nature COMMUNICATIONS

ARTICLE

Received 28 Oct 2014 | Accepted 16 Jan 2015 | Published 24 Feb 2015

DOI: 10.1038/ncomms7304

Genome-wide association study identifies peanut allergy-specific loci and evidence of epigenetic mediation in US children

Xiumei Hong[1,*], Ke Hao[2,*], Christine Ladd-Acosta[3,*], Kasper D. Hansen[4,5], Hui-Ju Tsai[6,7,8], Xin Liu[6,9], Xin Xu[10], Timothy A. Thornton[11], Deanna Caruso[1], Corinne A. Keet[3,12], Yifei Sun[4], Guoying Wang[1], Wei Luo[2,13], Rajesh Kumar[14], Ramsay Fuleihan[14], Anne Marie Singh[15], Jennifer S. Kim[14,16], Rachel E. Story[14,17], Ruchi S. Gupta[6], Peisong Gao[18], Zhu Chen[1], Sheila O. Walker[1], Tami R. Bartell[6], Terri H. Beaty[3], M. Daniele Fallin[19], Robert Schleimer[20], Patrick G. Holt[21], Kari Christine Nadeau[22], Robert A. Wood[12], Jacqueline A. Pongracic[14], Daniel E. Weeks[23] & Xiaobin Wang[1,24]

Food allergy (FA) affects 2%–10% of US children and is a growing clinical and public health problem. Here we conduct the first genome-wide association study of well-defined FA, including specific subtypes (peanut, milk and egg) in 2,759 US participants (1,315 children and 1,444 parents) from the Chicago Food Allergy Study, and identify peanut allergy (PA)-specific loci in the *HLA-DR* and *-DQ* gene region at 6p21.32, tagged by rs7192 ($P=5.5\times10^{-8}$) and rs9275596 ($P=6.8\times10^{-10}$), in 2,197 participants of European ancestry. We replicate these associations in an independent sample of European ancestry. These associations are further supported by meta-analyses across the discovery and replication samples. Both single-nucleotide polymorphisms (SNPs) are associated with differential DNA methylation levels at multiple CpG sites ($P<5\times10^{-8}$), and differential DNA methylation of the *HLA-DQB1* and *HLA-DRB1* genes partially mediate the identified SNP–PA associations. This study suggests that the *HLA-DR* and *-DQ* gene region probably poses significant genetic risk for PA.

[1] Department of Population, Family and Reproductive Health, Center on the Early Life Origins of Disease, Johns Hopkins University Bloomberg School of Public Health, 615 North Wolfe Street, E4132, Baltimore, Maryland 21205, USA. [2] Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. [3] Department of Epidemiology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland 21205, USA. [4] Department of Biostatistics, Johns Hopkins University Bloomberg School of Public Health Baltimore, Baltimore, Maryland 21205, USA. [5] McKusick-Nathans Insitute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. [6] Mary Ann and J. Milburn Smith Child Health Research Program, Department of Pediatrics, Northwestern University Feinberg School of Medicine and Stanley Manne Children's Research Institute, Ann and Robert H. Lurie Children's Hospital of Chicago, Chicago, Illinois 60611, USA. [7] Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, Zhunan 35053, Taiwan. [8] Department of Public Health, China Medical University, Taichung 40402, Taiwan. [9] Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois 60611, USA. [10] Guangdong Provincial Institute of Nephrology, Southern Medical University, Guangzhou 510515, China. [11] Department of Biostatistics, University of Washington, Seattle, Washington 98195, USA. [12] Division of Pediatric Allergy and Immunology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21287, USA. [13] College of Computer Science and Technology, Huaqiao University, Xiamen 361021, China. [14] Division of Allergy and Immunology, Ann and Robert H. Lurie Children's Hospital of Chicago, Chicago, Illinois 60611, USA. [15] Department of Pediatrics and Medicine, Ann and Robert H. Lurie Children's Hospital of Chicago, Northwestern Feinberg School of Medicine, Chicago, Illinois 61611, USA. [16] NorthShore University HealthSystem, Evanston, Illinois 60201, USA. [17] NorthShore University Health Systems, Pritzker School of Medicine, University of Chicago, Chicago, Illinois 60637, USA. [18] Division of Allergy and Clinical Immunology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21224, USA. [19] Department of Mental Health, Wendy Klag Center for Autism and Developmental Disabilities, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland 21205, USA. [20] Division of Allergy-Immunology, Northwestern University Feinberg School of Medicine, Chicago, Illinois 60611, USA. [21] Telethon Kids Institute, University of Western Australia; Perth and Queensland Children's Medical Research Institute, University of Queensland, Brisbane, Queensland 4029, Australia. [22] Division of Allergy, Immunology and Rheumatology, Stanford University School of Medicine, Stanford, California 94305, USA. [23] Departments of Human Genetics and Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, USA. [24] Division of General Pediatrics and Adolescent Medicine, Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, Maryland 21287, USA. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to X.W. (email: xwang82@jhu.edu).

Food allergy (FA), defined as immunoglobulin E (IgE)-mediated clinical reactivity to specific food proteins, affects 2%–10% of children in the United States[1,2]. Over the past 20 years, FA has grown from a relatively uncommon to a major clinical and public health problem worldwide, due to its increasing prevalence, potential fatality and enormous medical and economic impact[3–5]. FA accounts for more than $20 billion in overall annual healthcare costs in the United States[5]. To date, there is no safe and effective prevention or treatment for FA, which is approved by the Food and Drug Administration for use in clinical practice, except for emergency management of allergic reactions induced by accidental exposure. Strict food avoidance is the only effective strategy to prevent future allergic reactions among FA patients; however, this is exceedingly difficult, as peanut, egg and cow's milk (the three most common food allergens in the United States[3]) are ubiquitous in processed foods. Indeed, a growing body of literature has shown that FA significantly diminishes quality of life among affected patients and their caregivers[3,6,7], who live in constant fear of accidental ingestion and potentially life-threatening reactions.

A major obstacle to effective prevention and treatment of FA is our limited understanding of its causes and underlying biological mechanisms. Although available data support a role for genetic factors in FA based on familial aggregation studies[8,9] and heritability estimates (ranging from 15% to 82%)[8,10], few specific genes have been conclusively associated with FA, leaving its heritability largely unexplained. To date, all published genetic studies of FA have used a candidate gene approach, and have predominantly focused on peanut allergy (PA), a common type of FA that accounts for a disproportionate number of fatal and near-fatal food-induced episodes of anaphylaxis. Although multiple candidate genes have been reported for PA, few are considered established due to small sample sizes[11–17], lack of adjustment for multiple comparisons[12–14,16–18] and a high failure rate in genotyping calls[18]. To our knowledge, loss-of-function mutations in the gene encoding filaggrin (*FLG*) are among the few genetic risk factors replicated for PA in multiple populations[19–21]. For other allergic diseases, genome-wide association studies (GWASs) have shown promise in recent years in dissecting the genetic basis of asthma[22–24] and atopic dermatitis[25–27]. Two recent GWASs have examined allergen sensitization[28] and self-reported environmental allergy[29]. These previous GWASs did not specifically examine FA, but did reveal substantial differences in genetic effects across various allergies[29]. Clinical FA is distinct from sensitization to foods or aeroallergens: a sensitized child may or may not manifest clinical signs or symptoms of FA, underscoring the need to explore genetic variants specifically associated with clinical FA (a growing clinical and public health challenge).

Epigenetic mechanisms by which genetic variants affect FA are largely unexplored. DNA methylation (DNAm), a type of epigenetic mark, regulates gene expression. There is growing evidence that genetic variants can affect DNAm[30,31], and that DNAm may mediate genetic susceptibility to autoimmune disease[32]. To date, no published study on FA has simultaneously considered genome-wide genetic and epigenetic factors, nor have they assessed whether DNAm could mediate genetic susceptibility to FA.

To our knowledge, this is the first GWAS of well-defined FA in a US cohort of children and their biological parents. This study comprises three stages. In stage I, we use the modified quasi-likelihood score (MQLS) test[33,34] to detect genetic associations with any FA (including nine foods as follows: peanut, egg white, cow's milk, soy, wheat, walnut, fish, shellfish and sesame seed) and the three most common types of FA (peanut, egg white and cow's milk). We identify PA-specific loci in the human leukocyte antigen (*HLA*)*-DQ* and *-DR* region at 6p21.32, tagged by rs7192 and rs9275596, respectively. In stage II, we perform a replication study of identified single-nucleotide polymorphisms (SNPs) from stage I in an independent sample from the same cohort, and confirm that both rs7192 and rs9275596 are significantly associated with PA. In stage III, we examine relationships between the two PA-associated SNPs and DNAm, between genotype-dependent DNAm and PA, and also whether DNAm mediates identified SNP–PA associations. We find that both SNPs are significantly associated with differential DNAm at multiple CpG sites and differential DNAm of the *HLA-DQB1* and *HLA-DRB1* genes partially mediate the identified SNP–PA association. The population attributable risk (PAR) is 21% and 19%, respectively, for rs7192 and rs9275596. Taken together, this study indicates the possibility that the *HLA-DR* and *-DQ* gene region likely poses the single greatest genetic risk for PA.

Results

Phenotype definition and population characteristics. This study includes three stages as shown in Fig. 1. Both discovery and replication samples were from the Chicago Food Allergy Study and were collected under a standard study protocol, as described in the Methods. Main phenotypes of interest included 'any FA' and the three most common types of FA: PA, egg allergy and milk allergy. As described in our previous report[35], we applied stringent clinical criteria to define specific types of FA as follows: (1) a convincing history of clinical allergic reaction on ingestion of a specified food and (2) evidence of sensitization to the same food, defined as having a detectable food-specific IgE $(\geq 0.10\,\mathrm{kU\,l^{-1}})$ and/or a positive skin prick test (SPT) with mean

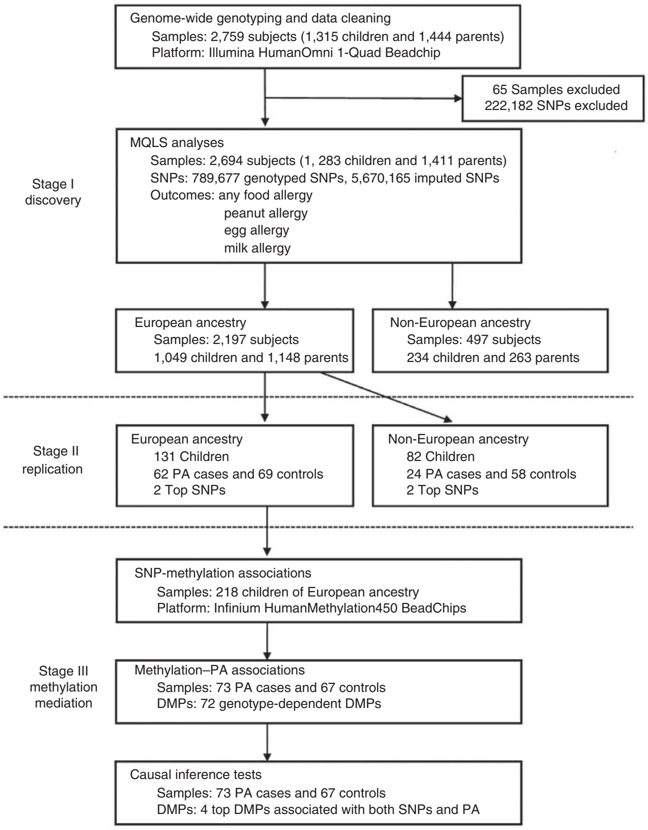
Genome-wide genotyping and data cleaning
Samples: 2,759 subjects (1,315 children and 1,444 parents)
Platform: Illumina HumanOmni 1-Quad Beadchip
65 Samples excluded
222,182 SNPs excluded
Stage I
discovery
MQLS analyses
Samples: 2,694 subjects (1, 283 children and 1,411 parents)
SNPs: 789,677 genotyped SNPs, 5,670,165 imputed SNPs
Outcomes: any food allergy
peanut allergy
egg allergy
milk allergy
European ancestry
Samples: 2,197 subjects
1,049 children and 1,148 parents
Non-European ancestry
Samples: 497 subjects
234 children and 263 parents
Stage II
replication
European ancestry
131 Children
62 PA cases and 69 controls
2 Top SNPs
Non-European ancestry
82 Children
24 PA cases and 58 controls
2 Top SNPs
SNP-methylation associations
Samples: 218 children of European ancestry
Platform: Infinium HumanMethylation450 BeadChips
Stage III
methylation
mediation
Methylation–PA associations
Samples: 73 PA cases and 67 controls
DMPs: 72 genotype-dependent DMPs
Causal inference tests
Samples: 73 PA cases and 67 controls
DMPs: 4 top DMPs associated with both SNPs and PA

Figure 1 | Study design and work flow diagram for the GWAS of FA. DMP, differentially methylated position; MQLS, modified quasi-likelihood score analyses; PA, peanut allergy; SNP, single-nucleotide polymorphism.

weal diameter (MWD) ≥ 3 mm to this specified food (see Methods). Accordingly, we defined specific types of FA to nine common foods (accounting for >95% of all FAs in the population) and defined a child as having any FA if she/he was allergic to any of the nine foods. In the genetic association analyses, we also performed sensitivity analysis using other cutoffs for food-specific IgE and SPT to define FA, for example, food-specific IgE ≥ 0.35 kU l^{-1}, SPT MWD ≥ 5 mm (ref. 36) or either food-specific IgE or SPT MWD ≥95% positive predictive value (PPV).

GWAS to identify susceptibility loci for FA (Stage I). A total of 2,759 samples (1,315 children and 1,444 biological parents) from the Chicago Food Allergy Study[9,35] were genotyped using the Illumina HumanOmni1-Quad BeadChip. We performed vigorous quality control and thorough data cleaning as detailed in Methods. The final sample size for gene discovery (Stage I) included 2,197 individuals of European ancestry (671 FA-affected children) and 497 individuals of non-European ancestry (155 FA-affected children). Each individual's genetic ancestry was estimated by principal component analysis (PCA)[37] by using the 1000 Genomes Project as a reference. The demographic and clinical characteristics of all study participants are provided in Supplementary Table 1.

The Stage I analyses included primarily family-based samples, with a small number of case–control samples. The MQLS test[33,34] was applied to test for genetic associations with FA due to its capacity to maximally use available information contained in the complex family data set (see Methods). The MQLS test allows for two types of controls: unaffected controls and controls of uncertain phenotypes. In this study, the children who did not meet FA case or control definitions and all parents were coded as controls of uncertain phenotypes (see Methods). We also performed sensitivity tests in different ways to incorporate controls of uncertain phenotypes.

To minimize population stratification, we first performed the MQLS test in 2,197 individuals of European ancestry, and then examined whether the identified genome-wide significant association signals were also present in 497 individuals of non-European ancestry. The quantile–quantile plots for association analyses in 2,197 individuals of European ancestry indicated no inflation of the MQLS test due to cryptic population structure or unaccounted relatedness among individuals (Fig. 2 and Supplementary Fig. 1).

We performed the MQLS test to examine genome-wide associations for any FA in 2,197 individuals of European ancestry, including 671 FA cases, 144 non-allergic non-sensitized normal controls and 1,382 controls of uncertain phenotypes (234 children and 1,148 parents). Under an additive model for the minor allele of each SNP, we found that no SNP reached genome-wide significance ($P<5\times10^{-8}$) or suggestive ($P<1\times10^{-7}$) threshold (Table 1). Removing the 234 children with uncertain phenotypes from these analyses did not substantially change the results.

We further examined genome-wide associations for PA, egg allergy and milk allergy among 2,197 individuals of European ancestry. In 316 PA cases, 144 non-allergic non-sensitized controls and 1,737 controls of uncertain phenotypes (589 children and 1,148 parents), the MQLS test identified genome-wide significant associations for 40 SNPs spanning the *HLA class II DQ* genes at the 6p21.32 region (Fig. 2 and Supplementary Table 2). An intergenic SNP, rs9275596, between the *HLA-DQB1* and *HLA-DQA2* genes, showed the most significant association with PA ($P=6.8\times10^{-10}$; Fig. 3 and Supplementary Table 2). The other 39 SNPs were predominantly in moderate-to-strong linkage disequilibrium (LD) with rs9275596 (Fig. 3 and Supplementary Table 2), and their associations with PA were no longer significant when conditioning on rs9275596 (all $P>0.001$), suggesting that this group of SNPs represents a single significant genetic signal for PA. At >200 kb upstream from this significant signal, eight SNPs located within or clustered around the *HLA-DRA* gene were in strong LD with each other ($r^2=1$) and showed suggestive associations with PA ($P<1\times10^{-7}$; Fig. 3). Among these identified SNPs, rs7192 ($P=5.5\times10^{-8}$) was the only coding SNP that leads to a Leu242Val change in the *HLA-DRA* gene product. The odds ratio (OR) and 95% confidence intervals (CIs) were the same for one copy of the *rs7192-T* allele and one copy of the *rs9275596-C* allele: 1.7 (95% CI: 1.4–2.1), as estimated by a generalized estimating

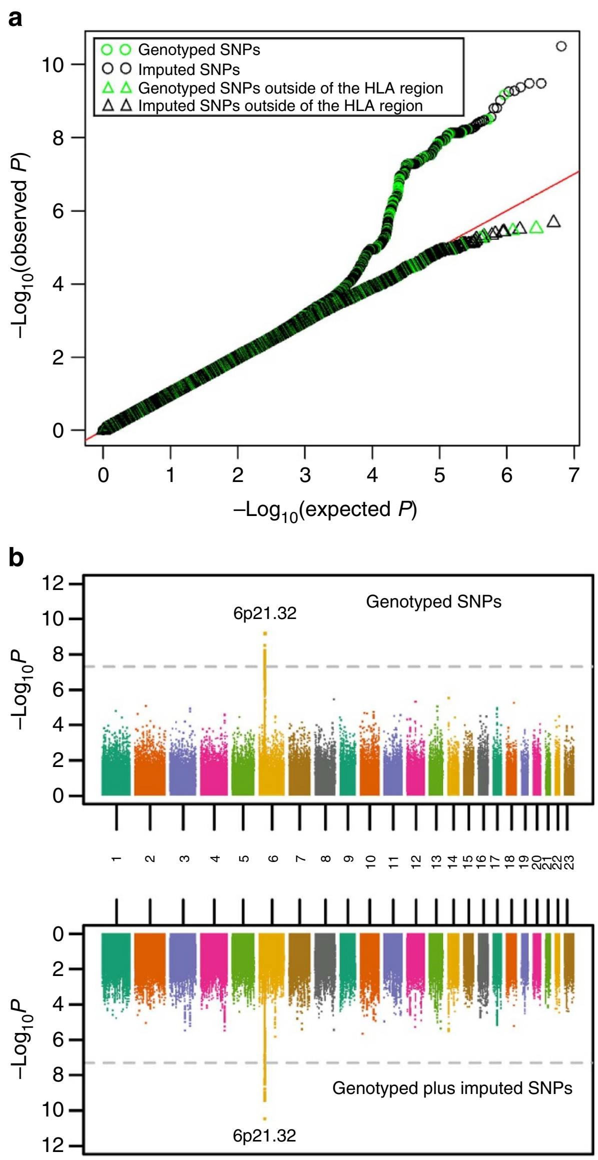

Figure 2 | Quantile-quantile (Q-Q) and Manhattan plots for genome-wide associations of PA in 2,197 discovery samples of European ancestry. (**a**) Q-Q plots for all of the genotyped and/or imputed SNPs and SNPs outside of the HLA-DQ and -DR region. (**b**) Manhattan plots for the genotyped SNPs alone (upper panel) and for the genotyped plus imputed SNPs. SNPs above the dashed line reached genome-wide significance at 5×10^{-8}, based on the MQLS test.

Table 1 | The top loci associated with FA and the three most common types of FA in 2,197 discovery samples of European ancestry.

SNP	CHR	Position	Nearest gene	Allele*	MAF†	*P* for SNP-phenotype association‡			
						FA	PA	Egg allergy	Milk allergy
The five top loci for FA									
rs12121623	1	54931396	*SSBP3|ACOT11*	G/T	0.13	3.1×10^{-7}	0.024	0.046	0.053
rs1318710	4	101436201	*EMCN*	A/G	0.10	2.6×10^{-6}	0.002	0.020	0.207
rs777717	2	195954734	*LOC645314|SLC39A10*	C/T	0.08	4.7×10^{-6}	0.008	0.057	0.214
rs10994607	10	62760742	*RHOBTB1*	C/T	0.04	7.1×10^{-6}	0.250	0.028	4.6×10^{-5}
rs6942407	7	86861313	*LOC100289677|TP53TG1*	G/A	0.21	8.2×10^{-6}	7.6×10^{-4}	0.016	0.003
The five top loci for PA									
rs9275596	6	32681631	*HLA-DQB1|HLA-DQA2*	T/C	0.35	0.006	6.8×10^{-10}	0.509	0.247
rs7192	6	32411646	*HLA-DRA*	G/T	0.39	0.175	5.5×10^{-8}	0.293	0.468
rs862942	14	26492233	*STXBP6|NOVA1*	T/C	0.07	2.7×10^{-4}	3.0×10^{-6}	0.169	0.026
rs4584173	8	135336557	*LOC100129104|ZFAT*	T/C	0.40	0.031	3.6×10^{-6}	0.168	0.386
rs10878354	12	66384885	*HMGA2|LLPH*	G/A	0.23	0.031	5.1×10^{-6}	0.603	0.451
The five top loci for egg allergy									
rs7717393	5	155753914	*SGCD*	C/G	0.07	0.067	0.627	1.4×10^{-6}	2.7×10^{-4}
rs5961136	23	54802520	*ITIH5L*	T/G	0.39	0.251	0.067	2.4×10^{-6}	0.900
rs250585	16	23401076	*COG7*	G/A	0.17	0.003	0.101	3.8×10^{-6}	0.021
rs16823014	2	169817713	*ABCB11*	G/A	0.05	0.352	0.563	4.4×10^{-6}	0.052
rs6498482	16	13987719	*LOC729993|ERCC4*	T/C	0.39	0.022	0.158	4.8×10^{-6}	0.006
The five top loci for milk allergy									
rs9898058	17	47818821	*FAM117A*	C/T	0.15	7.7×10^{-5}	0.014	0.016	1.1×10^{-6}
rs17032597	2	67055115	*LOC100289292|ETAA1*	C/A	0.29	0.003	0.568	0.074	1.6×10^{-6}
rs78405116	11	1892562	*LSP1*	G/T	0.03	0.015	0.423	0.775	1.7×10^{-6}
rs10994613	10	62780127	*RHOBTB1|TMEM26*	G/A	0.03	1.6×10^{-5}	0.997	0.036	4.8×10^{-6}
rs7833294	8	58008281	*IMPAD1|LOC286177*	C/T	0.02	0.281	0.504	0.066	7.3×10^{-6}

CHR, chromosome; FA, food allergy; MAF, minor allele frequency; MQLS, modified quasi-likelihood score; PA, peanut allergy; SNP, single-nucleotide polymorphism.
Only the genotyped SNP with the minimum *P*-value based on the MQLS test is shown for each gene.
*Major/minor allele.
†The MAF was calculated using the genotyping data from parents of European ancestry.
‡*P*-value was generated using the MQLS test in the 2,197 discovery samples.

equation (GEE) regression model (Table 2). The association between rs9275596 and PA was significantly reduced (OR = 1.4, 95% CI = 1.1–1.8, $P = 0.01$) when conditioned on rs7192, suggesting that SNP rs9275596 and rs7192 may represent a single risk factor for PA.

We performed a sensitivity test to examine whether the identified genetic associations for PA varied by IgE or SPT cutoffs. As shown in Table 3, the estimated ORs for either rs7192 or rs9275596 remained similar when more stringent IgE or SPT cutoffs were applied. We then tested associations for rs7192 and rs9275596 with PA, in 497 individuals of non-European ancestry. No significant associations were detected (Table 2). These results were unchanged when a quantitative version of MQLS analyses was applied to adjust for ancestry based on the first three principal components (PCs) from the genome-wide SNP genotypes.

The MQLS test for egg allergy and milk allergy did not identify any genome-wide significant or suggestive SNPs in the 2,197 individuals of European ancestry, and the two PA-associated SNPs (rs7192 and rs9275596) showed no evidence of association with either egg or milk allergy (Table 1). Neither rs7192 nor rs9275596 showed associations with other allergic phenotypes in 2,197 individuals of European ancestry (Supplementary Table 3).

SNP imputation was also conducted in this study. With the latest versions of SHAPEIT[38] and IMPUTE2 (ref. 39), a total of 6,459,842 genotyped and/or imputed SNPs passed post-imputation quality control steps (see Methods) and were then tested for their associations with each outcome (any FA, PA, egg allergy and milk allergy) among 2,197 individuals of European ancestry. The MQLS test for PA revealed a single genome-wide significant peak at 6p21.32, the same region identified in our original analysis of genotyped SNPs alone (Fig. 2), but the peak now included an additional 99 imputed SNPs (Supplementary Table 2). Imputed SNP rs33980016 (an insertion/deletion SNP in the intronic region of the *HLA-DQB1* gene) showed the most significant association with PA ($P = 3.2\times10^{-11}$; Fig. 3); it was in moderate LD ($r^2 = 0.54$) with genotyped SNP rs9275596. When conditioned on the genotyped SNP (rs9275596 or rs7192), the association between rs33980016 and PA was largely reduced ($P > 0.001$), suggesting that these imputed and genotyped SNPs may represent one single genetic region. No genome-wide significant or suggestive associations were identified for any FA and egg allergy (Supplementary Fig. 1). Four imputed SNPs on chromosome 3, located between the *C3orf67* and *LOC339902* genes, showed suggestive associations ($P < 5\times10^{-7}$) with milk allergy. However, this suggestive peak appears to be driven by imputed SNPs, because there is no such association based on genotyped SNPs in this region (Supplementary Fig. 1).

As no genome-wide significant associations were found for any FA, egg allergy or milk allergy, we narrowed the scope of our replication and DNAm mediation analyses to focus on PA.

Replication and meta analyses (Stage II). We performed a replication study for two genotyped PA-associated SNPs, rs9275596 (the most significant genotyped SNP for PA) and rs7192 (a potential functional SNP), in an independent sample (86 PA cases and 127 controls) from the same Chicago Food

Allergy cohort. Sample selection criteria and data cleaning procedures are provided in Methods. The demographic characteristics of the replication sample are presented in Supplementary Table 4. There were 131 children (62 PA cases and 69 controls) of European ancestry and 82 of non-European ancestry (24 PA cases and 58 controls) based on PCA analysis. Using GEE models to account for correlations among the 23 sibling pairs and to adjust for age and gender, we found that both rs7192 (OR = 1.8, 95% CI = 1.2–2.7, $P = 0.005$) and rs9275596 (OR = 1.7, 95% CI = 1.1–2.6, $P = 0.\ 022$) were significantly associated with PA after Bonferroni correction ($P < 0.025$ for two SNP tests) in children of European ancestry, and that both SNPs had a similar effect size as seen in the Stage I GWAS results (Table 2). No such associations were detected in children of non-European ancestry (Table 2).

Using allele frequencies reported in the HapMap Utah Residents with Northern and Western European Ancestry (CEU) and ORs derived from the replication sample of European ancestry (to avoid 'winner's curse' bias), we estimated the PAR, which was 21% for rs7192 and 19% for rs9275596.

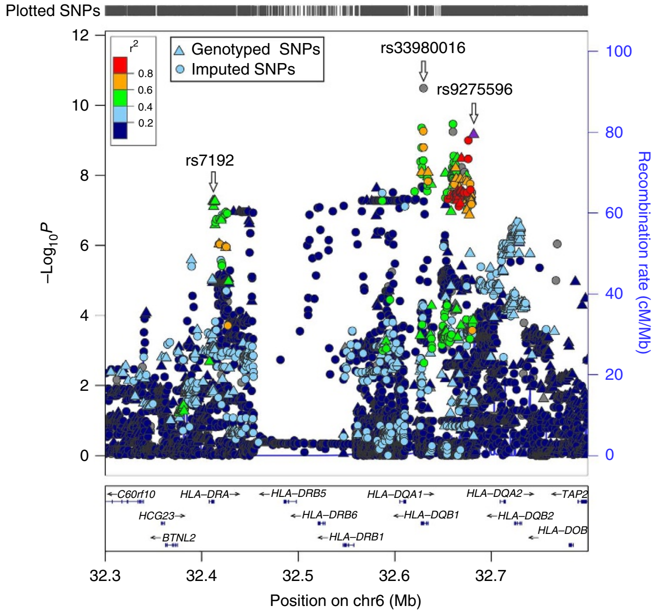

Figure 3 | Locus-specific plot of PA-associated loci reaching genome-wide significance. Each dot represents the $-\log_{10}$ (P-value) for one genotyped or imputed SNP based on the MQLS test in 2,197 discovery samples of European ancestry. The estimated recombination rates from the 1000 Genomes Project data are shown as blue lines and the genomic locations of genes within the regions of interest are shown at the bottom. SNP colour represents LD with the most significant genotyped SNP (rs9275596). SNP annotations are indicated as follows: triangles, genotyped SNPs; circles, imputed SNPs.

DNAm mediation (Stage III). In 218 unrelated children of European ancestry (including 73 PA cases, 67 non-PA controls and 78 children with uncertain PA phenotype), we tested associations between the top genotyped SNPs (rs7192 and rs9275596) and whole-blood DNAm levels at genome-wide CpG sites derived from the Infinium HumanMethylation450 BeadChip. When adjusting for age, gender and estimated cell composition (see Methods), we identified 72 differentially methylated positions (DMPs, $P < 5 \times 10^{-8}$): 69 located at chromosome 6p21.32 (Fig. 4a), 1 at 7q22 (cg03324851 in the *GNB2* gene), 1 at 12q24.33 (cg01256320 in the *FBRSL1* gene) and 1 at 17q25.1 (cg12311094 in the *C17orf77* gene). Of these 72 DMPs, 29 were significantly associated with rs7192 and rs9275596 (the top PA-associated genotyped SNPs), 17 were significantly associated with rs7192 and the remaining 26 DMPs were significantly associated with rs9275596.

The 72 identified genotype-dependent DMPs were then tested for their associations with PA in 73 PA cases and in 67 non-PA controls. The estimated proportion of CD4T cells in whole blood was slightly lower in PA cases than in controls ($P = 0.02$). With the adjustment for age, gender and estimated cell composition, a total of 18 DMPs, located in the *c6orf10* ($N = 7$), *HLA-DRB5* ($N = 2$), *HLA-DRB1* ($N = 8$) and *HLA-DQB1* ($N = 1$) genes, respectively, were significantly associated with the risk of PA after Bonferroni correction ($P < 0.0005$) and showed a ≥5% adjusted methylation level difference between PA cases and controls (Fig. 4a).

For each of the four identified genes (*c6orf10*, *HLA-DRB5*, *HLA-DRB1* and *HLA-DQB1*), the top DMP, which was (1) significantly associated with both rs7192 and rs9275596 ($P < 5 \times 10^{-8}$), and (2) yielded the smallest P-value in association tests with PA, was further tested for its role in mediating the

Table 2 | The estimated effect sizes of the two top SNPs on risk of PA in the discovery and replication samples.

SNP	Allele*	Discovery					Replication				P_{meta}¶
		Case†	Control†	Uncertain†,‡	OR (95% CI)§	P_{MQLS}‖	Case†	Control†	OR (95% CI)§	P_{GEE}§	
Subjects with European ancestry											
		$N = 316$	$N = 144$	$N = 1{,}737$‡			$N = 62$	$N = 69$			
rs7192	G/T	0.49	0.42	0.40	1.7 (1.4–2.1)	5.5×10^{-8}	0.50	0.31	1.8 (1.2–2.7)	0.005	2.7×10^{-9}
rs9275596	T/C	0.46	0.37	0.35	1.7 (1.4–2.1)	6.8×10^{-10}	0.44	0.31	1.7 (1.1–2.6)	0.022	6.3×10^{-11}
Subjects with non-European ancestry											
		$N = 80$	$N = 15$	$N = 402$#			$N = 24$	$N = 58$			
rs7192	G/T	0.43	0.46	0.36	1.2 (0.8–1.8)	0.198	0.42	0.36	1.4 (0.7-3.1)	0.375	0.147
rs9275596	T/C	0.34	0.41	0.30	1.2 (0.8–1.8)	0.327	0.25	0.40	0.6 (0.2-1.3)	0.176	0.420

CI, confidence interval; GEE, generalized estimating equation; MQLS, modified quasi-likelihood score; OR, odds ratio; PA, peanut allergy; SNP, single-nucleotide polymorphism.
*Major/minor allele. The major allele is the reference allele and the minor allele is the effective allele.
†Minor allele frequency is shown in each group.
‡Controls of uncertain phenotype, which included 1,148 parents and 589 children.
§The GEE model was applied to estimate the effect size of each SNP (additive genetic model) on the risk of PA in children with and without PA, with adjustment for age and gender. For analyses in non-European subjects, besides age and gender, we also included as covariates the first three principal components from the genome-wide SNP genotypes, to control for potential population stratification.
‖P-value from the MQLS analyses in the 2,197 discovery samples.
¶Meta-analysis was performed based on the Stouffer's weighted z-score method to combine association results from the MQLS analyses in the discovery sample and from the GEE analyses in the replication sample.
#Controls of uncertain phenotype, which included 263 parents and 139 children.

Table 3 | Associations of the two top SNPs with PA based on various definitions in 2,197 discovery samples of European ancestry.

PA definitions	Cases	rs7192*				rs9275596†			
		OR‡	95% CI‡	P_{GEE}‡	P_{MQLS}§	OR‡	95% CI‡	P_{GEE}‡	P_{MQLS}§
CR, plus psIgE $\geq$0.1 kU l^{-1} or SPT $\geq$3 mm	316	1.7	1.4–2.1	1.7×10^{-7}	5.5×10^{-8}	1.7	1.4–2.1	9.6×10^{-8}	6.8×10^{-10}
CR in 2 h, plus psIgE $\geq$0.1 kU l^{-1} or SPT $\geq$3 mm	286	1.8	1.4–2.2	1.5×10^{-7}	6.7×10^{-8}	1.8	1.4–2.2	1.8×10^{-7}	1.9×10^{-9}
CR in 2 h, plus psIgE $\geq$0.35 kU l^{-1} or SPT $\geq$3 mm	278	1.8	1.4–2.2	2.0×10^{-7}	1.0×10^{-7}	1.8	1.4–2.2	1.8×10^{-7}	2.5×10^{-9}
CR in 2 h, plus psIgE $\geq$0.35 kU l^{-1} or SPT $\geq$5 mm	276	1.7	1.4–2.2	4.2×10^{-7}	3.6×10^{-7}	1.7	1.4–2.2	3.0×10^{-7}	6.7×10^{-9}
CR in 2 h, plus psIgE $\geq$15 kU l^{-1} or SPT $\geq$8 mm	216	1.8	1.4–2.3	7.5×10^{-7}	1.4×10^{-6}	1.8	1.4–2.3	7.0×10^{-7}	8.9×10^{-8}
CR in 2 h, plus psIgE $\geq$57 kU l^{-1} or SPT $\geq$8 mm	187	1.8	1.4–2.3	4.2×10^{-6}	2.3×10^{-5}	1.8	1.4–2.2	9.2×10^{-6}	4.2×10^{-6}

CI, confidence interval; CR, clinical allergic reaction to peanut ingestion; GEE, generalized estimating equation; MQLS, modified quasi-likelihood score; OR, odds ratio; PA, peanut allergy; psIgE, peanut-specific IgE; SNP, single-nucleotide polymorphism; SPT, skin prick test.

*Using the minor allele (T allele) as the effective allele and the major allele (G allele) as the reference allele.

†Using the minor allele (C allele) as the effective allele and the major allele (T allele) as the reference allele.

‡The GEE model was applied to estimate the effect size of each SNP (additive genetic model) on the risk of PA in PA cases versus 733 non-PA children, with adjustment for age and gender, and controlling for within-family relationships.

§The MQLS test was applied to estimate the *P*-value by comparing PA cases, 144 non-allergic, non-sensitized normal controls and 1,737 controls of uncertain phenotype.

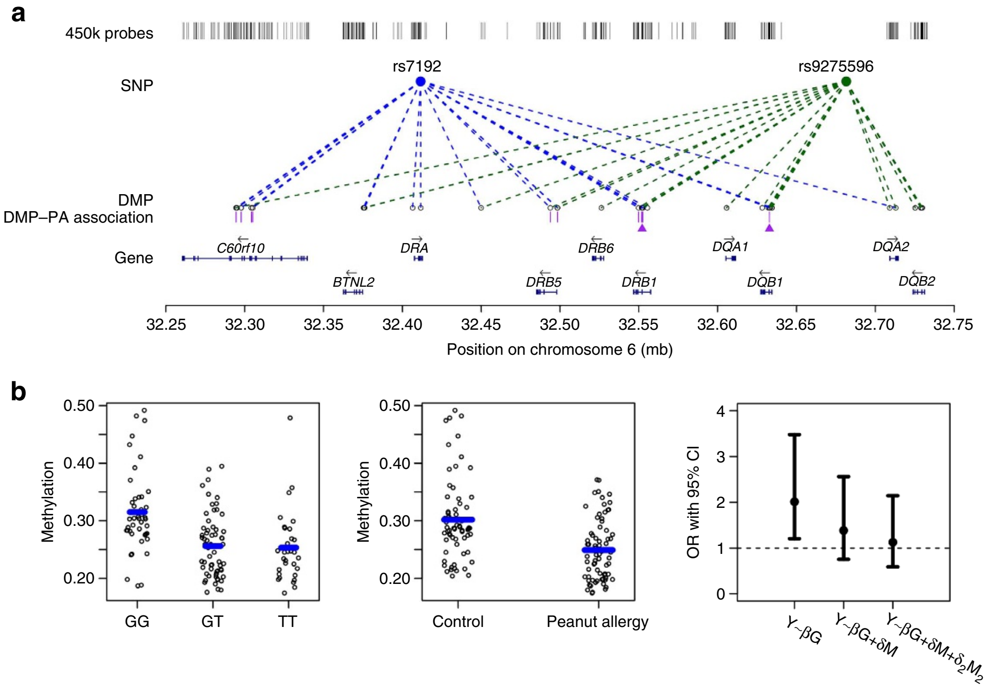

Figure 4 | DMPs associated with rs7192, rs9275596 and with PA, as well as DMPs that mediate genetic risk in PA. (**a**) Diagram showing associations between genotype (rs7192 (blue circle) and rs9275596 (green circle)) and DMPs, and between the genotype-dependent DMPs and PA. The DMPs are denoted by empty circles. Dashed blue (for rs7192) and green (for rs9275596) lines represent significant associations between genotypes and DMPs at $P < 5 \times 10^{-8}$ based on linear regression models ($N = 218$). In between the DMPs and gene tracks are purple bars that represent the genotype-dependent DMPs that are associated with PA based on linear regression models in 73 PA cases and 67 non-PA controls (DMPs as the outcomes). The two DMPs marked with purple triangles are DMP cg15982117 in the *HLA-DRB1* gene and DMP cg18024368 in the *HLA-DQB1* gene that may mediate genetic risk in PA, as determined by the CIT. (**b**) DMP cg18024368 in the *HLA-DQB1* gene that may mediate the association between the rs7192 genotype and PA risk in 73 PA cases and 67 non-PA controls. The left panel shows the association between the methylation level at DMP cg18024368 and the rs7192 genotype. The middle panel depicts the association between the methylation level at DMP cg18024368 and PA. The blue bars in the left and middle panels represent median methylation levels. The right panel shows the effect size of the observed rs7192–PA associations (represented by ORs based on the logistic regression model) before and after adjusting for the methylation level at DMP cg18024368 (*M*), as well as adjusting for both the methylation level at DMP cg18024368 (*M*) and at DMP cg15982117 (*M2*). Error bars represent the 95% CIs of the estimated ORs.

SNP–PA association via a causal inference test (CIT) (Table 4). Briefly, to be a qualified mediator, this CIT must meet the following criteria[32,40]: (i) genotype and PA are associated; (ii) genotype is associated with DMP independent of PA; (iii) DMP is associated with PA independent of genotype; and (iv) genotype is not independently associated with PA after adjusting for DMP. We found cg15982117 in the *HLA-DRB1* gene and cg18024368 in the *HLA-DQB1* gene significantly mediated the effects of rs7192 and rs9275596 on PA (Table 4, all $P_{CIT} < 0.005$) after Bonferroni correction for four different DMPs and two different SNPs. As an example, Fig. 4b shows that DMP cg18024368 in the *HLA-DQB1* gene was significantly hypomethylated in children carrying the T risk allele at rs7192 ($P = 2.7 \times 10^{-10}$; Fig. 4b, left panel) and also in PA cases ($N = 73$) compared with non-PA controls ($N = 67$, $P = 2.4 \times 10^{-6}$; Fig. 4b, middle panel). The estimated OR for the rs7192–PA association was substantially reduced after adjusting

Table 4 | Differentially methylated positions that mediate genetic risk in PA.

DMP†	DMP annotation			DMP-rs7192 associations§		DMP-rs9275596 associations§		DMP-PA associations¶		*P* for the CIT test††	
	Position	Nearest gene	Location‡	β‖	*P*	β‖	*P*	β#	*P***	rs7192	rs9275596
cg17039645	32294503	*C6orf10*	Gene body	0.04	4.4×10^{-14}	0.03	7.1×10^{-9}	0.05	5.4×10^{-7}	0.162	0.155
cg18111114	32498493	*HLA-DRB5*	TSS1500	0.10	5.5×10^{-25}	0.09	3.2×10^{-22}	0.10	1.6×10^{-5}	0.098	0.037
cg15982117	32552106	*HLA-DRB1*	Gene body	−0.09	8.9×10^{-10}	−0.09	3.1×10^{-10}	−0.15	3.8×10^{-7}	0.002*	0.002*
cg18024368	32632848	*HLA-DQB1*	Gene body	−0.04	2.7×10^{-10}	−0.05	3.1×10^{-13}	−0.06	2.4×10^{-6}	0.002*	0.003*

CIT, causal inference tests; DMP, differentially methylated position; GWAS, genome-wide association study; PA, peanut allergy; SNP, single-nucleotide polymorphism; TSS, transcription start site.

*$P<0.05$ after Bonferroni correction for four tested DMPs and two SNPs (cutoff: $P<0.006$ for 8 tests).

†Represented by Infinium HumanMethylation450 BeadChip probe name. For each gene, only the top DMP that was genome-wide significantly associated with both SNPs ($P<5\times10^{-8}$) and that had the strongest association with risk of PA is shown.

‡Location of the methylation CpG site, in relation to the nearest gene.

§Linear regression models were applied to test SNP–DMP associations in a subset of 218 children of European ancestry with available GWAS data and DNA methylation data, with adjustment of age, gender and the estimated cell composition.

‖Adjusted methylation difference with an increase of one copy of the risk allele (T allele for rs7192; or C allele for rs9275596).

¶Linear regression models were applied to test the DMP-PA associations in 73 PA cases and 67 controls of European ancestry, with adjustment of age, gender and the estimated cell composition.

#Adjusted methylation difference between 73 PA cases and 67 controls of European ancestry.

**$P<0.00069$ (=0.05/72) represents the significance level after Bonferroni correction.

††Causal inference tests (CIT) were performed for the 4 DMPs in 73 peanut allergic cases and 67 controls of European ancestry using the modified version for binary outcomes.

for DMP cg18024368, suggesting that this DMP acts as a mediator ($P_{\mathrm{CIT}}=0.002$, Bonferroni adjusted $P_{\mathrm{CIT}}=0.016$; Table 4). Similar associations were found for DMP cg15982117 in the *HLA-DRB1* gene (Table 4 and Supplementary Fig. 3), and the estimated OR for the rs7192–PA association was close to 1.0 after adjusting for both DMPs (cg18024368 and cg15982117; Fig. 4b, right panel).

Additional supporting analyses. To maximize the study power, we performed genome-wide meta-analysis for PA across the discovery and the replication samples of European ancestry (total $N=2{,}328$) using 5,693,167 imputed and genotyped autosomal SNPs (see Methods). Imputed SNP rs33980016 in the *HLA-DQB1* gene remained the top SNP for PA ($P_{\mathrm{meta}}=2.9\times10^{-12}$), followed by imputed SNP rs9273841 ($P_{\mathrm{meta}}=5.2\times10^{-11}$) and genotyped SNP rs9275596 ($P_{\mathrm{meta}}=6.3\times10^{-11}$; Supplementary Table 2 and Supplementary Fig. 2). SNP rs7192 was also significantly associated with PA ($P_{\mathrm{meta}}=2.8\times10^{-9}$) (Table 2). These findings further support that the 6p21.32 region is significantly associated with the risk of PA. No additional significant/suggestive signals outside of the 6p21.32 region were identified.

Given the unique LD pattern in the HLA region, we imputed the classical HLA alleles in individuals of European ancestry using the HLA*IMP framework[41]. In the discovery stage, 50 HLA alleles in the major histocompatibility complex (*MHC*) class II region with a frequency of 0.02 or above were analysed for possible association with PA. *HLA-DQA1*0102* was the only risk allele significantly associated with PA ($P=4.5\times10^{-8}$). This was confirmed in the replication sample of European ancestry ($P=0.025$) and in meta-analysis ($P_{\mathrm{meta}}=5.0\times10^{-9}$; Supplementary Table 5). To compare the relative effects of different HLA variants in association with PA, we imputed classical two-digit and four-digit HLA alleles, and amino acid (AA) polymorphisms in the MHC class II region, using the SNP2HLA framework[42]. The four-digit classical HLA alleles imputed by SNP2HLA were concordant with those imputed by HLA*IMP and *HLA-DQA1**0102 continued to be significantly associated with PA. A two-digit classical HLA allele, *HLA-DQB1*06*, was also significantly associated with PA in both the discovery and the replication samples of European ancestry (Supplementary Table 5). Similar associations were found for these two genes when AA polymorphisms were analysed instead (Supplementary Table 5). Among all of the classical HLA alleles and AA polymorphisms tested, the AA polymorphism at position 71 in the *HLA-DRB1* gene showed the top association signal with PA in both the discovery stage ($P=2.3\times10^{-10}$) and in the meta-analysis ($P_{\mathrm{meta}}=9.8\times10^{-11}$), in which the presence of *Arg* at position 71 was associated with a decreased risk of PA (Supplementary Table 5). A similar association trend was also found for this polymorphism in the replication sample, although it was statistically insignificant ($P=0.189$). All of these associations were greatly reduced by conditioning on rs9275596 or rs7192 ($P>0.008$), suggesting that these identified classical HLA alleles and/or AA polymorphisms may not be independent of the two validated PA-associated SNPs reported here.

By querying existing expression quantitative trait loci (eQTL) databases for populations of European ancestry, we found both rs7192 and rs9275596 were significantly associated with expression levels of the *HLA-DRA*, *HLA-DRB5*, *HLA-DRB1*, *HLA-DQA1*, *HLA-DQB1* and *HLA-DQA2* genes in subcutaneous/omental adipose[43], liver[43] and lymphoblastoid cell lines (http://regulome.stanford.edu; Supplementary Table 6)[44–46]. We then explored the most significant *cis*-eQTLs for these six genes in subcutaneous/omental adipose and liver cell lines, separately (Supplementary Table 7). SNP rs7192 was in almost complete LD ($r^2=0.99$) with rs3763327 (an intergenic SNP between the *HLA-DRA* and *HLA-DQB9* genes), the most significant *cis*-eQTL of the *HLA-DRA* gene in the liver, and the association between rs7192 and *HLA-DRA1* gene expression ($P=1.9\times10^{-16}$) was comparable to that of rs3763327($P=1.8\times10^{-16}$). These results indicate that rs7192 is a potential *cis*-QTL of the *HLA-DQA1* gene in the liver (Supplementary Table 7). In contrast, rs9275596 was in low-to-moderate LD ($r^2<0.61$) with the most significant eQTL for each gene–tissue combination.

Discussion

There is growing evidence that genetic factors may play a role in FA; however, there is a particular lack of knowledge regarding the genetic and epigenetic underpinnings of FA as a whole and also its subtypes[47]. We conducted the first GWAS study of well-defined FA, including specific subtypes (peanut, milk and egg) in US children, a particularly important age group given that FA most commonly develops in early childhood. In addition, this is the first study to demonstrate the key role of differential DNAm in mediating identified genetic risk factors for PA. Specifically, we identified and replicated genetic variants in the *HLA-DR* and *-DQ* gene region that were significantly associated with PA in children of European ancestry, tagged by rs7192 (a non-synonymous SNP of the *HLA-DRA* gene) and/or rs9275596 (intergenic between the *HLA-DQB1* and *HLA-DQA2* genes). Both rs7192 and rs9275596 significantly affect DNAm in several nearby genes. DNAm in the *HLA-DRB1* and *HLA-DQB1* genes, in turn, mediate the detected SNP–PA associations. Taken together with a PAR of 19%–21%, this GWAS indicates the possibility that the *HLA-DR* and *-DQ* gene region likely poses the single greatest genetic risk for PA.

The role of HLA variants in PA has been examined by previous candidate gene studies via direct assessment of HLA classical alleles[12–14,16–18], but the results were inconclusive, partly due to relatively small sample sizes and inadequate control of potential population stratification. This GWAS provides convincing evidence that the *HLA-DR* and *-DQ* gene region, as tagged by rs7192 or rs9275596, harbours significant genetic risk for PA in subjects of European ancestry. These two SNPs may represent one single risk factor, as the association between rs9275596 and PA was not independent of rs7192. To maximize the study power, we conducted a meta-GWAS across the discovery and the replication samples on a combined set of genotyped and imputed SNPs, which further supports this *HLA-DR* and *-DQ* gene region as a single significant region for PA.

Our findings are biologically plausible. The HLA-DR and -DQ molecules, which are expressed in a range of cells, including B cells, activated T cells and the monocyte/macrophage lineage, are known to play a critical role in the development of allergy[48]. These molecules present antigen-derived peptides, mostly of exogenous origin, to CD4+ helper T cells. Antigen presentation by HLA molecules is a defining step in the development of antigen-specific immune responses. These molecules have extensive molecular polymorphisms confined to the peptide-binding groove. These polymorphisms may determine which antigen-derived peptides are bound and presented to T cells via T-cell receptors, and may account for allergen-specific sensitivities. The top PA-associated AA polymorphism, at position 71 in the *HLA-DRB1* gene (imputed by SNP2HLA), is one such polymorphism located in the peptide-binding groove and may partly account for the identified associations. Previous studies have shown that this AA position, together with positions 13, 70 and 74, play important roles in the binding-specificity profile of pocket 4, which is one of the most important pockets for antigen interaction and presentation by the HLA-DR molecule[49]. Another possible explanation for our findings is that rs7192, a missense SNP in the *HLA-DRA* gene, may directly affect HLA-DRA protein function and/or expression and thus affect binding of HLA molecules with peanut allergens. This SNP was significantly associated with *HLA-DRA* gene expression in multiple tissues including adipose, liver and lymphoblastoid cell lines, which probably represent associations in antigen presenting cells throughout the body. In the liver, rs7192 may be an actual eQTL of the *HLA-DRA* gene, because (1) it is in almost complete LD with the top *cis*-eQTL (rs3763327) of the *HLA-DRA* gene ($r^2 = 0.99$) and (2) it induces a Leu242Val change in the *HLA-DRA* gene, while rs3763327 is an intergenic SNP between the *HLA-DRA* and *HLA-DQB9* genes with unknown functionality. The liver has been demonstrated to play a critical role in oral tolerance induction[50]. It would be of great interest to further examine these associations in additional tissues such as the intestinal mucosa and skin, which are likely to be critical to the pathogenesis of PA. We still cannot exclude the possibility that other untyped variants that are in high LD with rs7192 or rs9275596 could be the causal SNP(s). Although the imputed SNP, rs33980016, showed the strongest association with PA, our conditional analyses indicate that this top SNP signal is not distinguishable from the signals at rs9275596 or rs7192 with our current data. Our data provide strong evidence that the 6p21.32 region poses significant risk for PA. However, targeted sequencing in this region is needed to more precisely identify and validate the causal variant(s) for PA.

There is growing evidence that genotype may control DNAm levels[30,31]. DNAm, which regulates gene expression, might influence disease development in a manner complementary to direct mutation of the DNA sequence itself. A recent small epigenome-wide association study identified DMPs in the *HLA-DQB1* gene for IgE-mediated FA[51]. Genetic and epigenetic modification may also interact biologically[52]. We showed that both rs7192 and rs9275596 were methylation quantitative loci for the *HLA-DRB1* and *HLA-DQB1* genes, and that there are significant causal relationships among the genotypes, DNAm and the risk of PA. This indicates that DNAm may regulate the expression levels of these genes and, subsequently, may partly mediate the genetic risk of PA. A similar linkage was observed for rheumatoid arthritis in a recent study, although the involved SNPs and DMPs in the HLA region were not the same as those identified in our study[32]. Given the genetic associations with the HLA region and disease pathogenesis that have already been linked to specific HLA protein epitopes, the methylation mediation observed here implies an additional complementary mechanism by which the HLA variants may influence PA. This study represents the first step in understanding the role of DNAm mediation effects on PA. Our findings provide clues and underscore the need for additional functional studies, including follow-up data on independent PA subjects in a clinical setting, to show how genotype-dependent DNAm could regulate the expression of key genes, and how these expression patterns may correlate with clinical outcomes. Longitudinal cohort studies on DNAm at multiple time points are also needed to assess dynamic changes in DNAm and its temporal relationship with the risk of PA.

We showed that the identified SNP–PA associations were not observed for milk allergy, egg allergy or other allergic phenotypes (including allergic sensitization, self-reported physician-diagnosed asthma, eczema and allergic rhinitis; Supplementary Table 3), raising the possibility that PA may be under distinct genetic control. However, this study is underpowered on a genome-wide level, both for 'any FA' and for its subtypes. The finding of an association with PA should be interpreted cautiously in the context of limited study power. The lack of association between the two PA-associated SNPs, and other FA and/or other allergic phenotypes may be due to limited sample size and limited study power, hence requiring further studies in a larger sample.

Our study is also limited in the following aspects. As is the case in most GWAS studies, we examined the genetic associations for common variants. Substantially larger sample sizes will be required to identify rare variants or common variants with small effects. We replicated our GWAS results in an independent set of samples from the same Chicago Food Allergy Study. Our findings could be further strengthened by additional replications in other independent populations. The significant associations that we identified between rs7192 and rs9275596 and PA were not present in the participants of non-European ancestry; however, we cannot firmly conclude that these effects are specific to population of European ancestry due to a limited sample size of non-Europeans.

Batch effect is one of the major problems often encountered in epigenetic studies[53]. We used ComBat[54] transformation of methylation data to minimize potential confounding by batch effects. After transformation, the average DNAm level of the reported DMPs in PA cases or controls did not vary significantly across plates. Another issue linked to epigenetic studies is related to tissue- and cell-specific characteristics. FA is a systemic condition for which the study of methylation patterns in blood may be feasible, although cell heterogeneity in blood may act as a potential confounder[55,56] due to the cell-specific pattern of DNAm[57]. Accordingly, we adjusted for estimated cell composition using the 'limma' package[58] in all epigenetic association tests. However, we could not exclude the possibility of some residual confounding. The identified DNAm mediation effects will require replication and verification in future studies.

In summary, this GWAS of FA revealed one significant peak at 6p21.32 for PA, and the finding appears to be consistent based on analyses of genotyped SNPs, imputed SNPs, imputed classical HLA alleles and AA polymorphisms, and a meta-GWAS across the discovery and replication samples. Specifically, this study identified PA-specific susceptibility loci in the *HLA-DQ* and *-DR* region at 6p21.32, tagged by rs7192 and rs9275596. Both SNPs were associated with differential DNAm levels at multiple CpG sites and differential DNAm of the *HLA-DQB1* and *HLA-DRB1* genes partially mediated the identified SNP–PA association. Taken together with a PAR of 19%–21%, this study indicates the possibility that the *HLA-DR* and *-DQ* gene region likely poses the single greatest genetic risk for PA. Findings from this study warrant additional replication, validation and functional studies, which will have the potential to improve our understanding of the genetic factors and epigenetic mechanism underlying the risk of PA, and may inform future development of new strategies for the prediction, prevention and treatment of PA.

Methods

The Chicago Food Allergy Study. Both the discovery and replication samples were enrolled as part of the Chicago Food Allergy Study under a standard study protocol. All participants were recruited from the Chicago area from August 2005 to June 2011. Eligible families were those having either one or both parents with at least one biological child (aged 0–21 years) with or without FA, willing to participate in the study. Eligible FA case or control children (aged 0–21 years) were those with or without FA. For each family or participant, the following procedures were completed: (1) questionnaire interview by trained research staff to obtain information on each family member's home environment, diet, lifestyle, history of FA and other allergic diseases; (2) clinical evaluation by nurse or trained research staff to obtain height, weight, waist and hip circumference, blood pressure measurement and lung function test; (3) allergy SPT; and (4) collection of venous blood samples for food specific IgE (sIgE) measurement, DNA extraction and subsequent laboratory assays. Detailed information on SPT and sIgE measurement is given in the Supplementary Methods. For each child, we also collected a detailed history of clinical allergic reaction on ingestion of specific foods. The study protocol was approved by the Institutional Review Board of Ann and Robert H. Lurie Children's Hospital of Chicago and the Institutional Review Board of Johns Hopkins Bloomberg School of Public Health. Written informed consents were obtained from all participants or their legal guardian (for children aged <18 years).

Study sample included in the current GWAS of FA. In the discovery stage, we primarily used samples from nuclear families. A total of 2,759 subjects (853 families) were included. Among these families, 780 families ($n=2,678$) were included based on the following criteria: (1) at least one child had a convincing history of clinical allergic reaction on ingestion of specific foods and (2) two or more additional family members (parents/siblings) had archived DNA samples. Another 81 children from 73 families without parental genotyping data were also included (29 FA cases and 52 controls).

In the replication stage, we aimed to replicate the identified genetic associations with PA. We included 216 case–control samples (88 PA cases and 128 controls) from the Chicago Food Allergy Study, all independent of the discovery sample.

Definitions of phenotypes of interest. The main phenotypes of interest included 'any FA' and the three most common types of FA: PA, egg allergy and milk allergy. As we reported previously, we adopted stringent clinical criteria to define a specific type of FA[35]: (1) a convincing history of clinical allergic reaction on ingestion of specific foods[35] and (2) evidence of sensitization to the same food, defined as having a detectable sIgE ($\geq 0.10\,\mathrm{kU\,l^{-1}}$; detection limit of the instrument was $<0.10\,\mathrm{kU\,l^{-1}}$) and/or a positive SPT to this specified food. A positive SPT for a specific allergen was defined based on criteria as follows: (1) the MWD for the negative control was <3 mm, the positive control was ≥ 3 mm and the difference of positive minus negative control was ≥ 3 mm; and (2) MWD was ≥ 3 mm for the specified allergen. Accordingly, we defined allergy to nine common foods (peanut, egg white, cow's milk, soy, wheat, walnut, fish, shellfish and sesame seed) and 'any FA' if a child was allergic to any of these foods. Normal controls were defined if a child had neither clinical allergic reaction nor evidence of sensitization to any of the nine foods. All parents were defined as having uncertain FA phenotypes, as data on history of clinical allergic reaction subsequent to ingestion of specific foods were unavailable. We also performed sensitivity tests on FA definitions using other cutoffs for food-specific IgE and SPT, for example, food-specific IgE $\geq 0.35\,\mathrm{kU\,l^{-1}}$, SPT MWD ≥ 5 mm (ref. 36) or either food-specific IgE or SPT MWD $\geq 95\%$ PPV.

Genotyping and quality control steps in the discovery GWAS. Genomic DNA was isolated from EDTA-treated peripheral white blood cells. The concentration and purity was determined using a Quant-iT Broad-range dsDNA Assay Kit on a SpectraMax M2 micro-plate reader. Genotyping was performed using the Illumina HumanOmni1-Quad BeadChip in the Genome Technology Access Center, Washington University in St Louis, MO, according to specifications listed in Illumina's protocol (Illumina, Inc.). Among 2,759 genotyped samples, 12 failed to yield high-quality genotyping calls (Supplementary Methods), resulting in an overall genotyping success rate of 99.6%.

Genotypes for 2,747 subjects were exported, with a total of 1,011,859 SNPs. We performed rigorous quality-control steps as suggested by Laurie *et al.*[59], using the R/bioconductor package 'GWASTools'[60]. Briefly, we examined the following parameters: (1) missing call rate per SNP, per chromosome and per sample; (2) the reproducibility rate among the 100 duplicated samples; (3) duplicate discordance estimates for each SNP to infer SNP quality; 4) genotyping batch effects: measured by comparing the difference in allelic frequencies between each plate and a pool of the other plates, and by comparing variation in $\log_{10}$ of the autosomal missing call rate in each plate (no significant batch effects were detected); (5) gender identity: based on X chromosome heterozygosity and the means of the intensities of SNP probes on the X and Y chromosome; (6) autosomal heterozygosity; (7) Hardy–Weinberg equilibrium (HWE) test: performed among self-reported Caucasian parents or a sibling without FA if no parent was available. Sex-specific HWE tests were also performed; (8) Mendelian error check of 650 families with both parents available; and (9) pair-wise sample relatedness: pair-wise kinship estimates between every subject were computed using PLINK[61].

We filtered 45,100 monomorphic SNPs and 14,948 SNPs with a $>5\%$ missing genotyping rate. A total of 595 SNPs with duplicate discordance estimates $>2\%$ in 98 pairs of duplicates and 1,784 SNPs that deviated from the HWE test ($P<1\times10^{-6}$) were also filtered. Mendelian error checks filtered 2,145 SNPs with Mendelian errors in ≥ 10 families ($>1.5\%$ families). Some SNPs were filtered under more than one criterion. We also removed 162,283 SNPs with minor allele frequency (MAF) $<2\%$ and 2,086 SNPs on the Y chromosome or on mitochondrial chromosomes. Finally, a total of 772,141 autosomal SNPs and 17,536 SNPs on the X chromosome were used in the downstream analyses.

We removed one subject with a missing genotyping call rate $>5\%$, 12 subjects with gender discrepancies and 6 subjects with Mendelian errors in $>5,000$ SNPs. Pair-wise relatedness was checked for each pair of subjects by plotting the proportion of loci where the pair shared one allele identical by descent versus the proportion of loci where the pair shared zero allele identical by descent. A total of 34 subjects for whom the degree of relatedness was inconsistent with self-reported relationship were then removed. In total, 2,694 subjects were available for downstream data analyses.

Genetic ancestry was carefully computed by PCA using Eigenstrat[37] and all European, American, African and Asian individuals in the 1000 Genomes Project were used as a reference (phase I, release_v3.20101123), as detailed in Supplementary Methods.

Statistical analyses in the discovery GWAS. To leverage the family-based data with a small number of case–control samples, the MQLS test (for autosomal markers)[33] and its' extension, the XM test[34] (for X-linked markers) were applied to explore genetic associations for each phenotype of interest using MQLS-XM (http://www.stat.uchicago.edu/~mcpeek/software/MQLS_XM/download.html), a programme for dichotomous outcomes. The MQLS can maximally use information available in a complex family structure by the following: (1) distinguishing between unaffected controls and controls of uncertain phenotypes (that is, individuals with unmeasured phenotypes), and incorporating both into the analyses; and (2) incorporating phenotype data for relatives with missing genotype data at each marker tested[33]. MQLS is a retrospective score test that treats the genotype data on sample individuals as random and the available phenotype information as fixed in the analysis, thus allowing for valid association testing in the presence of phenotype misspecification, and hence the method provides high power at the appropriate type I error rate[33]. Before the MQLS analysis, using PA as an example, the phenotype of interest was coded as follows: (1) PA-affected cases, (2) non-allergic non-sensitized normal controls and (3) controls of uncertain phenotypes (including children who did not meet the PA case or normal control definition, and all genotyped parents). We also performed a sensitivity test and found that the results were not significantly altered by removing children who did not meet the PA case or normal control definition from the analysis. The MQLS test was performed under an additive genetic model, with a specified prevalence of 5% for any FA or 1% for PA, milk allergy and egg allergy, separately, in the Europeans. We also repeated the analyses, while specifying a higher prevalence (10% for any FA, or 5% for PA, milk allergy and egg allergy, separately) and obtained very similar results. To perform MQLS analyses by conditioning on one of the top SNPs, we first calculated the residual using logit($Y=1$) $=\beta_0+\beta_G{}^*G$ for subjects with non-missing phenotypes, where Y is the disease status and G is the genotype of the selected top SNP. The residual was set to 0 for subjects with missing phenotype. Similarly, to perform MQLS analyses in 497 non-European subjects adjusting for ancestry, the residual PA status for subjects with non-missing phenotypes was calculated using the first three PCs from the GWAS genotyping data as covariates and the residual for subjects with missing phenotypes was set to 0. The calculated residual was then used as the outcome to perform MQLS analyses using the QM-QXM programme, an approach that is an extension of the MQLS test to quantitative traits (http://faculty.washington.edu/tathornt/software/QM_QXM/). As the MQLS is a score test and does not estimate effect size, the reported OR and 95% CIs were estimated using GEE models, with adjustment for age and gender in subjects of European ancestry. The first three PCs were also adjusted in the analyses for non-European subjects.

Genotyping and data analyses in the replication sample. The replication sample consisted of 88 PA cases and 128 normal controls from the same Chicago Food Allergy Study. SNPs rs7192 and rs9275596, which were suggestively or significantly associated with PA in the discovery GWAS ($P<1\times10^{-7}$), were selected for replication. As we needed to impute population ancestry using a similar strategy as was used for the discovery sample, and impute classical HLA alleles and AA polymorphisms based on a relatively dense SNP set, the Human OmniExpressExome BeadChip was selected for genotyping. DNA samples were prepared using the same lab procedures as for the discovery sample, and cases and controls were distributed evenly in each plate. Genotyping was performed according to specifications listed in Illumina's protocol (Illumina, Inc.) at the Genomics Core Facility of the Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai. Similar quality-control steps were applied to the replication sample. Two subjects with gender inconsistencies and one subject from a monozygotic twin pair were removed from the subsequent data analysis. GEE models, adjusting for age and gender, were performed to test the association between each SNP and PA under an additive genetic model, in samples with European ancestry. When analyses were performed in samples of non-European ancestry, the first three PCs from the genome-wide SNP genotyping data were also included as covariates to adjust for potential population stratification.

SNP imputation. In the discovery sample, we performed phasing using SHAPEIT[38] and imputation using IMPUTE2 (ref. 39) with all individuals in the 1000 Genomes Project as a reference panel. As MQLS does not support analyses using posterior probabilities, we computed best-guess genotypes, using a probability threshold of 0.95, as recently described in the literature[62]. We applied several post-imputation quality-control metrics including removal of SNPs with an IMPUTE2 info score <0.8, with a missing call rate >0.05, or with a MAF <0.02. A total of 6,459,842 genotyped or imputed SNPs were then analysed for their associations with any FA and three specific types of FA, respectively, using the MQLS test.

Meta-GWAS. SNP imputation was also performed for the replication sample of European ancestry, leading to a combined set of 6,174,271 genotyped or imputed SNPs. We performed the association tests for PA in the replication sample using the GEE model (in a case–control setting), adjusting for age and gender, similar to what was done for genotyped SNPs rs7192 and rs9275596. To maximize power, we performed meta-analysis based on the Stouffer's weighted z-score method to combine the association results for PA from the discovery and the replication samples. Our GEE analyses in the replication sample did not include SNPs on the X chromosome ($N=139{,}697$) because of the small replication sample size and need to perform gender-specific GEE analyses (57 females and 74 males), and thus the meta-analysis was limited to 5,693,167 autosomal SNPs that overlapped in both sample sets.

Imputation of HLA alleles and AA polymorphisms. We used the HLA*IMP[41] programme to impute classical HLA alleles from SNP genotyping data via reference to a training data set of over 2,500 samples of European ancestry with dense SNP data and classical HLA allele types. This framework is reported to have high imputation accuracy (92%–98% of imputations agree with lab-derived HLA types)[41]. We also applied the SNP2HLA framework to impute AA polymorphisms as well as classical HLA alleles, with genotype data from the Major Histocompatibility Complex Working Group of the Type I Diabetes Genetics Consortium as a reference panel[42]. The imputation was performed for subjects of European ancestry in the *MHC class II* gene region. We used best-guess genotypes for analyses, as MQLS does not support analysis using posterior probabilities. After applying several quality-control filters to the imputed data (that is, removal of imputed variants with call rate $<95\%$ and/or MAF <0.02), 50 four-digit HLA alleles from the HLA*IMP programme, 27 two-digit HLA alleles, 41 four-digit alleles and 165 AA polymorphisms from the SNP2HLA programme were analysed for their associations with PA, using the MQLS test in the discovery sample and using the GEE model in the replication sample, as described above. Multiallelic AA polymorphisms were analysed for associations with PA after converting K-alleles to K-bialleles.

DNA methylation measurement and quality-control steps. A total of 218 unrelated children of European ancestry in the discovery ($N=199$) or replication ($N=19$) samples had genome-wide DNAm data measured in genomic DNA isolated from EDTA-treated peripheral white blood cells. DNAm was measured using Infinium HumanMethylation450 BeadChips (including $>485{,}000$ CpG sites), according to the manufacturer's instructions at the Center for Genetic Medicine, Northwestern University Feinberg School of Medicine. Several quality-control steps were performed with the 'minfi' framework[63], as detailed in the Supplementary Methods. Both β-and M-values (representing methylation ratios) were computed for downstream analyses. M-values are reported as superior to β-values for identification of differential methylation[64].

To account for potential batch effects, β- and M-values were ComBat-transformed using the 'sva' package[65], with chip number as the surrogate for batches. The ComBat-transformed β- and M-values at each CpG site were applied to explore associations between DNAm, genotypes and PA.

Cell heterogeneity in blood may act as a potential confounder[55,56] due to cell-specific patterns of DNAm[57]. Thus, with estimateCellCounts() function included in the 'minfi' package[63], the distribution of six cell types (CD8T cells, CD4T cells, NK cells, B cells, monocytes and granulocytes) was inferred for each sample based on external reference DNAm signatures of the constituent cell type from Illumina HumanMethylation450 BeadChips[55,56]. The estimated cell composition was adjusted as a covariate in subsequent analyses.

Statistical analyses on DNA methylation mediation effects. To identify DMPs associated with the two validated PA-associated SNPs, we applied the 'limma' package[58] in R/bioconductor, to fit a linear regression model in 218 unrelated children of European ancestry, with ComBat-transformed M-values at each CpG site ($N=456{,}513$) as a function of each SNP (under an additive genetic model), adjusted for age, gender and estimated cell composition. Genome-wide significance ($P<5\times10^{-8}$) cutoffs were applied. To report adjusted methylation differences in each genotype, ComBat-transformed β-values were analysed instead of ComBat-transformed M-values, which did not significantly change the results.

The identified genotype-dependent DMPs were tested for associations with PA, by fitting a linear regression model with ComBat-transformed M-values as outcomes, adjusting for the covariates mentioned above. These analyses were conducted in 73 PA cases and 67 controls, while the remaining 78 children with uncertain PA phenotypes were removed from these analyses. Bonferroni correction was applied to adjust for multiple testing. To report adjusted methylation differences in each group, ComBat-transformed β-values were analysed instead of ComBat-transformed M-values, which did not significantly change the results.

The SNP–DMP–PA relationships were then assessed using the CIT classification as methylation mediated, consequential, or independent[40]. We focused on the top DMP from each gene that was significantly associated with both SNPs and PA risk. Briefly, the CIT performs statistical tests for four conditions, all of which must be met to conclude that methylation mediation is occurring: (i) genotype and phenotype of interest (PA in the current study) are associated; (ii) genotype is associated with DMP after adjusting for phenotype; (iii) DMP is associated with phenotype after adjusting for genotype; and (iv) genotype is independent of phenotype after adjusting for DMP. The CIT P-value is defined using the intersection-union test framework as the maximum of the four-component test P-values. As the CIT was originally designed for continuous phenotypes, we applied a modified version based on logistic regression to examine the causal relationship for each SNP–DMP–PA pair in this study, which has been reported previously[32].

Functional annotation using existing eQTL data sets. To identify potentially causal gene(s) underlying the identified genetic associations with PA, we queried existing eQTL databases in multiple tissues (including subcutaneous/omental adipose tissue[43], liver tissue[43] and lymphocytes[44–46] (http://regulome.stanford.edu)), to assess whether the top PA-associated SNPs were eQTL SNPs. We surveyed both *cis*- and *trans*-eQTLs of 10% false discovery rate and found that the two PA-associated SNPs influence gene expression mainly in a *cis* manner, and that corresponding *cis*-eQTLs were reported in the paper. For each gene whose expression level was significantly associated with the two PA-associated SNPs, the most significant *cis*-eQTL in the subcutaneous/omental adipose and liver tissues, separately, and its LD squared correlation coefficient with the PA-associated SNPs were also reported.

References

1. Boyce, J. A. *et al.* Guidelines for the diagnosis and management of food allergy in the United States: report of the NIAID-sponsored expert panel. *J. Allergy Clin. Immunol.* **126,** S1–58 (2010).
2. Sicherer, S. H. & Sampson, H. A. Food allergy: epidemiology, pathogenesis, diagnosis, and treatment. *J. Allergy Clin. Immunol.* **133,** 291–307 (2014).
3. Longo, G., Berti, I., Burks, A. W., Krauss, B. & Barbi, E. IgE-mediated food allergy in children. *Lancet* **382,** 1656–1664 (2013).
4. Patel, D. A., Holdford, D. A., Edwards, E. & Carroll, N. V. Estimating the economic burden of food-induced allergic reactions and anaphylaxis in the United States. *J. Allergy Clin. Immunol.* **128,** 110–115 e5 (2011).
5. Gupta, R. *et al.* The economic impact of childhood food allergy in the United States. *JAMA Pediatr.* **134**(Suppl 3): S149–S150 (2013).

6. Cohen, B. L., Noone, S., Munoz-Furlong, A. & Sicherer, S. H. Development of a questionnaire to measure quality of life in families with a child with food allergy. *J. Allergy Clin. Immunol.* **114,** 1159–1163 (2004).
7. Cummings, A. J., Knibb, R. C., King, R. M. & Lucas, J. S. The psychosocial impact of food allergy and food hypersensitivity in children, adolescents and their families: a review. *Allergy* **65,** 933–945 (2010).
8. Hong, X., Tsai, H. J. & Wang, X. Genetics of food allergy. *Curr. Opin. Pediatr.* **21,** 770–776 (2009).
9. Tsai, H. J. *et al.* Familial aggregation of food allergy and sensitization to food allergens: a family-based study. *Clin. Exp. Allergy* **39,** 101–109 (2009).
10. Sicherer, S. H. *et al.* Genetics of peanut allergy: a twin study. *J. Allergy Clin. Immunol.* **106,** 53–56 (2000).
11. Dreskin, S. C. *et al.* Association of genetic variants of CD14 with peanut allergy and elevated IgE levels in peanut allergic individuals. *Ann. Allergy Asthma Immunol.* **106,** 170–172 (2011).
12. Shreffler, W. G., Charlop-Powers, Z. & Sicherer, S. H. Lack of association of HLA class II alleles with peanut allergy. *Ann. Allergy Asthma Immunol.* **96,** 865–869 (2006).
13. Howell, W. M., Turner, S. J., Hourihane, J. O., Dean, T. P. & Warner, J. O. HLA class II DRB1, DQB1 and DPB1 genotypic associations with peanut allergy: evidence from a family-based and case-control study. *Clin. Exp. Allergy* **28,** 156–162 (1998).
14. Dreskin, S. C. *et al.* Peanut-allergic subjects and their peanut-tolerant siblings have large differences in peanut-specific IgG that are independent of HLA class II. *Clin. Immunol.* **137,** 366–373 (2010).
15. Amoli, M. M. *et al.* Polymorphism in the STAT6 gene encodes risk for nut allergy. *Genes Immun.* **3,** 220–224 (2002).
16. Boehncke, W. H. *et al.* Identification of HLA-DR and -DQ alleles conferring susceptibility to pollen allergy and pollen associated food allergy. *Clin. Exp. Allergy* **28,** 434–441 (1998).
17. Donovan, G. R. *et al.* A family study of allergy: segregation with HLA but not with T-cell receptor genes. *J. Allergy Clin. Immunol.* **97,** 712–713 (1996).
18. Madore, A. M. *et al.* HLA-DQB1*02 and DQB1*06:03P are associated with peanut allergy. *Eur. J. Hum. Genet.* **21,** 1181–1184 (2013).
19. Brown, S. J. *et al.* Loss-of-function variants in the filaggrin gene are a significant risk factor for peanut allergy. *J. Allergy Clin. Immunol.* **127,** 661–667 (2011).
20. Asai, Y. *et al.* Filaggrin gene mutation associations with peanut allergy persist despite variations in peanut allergy diagnostic criteria or asthma status. *J. Allergy Clin. Immunol.* **132,** 239–242 (2013).
21. Venkataraman, D. *et al.* Filaggrin loss-of-function mutations are associated with food allergy in childhood and adolescence. *J. Allergy Clin. Immunol.* **134,** 876–882 e4 (2014).
22. Torgerson, D. G. *et al.* Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nat. Genet.* **43,** 887–892 (2011).
23. Hirota, T. *et al.* Genome-wide association study identifies three new susceptibility loci for adult asthma in the Japanese population. *Nat. Genet.* **43,** 893–896 (2011).
24. Moffatt, M. F. *et al.* A large-scale, consortium-based genomewide association study of asthma. *N. Engl. J. Med.* **363,** 1211–1221 (2010).
25. Paternoster, L. *et al.* Meta-analysis of genome-wide association studies identifies three new risk loci for atopic dermatitis. *Nat. Genet.* **44,** 187–192 (2012).
26. Sun, L. D. *et al.* Genome-wide association study identifies two new susceptibility loci for atopic dermatitis in the Chinese Han population. *Nat. Genet.* **43,** 690–694 (2011).
27. Hirota, T. *et al.* Genome-wide association study identifies eight new susceptibility loci for atopic dermatitis in the Japanese population. *Nat. Genet.* **44,** 1222–1226 (2012).
28. Bonnelykke, K. *et al.* Meta-analysis of genome-wide association studies identifies ten loci influencing allergic sensitization. *Nat. Genet.* **45,** 902–906 (2013).
29. Hinds, D. A. *et al.* A genome-wide association meta-analysis of self-reported allergy identifies shared and allergy-specific susceptibility loci. *Nat. Genet.* **45,** 907–911 (2013).
30. Drong, A. W. *et al.* The presence of methylation quantitative trait loci indicates a direct genetic influence on the level of DNA methylation in adipose tissue. *PLoS ONE* **8,** e55923 (2013).
31. Liu, Y. *et al.* GeMes, clusters of DNA methylation under genetic control, can inform genetic and epigenetic analysis of disease. *Am. J. Hum. Genet.* **94,** 485–495 (2014).
32. Liu, Y. *et al.* Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat. Biotechnol.* **31,** 142–147 (2013).
33. Thornton, T. & McPeek, M. S. Case-control association testing with related individuals: a more powerful quasi-likelihood score test. *Am. J. Hum. Genet.* **81,** 321–337 (2007).
34. Thornton, T., Zhang, Q., Cai, X., Ober, C. & McPeek, M. S. XM: association testing on the X-chromosome in case-control samples with related individuals. *Genet. Epidemiol.* **36,** 438–450 (2012).
35. Hong, X. *et al.* IgE, but not IgG4, antibodies to Ara h 2 distinguish peanut allergy from asymptomatic peanut sensitization. *Allergy* **67,** 1538–1546 (2012).
36. Nelson, H. S., Rosloniec, D. M., McCall, L. I. & Ikle, D. Comparative performance of five commercial prick skin test devices. *J. Allergy Clin. Immunol.* **92,** 750–756 (1993).
37. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38,** 904–909 (2006).
38. Delaneau, O., Zagury, J. F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* **10,** 5–6 (2013).
39. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5,** e1000529 (2009).
40. Millstein, J., Zhang, B., Zhu, J. & Schadt, E. E. Disentangling molecular relationships with a causal inference test. *BMC Genet.* **10,** 23 (2009).
41. Dilthey, A. T., Moutsianas, L., Leslie, S. & McVean, G. HLA*IMP--an integrated framework for imputing classical HLA alleles from SNP genotypes. *Bioinformatics* **27,** 968–972 (2011).
42. Jia, X. *et al.* Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS ONE* **8,** e64683 (2013).
43. Greenawalt, D. M. *et al.* A survey of the genetics of stomach, liver, and adipose gene expression from a morbidly obese cohort. *Genome Res.* **21,** 1008–1016 (2011).
44. Montgomery, S. B. *et al.* Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* **464,** 773–777 (2010).
45. Stranger, B. E. *et al.* Population genomics of human gene expression. *Nat. Genet.* **39,** 1217–1224 (2007).
46. Veyrieras, J. B. *et al.* High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genet.* **4,** e1000214 (2008).
47. Hong, X. & Wang, X. Early life precursors, epigenetics, and the development of food allergy. *Semin. Immunopathol.* **34,** 655–669 (2012).
48. Robinson, J. H. & Delvig, A. A. Diversity in MHC class II antigen presentation. *Immunology* **105,** 252–262 (2002).
49. Sturniolo, T. *et al.* Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. *Nat. Biotechnol.* **17,** 555–561 (1999).
50. Rescigno, M. Dendritic cells in oral tolerance in the gut. *Cell Microbiol.* **13,** 1312–1318 (2011).
51. Martino, D. *et al.* Epigenome-wide association study reveals longitudinally stable DNA methylation differences in CD4+ T cells from children with IgE-mediated food allergy. *Epigenetics* **9,** 998–1006 (2014).
52. Rakyan, V. K., Down, T. A., Balding, D. J. & Beck, S. Epigenome-wide association studies for common human diseases. *Nat. Rev. Genet.* **12,** 529–541 (2011).
53. Leek, J. T. *et al.* Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat. Rev. Genet.* **11,** 733–739 (2010).
54. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8,** 118–127 (2007).
55. Jaffe, A. E. & Irizarry, R. A. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol.* **15,** R31 (2014).
56. Houseman, E. A. *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* **13,** 86 (2012).
57. Reinius, L. E. *et al.* Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS ONE* **7,** e41361 (2012).
58. Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3,** Article3 (2004).
59. Laurie, C. C. *et al.* Quality control and quality assurance in genotypic data for genome-wide association studies. *Genet. Epidemiol.* **34,** 591–602 (2010).
60. Gogarten, S. M. *et al.* GWASTools: an R/Bioconductor package for quality control and analysis of genome-wide association studies. *Bioinformatics* **28,** 3329–3331 (2012).
61. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).
62. Bigdeli, T. B. *et al.* Molecular validation of the schizophrenia spectrum. *Schizophr. Bull.* **40,** 60–65 (2014).
63. Aryee, M. J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30,** 1363–1369 (2014).
64. Du, P. *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* **11,** 587 (2010).
65. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28,** 882–883 (2012).

Acknowledgements

We gratefully acknowledge the individuals and families who participated in the Chicago Food Allergy Study. This research was supported in part by grants from the Bunning Family and their family foundations, Sacks Family Foundation Fund, Food Allergy Research and Education (FARE) and the National Institute of Allergy and Infectious Diseases (NIAID, PI: X.W., U01AI090727 from the Consortium of Food Allergy Research, R56AI080627 and R21AI088609). Dr Hui-Ju Tsai was supported by a grant MOST 103-2314-B-400-004-MY3. We also acknowledge Drs I-Shou Chang, Wen-Chang Wang and Mr Lan-Chao Wang for their assistance in GWAS data cleaning, and acknowledge Dr Jean-Philippe Fortin for his assistance in DNAm data cleaning.

Author contributions

X.W. is the principal investigator of this GWAS study and has full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. The subject recruitment, data collection and FA phenotyping of the Chicago Food Allergy Study (the parent study) was overseen by X.W. and J.A.P., and conducted by a team of investigators including X.H., H.T., X.L., R.K., R.F., A.M.S., J.S.K., R.E.S., R.S.G., G.W. and R.S., along with a dedicated field team led by D.C., X.X., K.H. and D.E.W. provided advice on study design. G.W., Z.C., X.L., H.T. and X.H. prepared samples for genotyping; X.H. and H.T. assumed primary responsibility for the genome-wide genotyping and data cleaning; K.H. and X.H. assumed primary responsibility for the replication genotyping and data cleaning; X.H., K.H., C.L.-A., H.T., X.X., X.L., Y.S. and W.L. analysed the data, with guidance from D.E.W., K.D.H., T.A.T., M.D.F., T.H.B. and X.W. X.H., C.A.K., C.L.-A., S.O.W., T.R.B. and X.W. wrote the manuscript. X.H., J.A.P., P.G., R.A.W., C.K., D.E.W., R.S., K.N., P.G.H. and X.W. interpreted the data and the research findings. All of the co-authors reviewed the manuscript. X.W. obtained the study funding.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Hong, X. *et al.* Genome-wide association study identifies peanut allergy-specific loci and evidence of epigenetic mediation in U.S. Children. *Nat. Commun.* 6:6304 doi: 10.1038/ncomms7304 (2015).