genetics of peanut allergy: a twin study. *J Allergy Clin Immunol.* 2000;106(1 Pt 1):53-6.
66. Li W, Christiansen L, Hjelmborg J, Baumbach J, and Tan Q. On the power of epigenome-wide association studies using a disease-discordant twin design. *Bioinformatics.* 2018;34(23):4073-8.
67. Sahu M, and Prasuna JG. Twin Studies: A Unique Epidemiological Tool. *Indian J Community Med.* 2016;41(3):177-82.
68. Levesque ML, Casey KF, Szyf M, Ismaylova E, Ly V, Verner MP, et al. Genome-wide DNA methylation variability in adolescent monozygotic twins followed since birth. *Epigenetics.* 2014;9(10):1410-21.

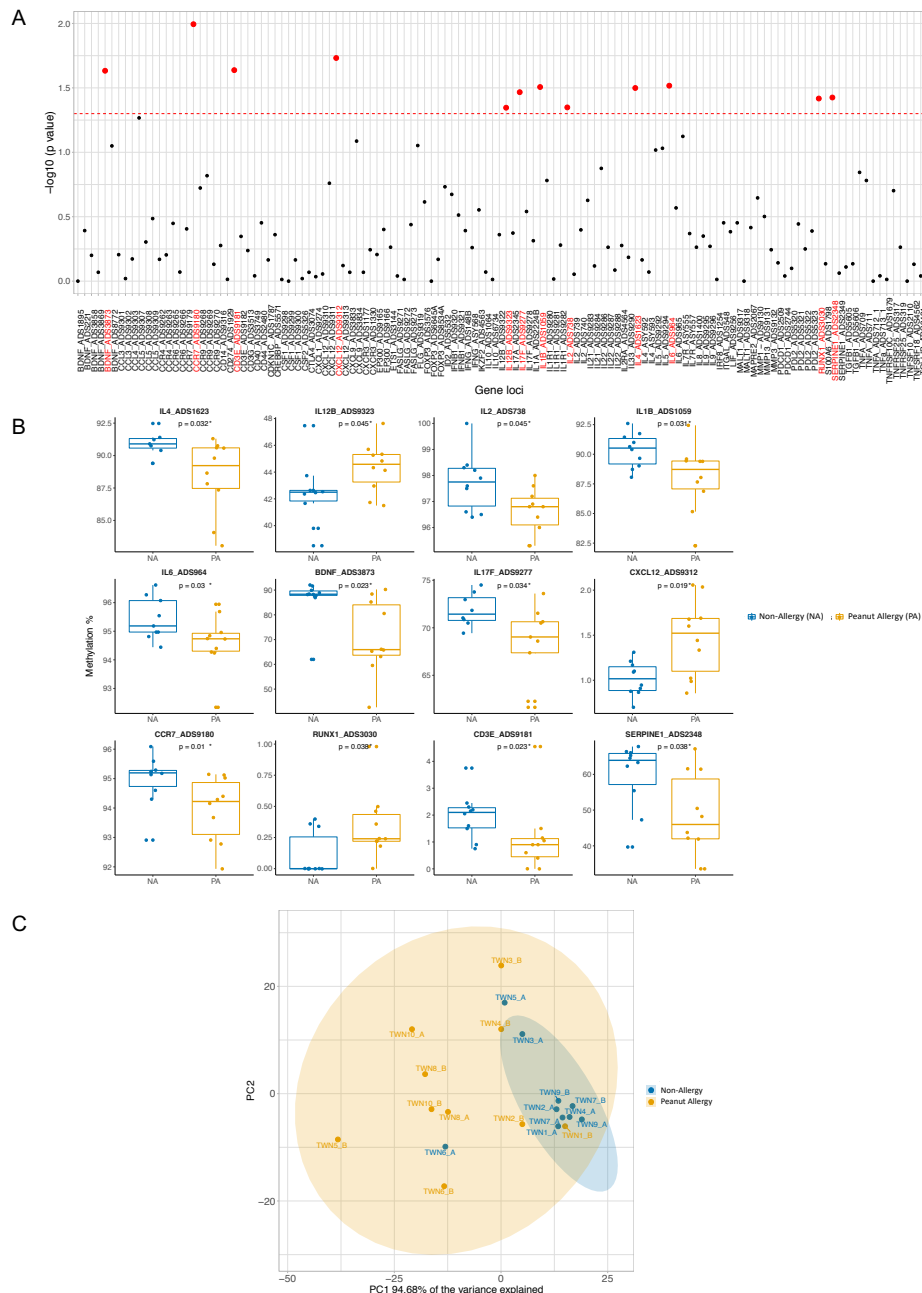


**Figure 1. The significant differences in the DNAm levels for targeted 12 genomic regions were observed between PA and NA participants. (A)** The comparison analysis on the average DNAm level in each of 125 targeted genomic regions between PA (n=10) and NA (n=10) participants was performed using nonparametric unpaired comparison test (Wilcoxon rank sum test). The y axis shows the  $-\log_{10} p$  value and the x axis shows 125 targeted genomic regions. Horizontal red dash line represents a statistical significance level of  $p = 0.05$ . Each dot represents the  $p$  value and the red dots indicate the  $p$  value less than of 0.05. **(B)** The box plots overlaid with dot plots showed the significant differences in the average methylation levels for 12 targeted genomic regions between PA (n=10) and NA (n=10) participants ( $*p < 0.05$ ) **(C)** PCA of the DNAm levels in the targeted 12 genomic regions shows the two distinct clusters formed from NA and PA individuals, respectively. The percentage variance explained by principal component (PC) 1 is indicated. Yellow circles represent PA samples; blue circles represent NA samples.

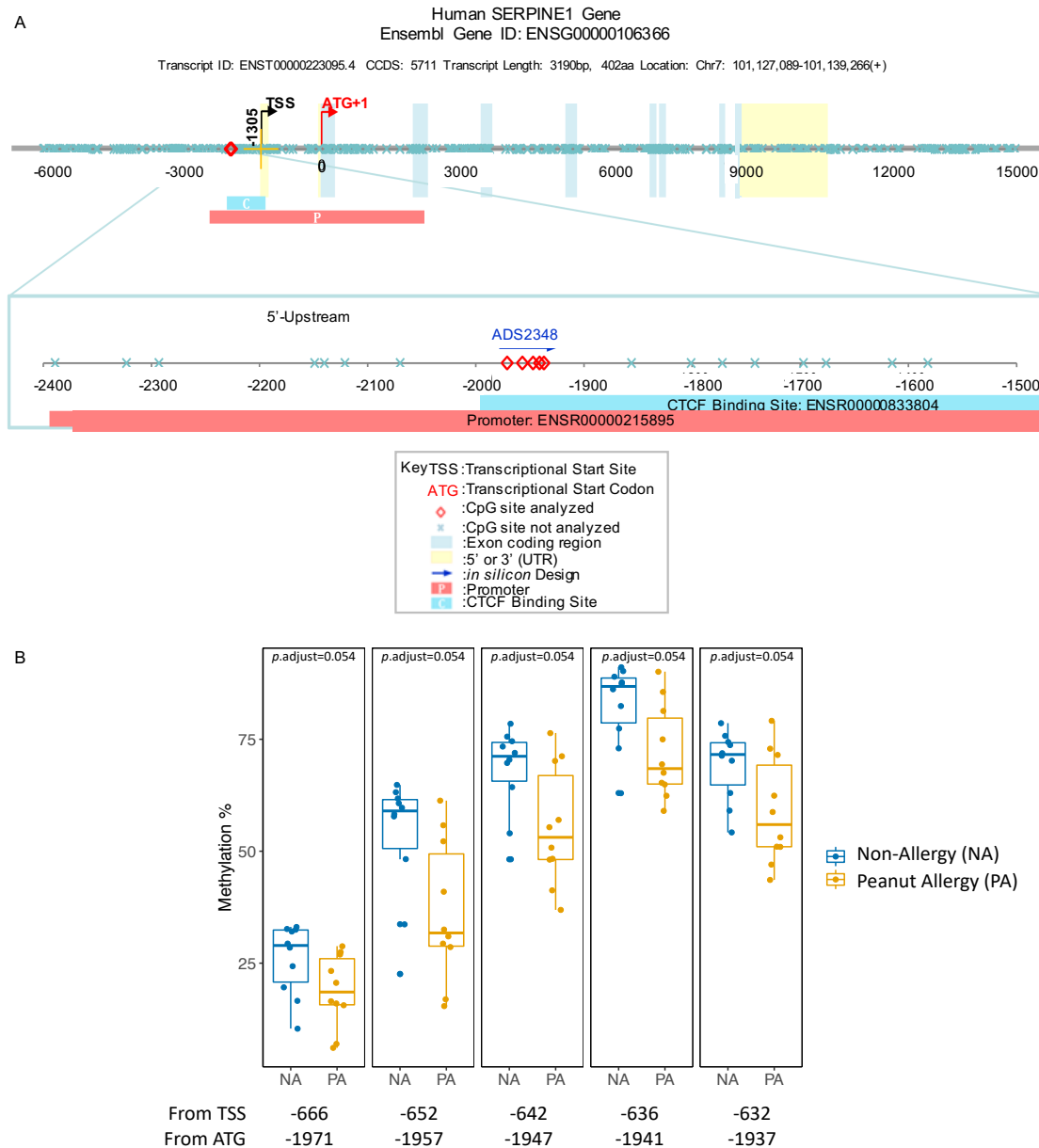


**Figure 2. (A)** Schematic of a genomic region for the *SERPINE1* gene with known CpG sites. The 5 CpG sites labeled with red diamond within the targeted genomic region (chr7:101126423-101126457) were analyzed by tNGBS in our study. Bottom: The box plots overlaid with dot plots showed the significant differences in methylation levels at each of 5 CpG sites in targeted *SERPINE1* region between PA (n=10) and NA (n=10) participants (adjusted  $p < 0.1$ ). Each dot represents one sample. (B) Box plots indicate the interquartile range (IQR) and median; whiskers extend to the farthest data point within a maximum of  $1.5 \times \text{IQR}$ . Sample sets were analyzed using the Wilcoxon rank sum test (two sided).

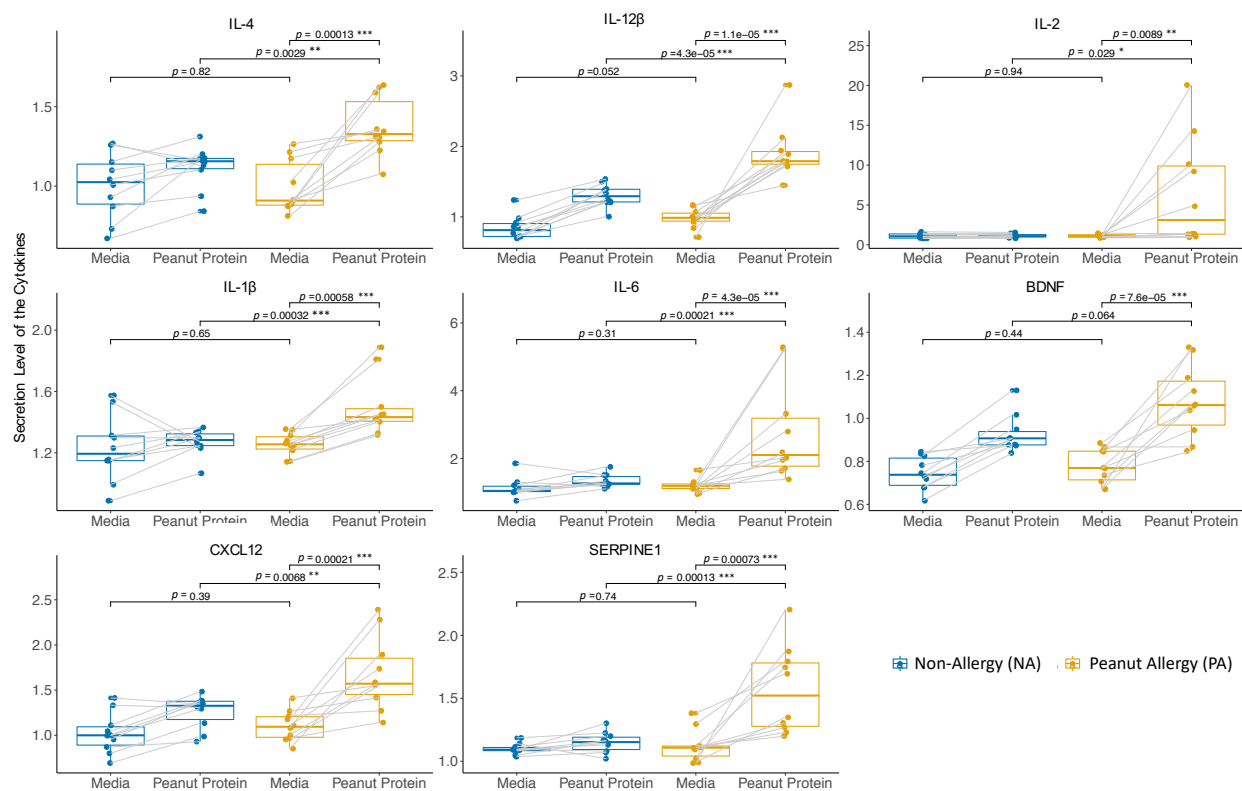
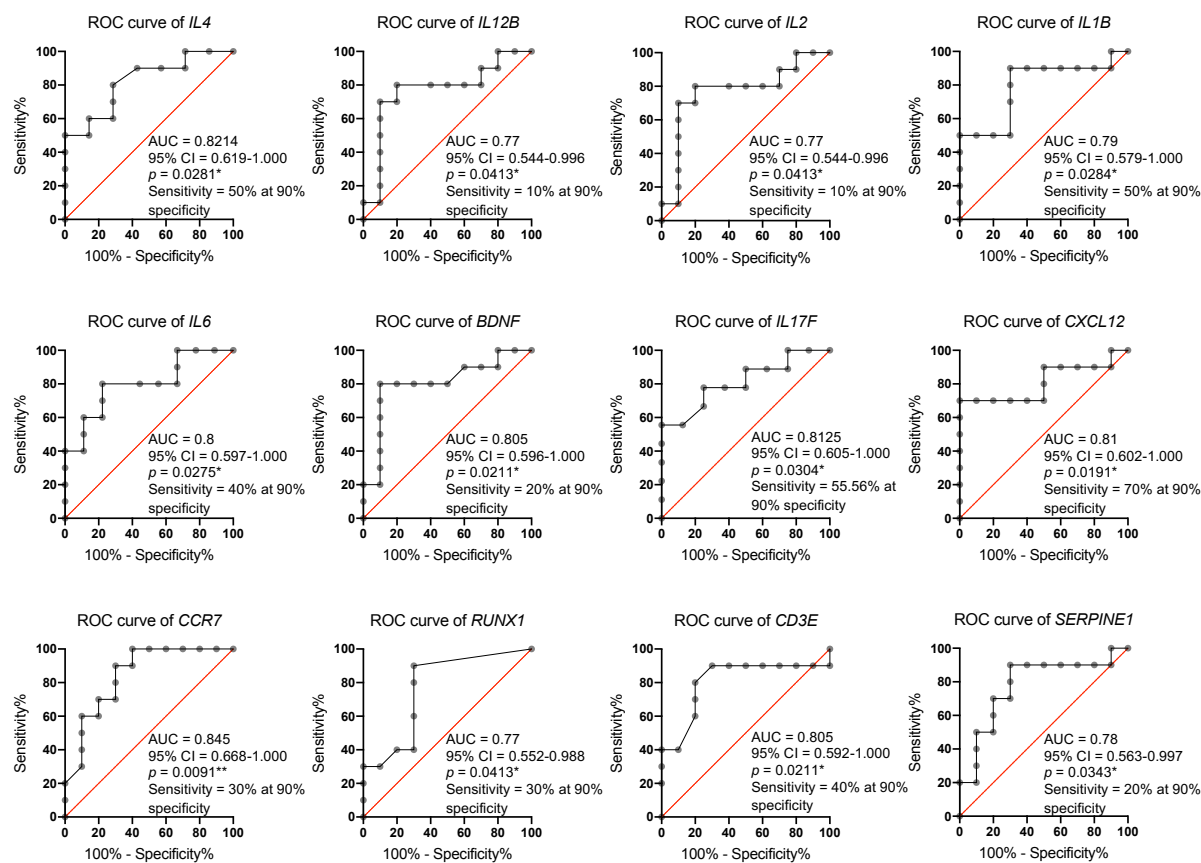


**Figure 3. PBMCs from PA participants secrete higher levels of cytokines (IL-4, IL-12 $\beta$ , IL-2, IL-1 $\beta$ , IL-6, CXCL12, and BDNF) and SERPINE1 protein compared to NA participants.** Secreted levels of cytokines and SERPINE1 protein from PBMCs stimulated with or without peanut protein for PA (n=10) and NA (n=10) participants are shown in box plots overlaid with dot plots (\*p<0.05, \*\*p<0.01, \*\*\*<0.001.). Each pair of points connected by a line represents one sample. Box plots indicate the IQR and median; whiskers extend to the farthest data point within a maximum of 1.5 $\times$  IQR. Sets of paired samples were analyzed using the Wilcoxon signed rank test (two sided). Unpaired sample sets were analyzed using the Wilcoxon rank sum test (two sided).

A



B

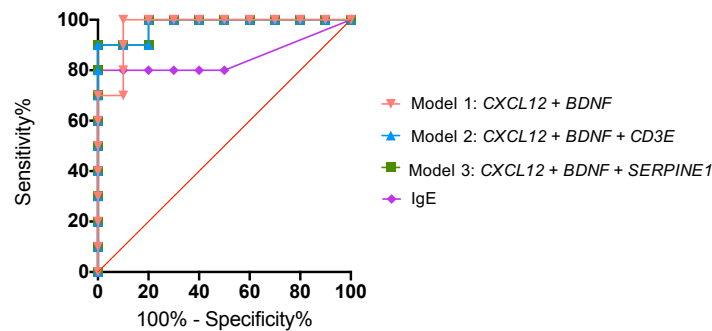
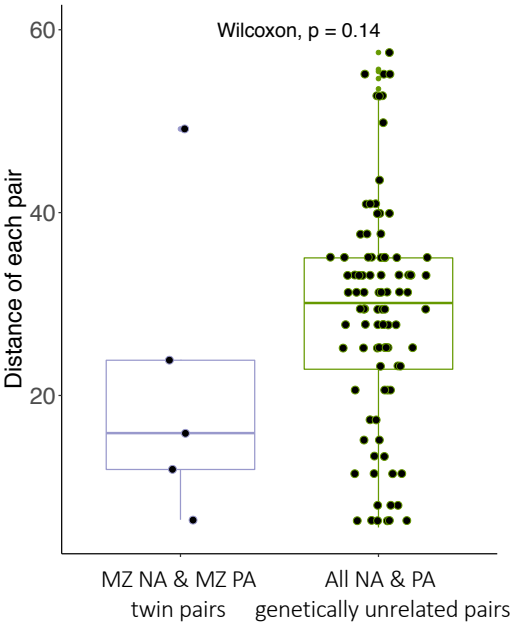


**Figure 4. ROC analysis is applied to evaluate the performance of DNAm signatures for discrimination of PA vs. NA groups. (A) ROC curves show the AUC and the sensitivity and specificity for each of 12 DNAm signatures. (B) ROC curves show the AUC, and the sensitivity and specificity, for comparison of 3 models with the combination DNAm signatures against peanut specific IgE in serum.**

A

|        | MZ PA   | All PA   | All NA |
|--------|---------|----------|--------|
| MZ NA  | 15.88 # | 23.86    | 27.92  |
| All NA | 28.64   | 30.11 \$ |        |
| All PA | 17.72   |          |        |

B



C

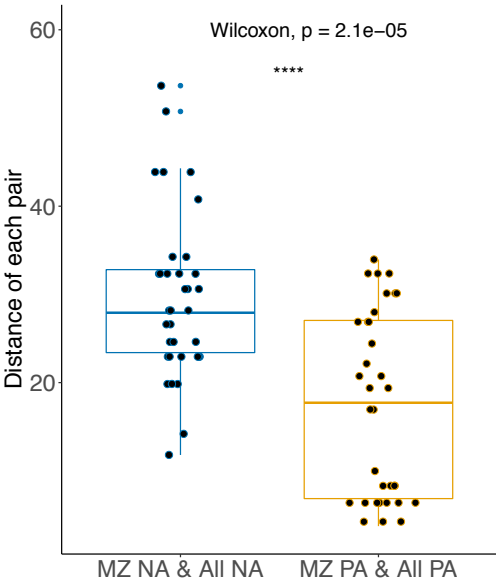


**Figure 5. Peanut allergy associated DNAm levels in 12 targeted genomic regions are genetically influenced.** (A) Euclidean distances of 12 DNAm signatures were calculated pairwise either between 5 MZ twin pairs (pairs=5) who are discordant for peanut allergy, or randomly selected genetically unrelated pairs (pairs=95) (i.e. one sample has peanut allergy and the other is nonallergy without peanut allergy). # median of the distance within 5 discordant MZ twin pairs. \$ median of the distance of all 95 genetically unrelated NA and PA pairs. (B) Box plots overlaid with dot plots represent the Euclidean distances between random pairs of NA MZ twin and another genetically unrelated NA individuals (pairs=35, left panel) and random pairs of PA MZ twin and another genetically unrelated PA individuals (pairs=35, right panel) (C) Box plots overlaid with dot plots represent the Euclidean distances between MZ twins and genetically unrelated individuals in PA participants and in NA individuals. Box plots indicate the interquartile range (IQR) and median; whiskers extend to the farthest data point within a maximum of 1.5× IQR. The Wilcoxon rank sum test (two sided) was used for comparison analysis, \*\*\*p<0.001.

**Table 1.** Demographics for the peanut allergic and nonallergic participants analyzed in the study.

| Twin pair | Sample ID | Peanut allergy assessment  | Zygosity    | Age at blood sample collection | Sex    | Race                      | Food allergies | Food allergens    | History of atopic dermatitis | History of asthma |
|-----------|-----------|----------------------------|-------------|--------------------------------|--------|---------------------------|----------------|-------------------|------------------------------|-------------------|
| 1         | TWN1_B    | Yes by oral food challenge | Monozygotic | 10                             | female | White                     | Allergic       | Peanut, tree nuts | yes                          | yes               |
|           | TWN1_A    | No by history              | Monozygotic | 10                             | female | White                     | Nonallergic    | no                | yes                          | no                |
| 2         | TWN2_A    | No by history              | Monozygotic | 5                              | male   | Asian                     | Nonallergic    | no                | no                           | no                |
|           | TWN2_B    | Yes by oral food challenge | Monozygotic | 5                              | male   | Asian                     | Allergic       | Peanut, tree nuts | no                           | yes               |
| 3         | TWN3_B    | Yes by oral food challenge | Monozygotic | 9                              | female | Black or African American | Allergic       | Peanut, tree nuts | no                           | yes               |
|           | TWN3_A    | No by history              | Monozygotic | 9                              | female | Black or African American | Nonallergic    | no                | yes                          | no                |
| 4         | TWN4_B    | Yes by oral food challenge | Monozygotic | 6                              | male   | Asian                     | Allergic       | Peanut, tree nuts | yes                          | no                |
|           | TWN4_A    | No by history              | Monozygotic | 6                              | male   | Asian                     | Nonallergic    | no                | no                           | no                |
| 5         | TWN5_B    | Yes by oral food challenge | Monozygotic | 5                              | male   | Asian                     | Allergic       | Peanut            | no                           | no                |
|           | TWN5_A    | No by history              | Monozygotic | 5                              | male   | Asian                     | Nonallergic    | no                | no                           | no                |
| 6         | TWN6_B    | Yes by oral food challenge | Dizygotic   | 5                              | male   | Asian                     | Allergic       | Peanut, tree nuts | yes                          | no                |
|           | TWN6_A    | No by history              | Dizygotic   | 5                              | female | Asian                     | Nonallergic    | no                | no                           | no                |
| 7         | TWN7_B    | No by history              | Dizygotic   | 6                              | male   | White                     | Allergic       | tree nuts         | no                           | no                |
|           | TWN7_A    | No by history              | Dizygotic   | 6                              | female | White                     | Nonallergic    | no                | no                           | no                |
| 8         | TWN8_A    | Yes by oral food challenge | Dizygotic   | 4                              | female | Asian                     | Allergic       | Peanut            | no                           | no                |
|           | TWN8_B    | Yes by oral food challenge | Dizygotic   | 4                              | female | Asian                     | Allergic       | Peanut            | yes                          | no                |
| 9         | TWN9_A    | No by history              | Dizygotic   | 10                             | male   | White                     | Nonallergic    | no                | no                           | no                |
|           | TWN9_B    | No by history              | Dizygotic   | 10                             | male   | White                     | Allergic       | tree nuts         | no                           | no                |
| 10        | TWN10_A   | Yes by oral food challenge | Dizygotic   | 5                              | female | White                     | Allergic       | Peanut            | no                           | yes               |
|           | TWN10_B   | Yes by oral food challenge | Dizygotic   | 5                              | female | White                     | Allergic       | Peanut, egg       | no                           | no                |

**Table 2.** Summary of the ROC curve analysis of the 3 models with the combination DNAm signatures selected from initial 12 DNAm signatures and peanut specific IgE in serum.

| Clinical Question              | Predictor                   | AUC  | p value | 95% CI         | Sensitivities(%) at 90 % specificity | AIC  |
|--------------------------------|-----------------------------|------|---------|----------------|--------------------------------------|------|
| Peanut allergy vs. Non-Allergy | IgE                         | 0.85 | 0.0082  | 0.6528 ~ 1.000 | 80                                   | 14.3 |
|                                | <i>CXCL12+BDNF</i>          | 0.97 | 0.0004  | 0.9019 ~ 1.000 | 70                                   | 13.8 |
|                                | <i>CXCL12+BDNF+CD3E</i>     | 0.98 | 0.0003  | 0.9295 ~ 1.000 | 90                                   | 15.4 |
|                                | <i>CXCL12+BDNF+SERPINE1</i> | 0.98 | 0.0003  | 0.9295 ~ 1.000 | 90                                   | 15.4 |

AUC: The area under the ROC curve

95% CI: 95% Confidence Intervals

AIC: Akaike information criterion