Genome-wide association study and meta-analysis in multiple populations identifies new loci for peanut allergy and establishes C11orf30/EMSY as a genetic risk factor for food allergy

Yuka Asai, MD, MSc,a,* Aida Eslami, PhD,b,* C. Dorien van Ginkel, BSc,c Loubna Akhabir, PhD,b Ming Wan, BSc,b George Ellis, BSc,b Moshe Ben-Shoshan, MD, MSc,d David Martino, PhD,e Manuel A. Ferreira, PhD,f Katrina Allen, MD, PhD,g Bruce Mazier, MD,h Hans de Groot, MD, PhD,i Nicolette W. de Jong, PhD,j Roy N. Gerth van Wijk, MD, PhD,k Anthony E. J. Dubois, MD, PhD,l Rick Chin, MSc,m Stephen Cheuk, MD,n Joshua Hoffman, PhD,o Eric Jorgensen, PhD,p John S. Witte, PhD,q Ronald B. Melles, MD,r Xiumei Hong, MD, PhD,s Xiaobin Wang, MD, MPH, ScD,o Jennie Hui, PhD,p Andrew J. Sandford, PhD,b Ann E. Clarke, MD, MSc,‡ and Denise Daley, PhD,‡+

Background: Peanut allergy (PA) is a complex disease with both environmental and genetic risk factors. Previously, PA loci were identified in flagggin (FLG) and HLA in candidate gene studies, and loci in HLA were identified in a genome-wide association study and meta-analysis.

Objective: We sought to investigate genetic susceptibility to PA.

Methods: Eight hundred fifty cases and 926 hyper-control subjects and more than 7.8 million genotyped and imputed single nucleotide polymorphisms (SNPs) were analyzed in a genome-wide association study to identify susceptibility variants for PA in the Canadian population. A meta-analysis of 2 phenotypes (PA and food allergy) was conducted by using 7

From 1the Division of Dermatology, Department of Medicine, Queen’s University, Kingston, and the Division of Experimental Medicine, Department of Medicine, McGill University, Montreal; 2the Centre for Heart Lung Innovation, University of British Columbia, Vancouver; 3the University of Groningen, University Medical Center Groningen, Department of Pediatric Pulmonology and Pediatric Allergology and University of Groningen, University Medical Center Groningen, GRIAC Research Institute, Groningen; 4the Division of Allergy and Immunology, Department of Pediatrics, Montreal Children’s Hospital, Research Institute of the McGill University Health Centre, Montreal; 5Murdoch Children’s Research Institute, Royal Children’s Hospital, University of Melbourne; 6QIMR Berghofer Medical Research Institute, Brisbane; 7the Department of Pediatric Allergology, Diakonessenhuis Voorburg, Reintier de Gisaf Gasbuis, Delft; 8Erasmus MC, Department of Allergology, Rotterdam; 9the Division of Rheumatology, Department of Medicine, Cumming School of Medicine, University of Calgary; 10Private Practice, Calgary, Alberta; 11the Department of Epidemiology and Biostatistics, University of California, San Francisco; 12the Division of Research, Kaiser Permanente Northern California, Redwood City Medical Center; 13the Department of Population, Family and Reproductive Health, Center on the Early Life Origins of Disease, Johns Hopkins University Bloomberg School of Public Health, Baltimore; 14the School of Population Health, University of Western Australia, Australia and Pathology and Laboratory Medicine, University of Western Australia; 15the Department of Respiratory Medicine, Sir Charles Gairdner Hospital, and School of Population Health, University of Western Australia, Perth; 16the School of Population Health, University of Western Australia, Perth; and 17the Department of Pulmonary Physiology and Sleep Medicine, Sir Charles Gairdner Hospital, and School of Medicine and Pharmacology, University of Western Australia, Perth.

These authors contributed equally to this work as joint first authors.

These authors contributed equally to this work as joint senior authors.

Supported by the Allergy, Genes, and Environment Network of Centres of Excellence (AllerGEN NCE); the Canadian Institutes of Health Research; Canadian Research Chairs Program (salary award to D. Daley); the Natural Sciences and Engineering Research Council; the Michael Smith Foundation for Health Research/AllerGEN NCE post-doctoral research fellowship awards (A.E.); the Division of Experimental Medicine Entrance Fellowship (V.A.); the Montreal Children’s Hospital Foundation; the McGill University Health Centre Foundation; the Canadian Dermatology Foundation; the Canadian Allergy, Asthma and Immunology Foundation; the Canadian Society of Allergy and Clinical Immunology; the National Institutes of Health (grant no. CA112355 awarded to J.S.W. and RC2 AG036607 awarded to C.A.S. and N.J.R.); HealthWay; the National Health and Medical Research Council (Australia); the Australian Research Council; and the US Department of Defense (grant no. W81XWH-10-1-0487 to K.A.). C.D.v.G., A.E.J.D., and G.H.K. received an unrestricted grant from the Nutricia Research Foundation to complete the Dutch GENEVA cohort. The Dutch IDEAL cohort of N.W.d.J., H.d.G., R.G.v.W., and A.E.J.D. was supported by the Dutch Technology Foundation STW, which is part of the Netherlands Organisation for Scientific Research (NOW) and partly funded by the Ministry of Economic Affairs (project number 11868) and the Food Allergy Foundation, Siemens Healthcare Diagnostics, HAL Allergy, Intermack the Netherlands B.V., ALK-Abelló B.V., and the Netherlands Anaphylaxis Network. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by the National Cancer Institute; National Human Genome Research Institute; National Heart, Lung, and Blood Institute; National Institute on Drug Abuse; National Institute of Mental Health; and National Institute of Neurological Disorders and Stroke.

Disclosure of potential conflict of interest: Y. Asai and A. E. Clarke have received a grant from the Allergy, Genes, and Environment Network of Centres of Excellence (AllerGEN NCE) and has received travel support from the Canadian Institutes of Health Research (CIHR). C. D. van Ginkel has received grants from JC de Cock Stichting and the Nutricia Research Foundation. K. Allen has received consulting fees or honoraria from Nestle and Thermo Fisher and has consultant arrangements with Before Brands. B. Mazer and D. Daley have received grants from ALLERGEN-NCE and the Canadian Institutes of Health Research (CIHR); R. N. Gerth van Wijk has received a grant from STW. S. Cheuk has consultant arrangements with Novartis, Sanofi, Graceway, Merck, and Pfizer and has received payment for lectures from GlaxoSmithKline, Sanofi, Graceway, Pfizer, Merck, Novartis, and PediaPharm. G. H. Koppelman has received grants from the Lung Foundation of the Netherlands, TEVA, UBOO Emius Foundation, TETRI Foundation, and GlaxoSmithKline. The rest of the authors declare that they have no relevant conflicts of interest.

Received for publication September 18, 2016; revised September 5, 2017; accepted for publication September 19, 2017.

Corresponding author: Denise Daley, PhD, Centre for Heart Lung Innovation, St Paul’s Hospital, Rm 166, 1081 Burrard St, Vancouver, British Columbia, V6Z 1Y6 Canada.
E-mail: denise.daley@hls.ubc.ca.
0091-6749 © 2017 The Authors. Published by Elsevier Inc. on behalf of the American Academy of Allergy, Asthma & Immunology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
https://doi.org/10.1016/j.jaci.2017.09.015
studies from the Canadian, American (n = 2), Australian, German, and Dutch (n = 2) populations. Results: An SNP near integrin α6 (ITGA6) reached genome-wide significance with PA (P = 1.80 × 10^{-8}), whereas SNPs associated with Src kinase–associated phosphoprotein 1 (SKAP1), matrix metallopeptidase 12 (MMP12)/MMP13, catenin α3 (CTNNA3), rho GTPase–activating protein 24 (ARHGAP24), angiopoietin 4 (ANGPT4), chromosome 11 open reading frame (C11orf30/EMSY), and exocyst complex component 4 (EXOC4) reached a threshold suggestive of association (P ≤ 1.49 × 10^{-6}). In the meta-analysis of PA, loci in or near ITGA6, ANGPT4, MMP12/MMP13, C11orf30, and EXOC4 were significant (P ≤ 1.49 × 10^{-8}). When a phenotype of any food allergy was used for meta-analysis, the C11orf30 locus reached genome-wide significance (P = 7.50 × 10^{-11}), whereas SNPs associated with ITGA6, ANGPT4, MMP12/MMP13, and EXOC4 and additional C11orf30 SNPs were suggestive (P ≤ 1.49 × 10^{-6}). Functional annotation indicated that SKAP1 regulates expression of CXB1, which co-localizes with the EMSY protein coded by C11orf30. Conclusion: This study identifies multiple novel loci as risk factors for PA and food allergy and establishes C11orf30 as a risk locus for both PA and food allergy. Multiple genes (C11orf30/EMSY, SKAP1, and CTNNA3) identified by this study are involved in epigenetic regulation of gene expression. (J Allergy Clin Immunol 2017; xxxx:xxxx-xxxx.)

Key words: Peanut allergy, food allergy, genome-wide association study, meta-analysis, EMSY, C11orf30, epigenetics

Peanut allergy (PA) is a main cause of anaphylaxis in North America. In Canada the prevalence of PA is 1% overall, with a prevalence of 2.2% in children. The self-reported prevalence of tree nut allergy and PA in the United States was 2.1% in children, whereas 3% of infants in an Australian study had a positive food challenge result to peanut. PA is highly heritable, with a concordance rate of 64% in monozygotic twins compared with 7% in dizygotic twins. Family studies have found the risk of PA in subjects with a sibling with PA to be significantly greater than in the general population, with odds ratios (ORs) ranging from 6.7 to 13.5.

The pathogenesis of PA involves both genetics and the environment. Involvement of environmental exposures is supported by (1) findings that early oral exposure to peanut leads to development of tolerance, whereas differences in PA prevalence and disease phenotype vary with age, (2) differences in PA prevalence internationally, and (3) the rapid increase in disease prevalence reported in some studies that cannot be explained by genetic changes.

Previous genetic work has found risk factors for PA in the innate and adaptive immune pathways, including HLA, CD14, IL9, and filaggrin (FLG). Recently, genome-wide association studies (GWASs) of food allergy identified associations between PA and the HLA region. We have previously identified HLA and FLG associations with PA in a well-characterized group of Canadian patients with PA from the Canadian Peanut Allergy Registry (CanPAR). As a follow-up to this work, we conducted a GWAS of PA, along with a meta-analysis of results from the previously published GWAS and other studies of food allergy. HLA variants were identified as significant risk factors for PA in the CanPAR GWAS (rs1049213, P = 1.82 × 10^{-11}) and in a meta-analysis (rs1063347, P = 3.67 × 10^{-23}), as reported in a separate publication, in which we narrowed the locus to HLA-DQB1 and showed that its relationship to PA is independent of asthma. Here we present novel non-HLA loci identified in a GWAS and meta-analysis in an additional 6 populations.

METHODS

Clinical characteristics

Inclusion criteria for CanPAR cases are found in Table E1 in this article’s Online Repository at www.jacionline.org. Salivary DNA was isolated from patients with PA in the CanPAR study. Hyper-control subjects were self-reported white subjects from the Busselton Health Study in Australia with no history of asthma, atopy, eczema, allergic rhinitis, or food allergy who had blood-derived DNA and data assessed by using methacholine challenge and skin prick tests. Genotyping of 1974 subjects (987 cases and 987 control subjects) was conducted on the Illumina Omni 2.5M+Exome 8v1.1 chip (Génome Québec Innovation Centre, Montreal, Quebec, Canada). Quality control (QC), including batch effects, single nucleotide polymorphisms (SNPs), and sample quality, are described in Fig E1 (this article’s Online Repository at www.jacionline.org). A total of more than 7.8 million SNPs (1,388,588 genotyped and 6,441,607 imputed) and 1,776 subjects (850 cases and 926 control subjects) passed QC (see Fig E2 in this article’s Online Repository at www.jacionline.org). Details on imputation are presented in the Methods section in this article’s Online Repository at www.jacionline.org.

Two analyses were performed (related and unrelated) because examination of alleles determined to be identical by using state and KING kinship coefficients identified related cases (siblings) and control subjects (first- to third-degree relatives). PC-Air and KING (http://people. virginia.edu/~wc/k/KING/) were used to estimate principal components and kinship coefficients for the related analysis. Association analyses were conducted with Stata software, with sandwich estimation to model the clustering of family genotypes with the addition of a family group identifier, 10 principal components to account for population stratification, and plate numbers to account for plate effects.
A secondary case-control study excluding related subjects was conducted with PLINK (version 1.07). To make the sample unrelated, 160 subjects were excluded (14 cases and 146 control subjects); the youngest subject in each family was retained. The unrelated analysis was performed with 834 cases and 781 control subjects (n = 1615; see Fig E2, B).

The analysis including related subjects is our primary analysis because it has the largest sample size and greatest power. Rank order and OR differences were evaluated between related and unrelated analyses. All subsequent analyses, including conditioning, were conducted by using the related analysis. A P value of 3.60 × 10−8 was considered the threshold for genome-wide significance (Bonferroni correction), with 1.49 × 10−6 being suggestive evidence for association. We chose 1.49 × 10−6 as our threshold based on significance levels presented in 2 previously published PA GWAS studies.

**Conditioning on HLA**

After identification of multiple SNPs in the HLA region, we conditioned on the top genotyped SNP (rs3134976) to investigate independence of signals from the rest of the genome and to determine the contribution of HLA associations to deviation from the expected line observed in the quantile-quantile plots (Fig 1).

**Meta-analysis**

A meta-analysis was conducted by using 2 phenotypes (PA and food allergy), including previously published PA GWAS results, and published data. The CanPAR study and 6 additional studies were included in the meta-analysis: 2 American studies (the Chicago Food Allergy [CFA] study [n = 2,197; 316 PA cases] and the Genetic Epidemiology Research on Allergies [GERA] cohort [n = 29,053; 5,108 self-reported food allergy]), the Australian HealthNuts study (n = 221; 73 patients with PA), and the German Understanding Food Allergy (UFA) study (n = 2,592; 205 patients with PA), which contributed 21 previously published SNPs. Genotyping for SNPs was conducted in 2 Dutch studies: IDEAL and GENEVA (n = 512; 138 patients with PA). Both IDEAL and GENEVA include cases with general food allergy. See Table E2 in this article’s Online Repository at www.jacionline.org for full study and phenotype descriptions. The meta-analysis for PA included 1,582 patients with PA and 5,446 control subjects, with more than half of the patients and control subjects coming from the CanPAR study. Because the GERA cohort used self-reported food allergy phenotypes with no additional diagnostic testing or history, the meta-analysis was conducted both with and without GERA data to evaluate the sensitivity of the meta-analysis results to stringent food allergy phenotyping. The meta-analysis for food allergy included 7,267 patients with food allergy and 29,084 control subjects, with inclusion of GERA data.

Fixed- and random-effects models evaluate heterogeneity but require point estimates and SEs. Because the CFA study provided P values and sample sizes only, for meta-analyses, P values were obtained by using the Stouffer weighted z score, which requires consistency in the direction of effects (for accurate P value estimation), and we were able to confirm that the direction of effect is the same for the CFA study because the investigators provided us with the case/control allele frequencies, which are consistent with the CanPAR associations.

**Identification of expression quantitative trait loci**

The Genotype-Tissue Expression (gtexportal.org) database was queried for novel regions.

**RESULTS**

**GWAS**

SNPs in HLA and an imputed SNP on chromosome 2 close to integrin α6 (ITGA6; rs115218289, P = 1.80 × 10−8; Fig 1 and Table 1 and see Table E3 in this article’s Online Repository at www.jacionline.org) reached genome-wide significance. Several SNPs with suggestive evidence for association were detected in novel loci (Table 1 and see Table E3), including multiple SNPs located in Src kinase–associated phosphoprotein 1 (SKAP1; chromosome 17), 1 located between matrix metalloproteinase 12 (MMP12) and MMP13 (rs144897250, chromosome 11; P = 2.90 × 10−7), multiple SNPs within catenin α3 (CTNN3; chromosome 10), rs744597 near rho GTPase–activating protein 24 (ARHGAP24; chromosome 4, P = 3.98 × 10−7), rs523865 in angiopoietin 4 (ANGPT4; chromosome 20, P = 4.42 × 10−7), multiple SNPs near the chromosome 11 open reading frame (C11orf30; chromosome 11, also known as EMSY), and rs78048444, which is located in a region between coiled-coil-helix-coiled-coil-helix domain containing 3 (CHCHD3) and exocyst complex component 4 (EXOC4; chromosome 7, P = 5.44 × 10−7).

After conditioning on the top genotyped HLA SNP (rs3134976, Fig 2), deviation observed in the quantile-quantile plot was largely resolved (Figs 1, A, and 2, A); residual deviation is primarily due to the number of SNPs supporting SKAP1, CTNNA3, and C11orf30/EMSY associations. Conditioning identified 16 additional SNPs near SKAP1 and CTNNA3 (rs139902172, see Table E4 in this article’s Online Repository at www.jacionline.org), likely because of the low MAF.

**Meta-analysis for PA**

We identified 85 SNPs in common between the CanPAR study and 1 or more of the previously reported PA GWASs. The top novel SNP identified in the meta-analysis for PA was rs115218289, which was located near ITGA6 and did not reach genome-wide significance but met the threshold suggestive for significance (P = 9.16 × 10−8; Table II and see Table E5 and full results in Table E6 in this article’s Online Repository at www.jacionline.org). Loci in ANGPT4 (rs523865, P = 1.54 × 10−7) and intragenic SNPs (rs144897250, P = 2.94 × 10−7) near MMP12/MMP13, C11orf30 (rs7936434, P = 3.13 × 10−7), and EXOC4 (rs78048444, P = 7.33 × 10−7) were suggestive of significance (P ≤ 1.49 × 10−6) in a meta-analysis for the PA phenotype.

**Meta-analysis of food allergy**

By using the phenotype of “any food allergy” in all 6 populations, both with and without GERA data, the top SNP identified in meta-analysis was rs7936434 near C11orf30 (P = 1.98 × 10−8 and P = 7.50 × 10−11 with and without GERA data, respectively; Table II and see Table E5). The SNPs associated with ITGA6, ANGPT4, and MMP12/MMP13 were suggestive of significance (P ≤ 1.49 × 10−6) in a meta-analysis for food allergy but only if GERA data were not included.
FIG 1. Quantile-quantile and Manhattan plots of related and unrelated analyses. A, Quantile-quantile plot of the expected distribution of test statistics (x-axis) versus observed P values (y-axis) for related (left) and unrelated (right) analyses. B, Manhattan plot: SNPs in 850 patients with PA and 926 hyper-control subjects for the related (upper) and unrelated (lower) analyses. The x-axis denotes the genomic location, and the y-axis denotes the association level. The solid line indicates the threshold for genome-wide significance \((P \leq 3.60 \times 10^{-8})\), and the dashed line indicates the suggestive association significance threshold \((P \leq 1.49 \times 10^{-6})\).
The nearest gene was used to determine genomic location.

Identification of expression quantitative trait loci

Many SNPs identified near SKAP1 by the CanPAR study were expression quantitative trait loci (eQTLs) regulating expression of 2 genes, sorting nexin 1 (SNX1) and chromobox protein homolog 1 (CBX1), in numerous tissues (sun-exposed skin, whole blood, transformed fibroblasts, testis, colon, and thyroid). Results are presented for tissues relevant to PA and food allergy (sun-exposed skin, whole blood, and transformed fibroblasts) with a P value of less than 1.0 × 10^{-6} (Table III and see Table E7 in this article’s Online Repository at www.jacionline.org). Little is known about SNX1; it belongs to a family of retrograde transport molecules, and its protein is involved in targeting cell-surface molecules to the lysosome. CBX1 is a member of the highly conserved heterochromatin protein family that binds to histones through methylated lysine residues, mediating gene silencing and alternative splicing. It is believed that CBX1 can play an important role in epigenetic regulation and gene expression.

**DISCUSSION**

This study identifies several novel loci for PA and food allergy. The effect sizes for the identified loci are large, with ORs ranging from 0.18 to 0.22 and 1.57 to 6.20; these effect sizes are particularly impressive for a complex disease and a small GWAS. The most significant novel PA locus from the CanPAR GWAS was rs115218289, which is located in ITGA6. This was an imputed SNP with low MAF (0.021), with no directly genotyped SNPs supporting the association (Fig 1, A). A second imputed SNP near MMP12/MMP13 also had a low MAF (0.02). Although significant or suggestive of significance for other loci identified in the CanPAR GWAS data compared with the meta-analysis.

**Table I.** Most significant SNPs from 10 genomic regions identified in the CanPAR GWAS listed by order of significance

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chromosome</th>
<th>Position</th>
<th>Allele</th>
<th>Source of SNPs</th>
<th>MAF</th>
<th>OR (Loci)</th>
<th>UCI</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs115218289</td>
<td>2</td>
<td>173265750</td>
<td>A/C</td>
<td>Imputed</td>
<td>0.02</td>
<td>2.16</td>
<td>1.90</td>
<td>5.17</td>
</tr>
<tr>
<td>rs72827854</td>
<td>17</td>
<td>46460525</td>
<td>T/C</td>
<td>Imputed</td>
<td>0.09</td>
<td>2.60</td>
<td>2.60×10^{-7}</td>
<td>2.16</td>
</tr>
<tr>
<td>rs144899250</td>
<td>10</td>
<td>102750264</td>
<td>A/C</td>
<td>Imputed</td>
<td>0.02</td>
<td>2.90</td>
<td>2.90×10^{-7}</td>
<td>2.90</td>
</tr>
<tr>
<td>rs7457217</td>
<td>10</td>
<td>6844401</td>
<td>T/C</td>
<td>Genotyped</td>
<td>0.38</td>
<td>1.98</td>
<td>1.98×10^{-7}</td>
<td>1.98</td>
</tr>
<tr>
<td>rs745497</td>
<td>8</td>
<td>86337028</td>
<td>A/G</td>
<td>Genotyped</td>
<td>0.40</td>
<td>0.74</td>
<td>0.74×10^{-7}</td>
<td>0.74</td>
</tr>
<tr>
<td>rs523865</td>
<td>20</td>
<td>894881</td>
<td>C/T</td>
<td>Genotyped</td>
<td>0.23</td>
<td>0.44</td>
<td>0.44×10^{-7}</td>
<td>0.44</td>
</tr>
<tr>
<td>rs7936434</td>
<td>10</td>
<td>76293805</td>
<td>C/G</td>
<td>Imputed</td>
<td>0.49</td>
<td>1.90</td>
<td>1.90×10^{-7}</td>
<td>1.90</td>
</tr>
<tr>
<td>rs78048444</td>
<td>8</td>
<td>132832218</td>
<td>C/T</td>
<td>Genotyped</td>
<td>0.22</td>
<td>0.39</td>
<td>0.39×10^{-6}</td>
<td>0.39</td>
</tr>
<tr>
<td>rs56151068</td>
<td>17</td>
<td>46381431</td>
<td>T/C</td>
<td>Genotyped</td>
<td>0.10</td>
<td>2.76</td>
<td>2.76×10^{-7}</td>
<td>2.76</td>
</tr>
<tr>
<td>rs139462954</td>
<td>17</td>
<td>46523678</td>
<td>A/C</td>
<td>Imputed</td>
<td>0.09</td>
<td>2.76</td>
<td>2.76×10^{-6}</td>
<td>2.76</td>
</tr>
</tbody>
</table>

Allele: Minor allele/major allele; LCI, lower 95% CI; UCI, upper 95% CI.

*The nearest gene was used to determine genomic location.

†Comparing related and unrelated analyses ORs.
**TABLE II. Meta-analysis of Canadian, American, Australian, German, and Dutch populations for association with PA and food allergy phenotypes**

| SNP* | Chromosome | Allele | P-CanPAR|$^a$ | P-CFA|$^b$ | P-HealthNuts|$^c$ | P-UFAl|$^d$ | P-IDEAL/GENEVA, case-control|$^e$ | P-Dutch GENEVA, family based|$^f$ |
|------|------------|--------|---------|--------|---------|-------------|--------|---------|----------------|----------------|----------------|
| rs115218289|$^g$ | 2 | A/C | $1.80 \times 10^{-8}$ | NA | 6.77 $\times 10^{-1}$ | NA | 7.18 $\times 10^{-1}$ | 3.54 $\times 10^{-1}$ |
| rs523865|$^g$ | 20 | C/T | $4.42 \times 10^{-7}$ | NA | NA | NA | 8.33 $\times 10^{-1}$ | 1.63 $\times 10^{-2}$ |
| rs144897250|$^g$ | 11 | A/C | $2.90 \times 10^{-7}$ | NA | 3.84 $\times 10^{-1}$ | NA | NA | NA |
| rs7936434|$^g$ | 11 | C/G | $5.17 \times 10^{-7}$ | $3.66 \times 10^{-2}$ | $1.43 \times 10^{-1}$ | NA | NA | NA |
| rs78048444|$^g$ | 7 | C/T | $5.44 \times 10^{-7}$ | NA | 2.13 $\times 10^{-1}$ | NA | 6.46 $\times 10^{-1}$ | 3.54 $\times 10^{-1}$ |
| rs744597|$^g$ | 4 | A/G | $3.98 \times 10^{-7}$ | $1.15 \times 10^{-1}$ | $5.57 \times 10^{-1}$ | NA | 8.20 $\times 10^{-2}$ | 9.62 $\times 10^{-1}$ |
| rs72827854|$^g$ | 17 | T/C | $2.60 \times 10^{-7}$ | $1.36 \times 10^{-1}$ | $3.55 \times 10^{-1}$ | NA | NA | NA |
| rs55765969|$^g$ | 17 | T/C | $1.23 \times 10^{-6}$ | $1.97 \times 10^{-1}$ | $2.81 \times 10^{-1}$ | NA | NA | NA |
| rs56151068|$^g$ | 17 | T/C | $9.58 \times 10^{-7}$ | $2.51 \times 10^{-1}$ | $3.57 \times 10^{-1}$ | NA | NA | NA |
| rs71193762|$^g$ | 10 | A/G | $3.77 \times 10^{-7}$ | $8.63 \times 10^{-1}$ | $3.62 \times 10^{-1}$ | NA | NA | NA |

Boldface rows indicate suggestive significance ($P \leq 1.49 \times 10^{-8}$) in patients with PA.

|$^a$P value from CanPAR (n = 1,776).

|$^b$P value from the Dutch GENEVA family study; number of informative families = 20, 112, 21, and 112 for rs115218289, rs523865, rs78048444, and rs744597, respectively.

|$^c$P value from the Stouffer weighted z score meta-analysis method for PA.

|$^d$P value from the Dutch GENEVA, Asthma, and rhinoconjunctivitis study; number of subjects = 226, 229, 227, and 217 for rs115218289, rs523865, rs78048444, and rs744597, respectively.

|$^e$P value from the Stouffer weighted z score meta-analysis method for food allergy.

|$^f$P value from the Australian HealthNuts study (n = 2,197).

|$^g$P value from the Australian HealthNuts study (n = 2,197).

|$^h$P value from the Dutch IDEAL and GENEVA case-control studies; number of subjects for SNPs = 479, 487, 482, and 466 for rs115218289, rs523865, rs78048444, and rs744597, respectively.

|$^i$P value from the Dutch GENEVA family study; number of informative families = 26, 196, 37, and 214 for rs115218289, rs523865, rs78048444, and rs744597, respectively.

|$^j$P value from the Australian HealthNuts study (n = 2,197).

|$^k$P value from the Stouffer weighted z score meta-analysis method for food allergy.

|$^l$P value from the Stouffer weighted z score meta-analysis method for food allergy without the GERA cohort.

|$^m$P value from the GERA food allergy study (n = 29,053).

|$^n$Imputed SNP from CanPAR.

|$^o$Used for both PA and food allergy (FA).

|$^p$SNPs in this table were selected to represent each of the 10 genomic regions identified in the CanPAR GWAS (see Table E4 for full results).

|$^q$Genotyped SNP from CanPAR.

Significance of the other loci in the meta-analysis should be interpreted with care because there is insufficient power in the other studies to replicate CanPAR PA findings.

Identification of C11orf30 as a susceptibility locus for food allergy fits within what is known about this area, as has been implicated in serum IgE levels and asthma. Recently, C11orf30 was identified as a risk factor for eosinophilic esophagitis (OR, 2.22; $P = 5.38 \times 10^{-10}$). C11orf30 is a chronic inflammatory disease of the esophagus that is mainly triggered by food proteins. Loci in this region have been significantly associated with atopic dermatitis, Asthma, and rhinoconjunctivitis, in general.

It could be argued that the genome-wide significance of C11orf30 in the food allergy meta-analysis with the addition of the GERA cohort indicates that the locus represents association with an allergic diathesis rather than with food allergy or PA specifically because of the potential misclassification rate for self-reported food allergy in the GERA cohort. The idea of C11orf30 being associated with an allergic phenotype is supported by data that show that C11orf30 is a risk factor for polysensitization to multiple allergens on skin prick testing. C11orf30 is associated with Crohn disease and ulcerative colitis, autoimmune inflammatory bowel diseases that are not classically part of the atopic march, although some epidemiologic data link eczema with inflammatory bowel diseases and other autoimmune diseases. Therefore C11orf30 could be a risk factor for immune dysfunction disorders in general.

**PA, food allergy, or atopy?**

The underlying genetic model for PA is unknown: food allergy, asthma, eczema, and allergic rhinitis can share common genetic susceptibility, whereas environmental factors determine which specific atopic disease develops; alternatively, each specific allergy can have its own risk variants. The genetic model and study design affect the power to identify risk variants. The
CanPAR GWAS subjects were recruited to investigate risk factors specific for PA, with the use of hyper-control subjects to increase power, in addition to a defined phenotype for allergy to peanut. The other food allergy case groups were recruited based on any food allergy or egg allergy and PA. This design might be more powerful if one assumes that all food allergies are influenced by the same genetic risk loci rather than specific risk loci for specific foods.

**Novel pathways in the pathogenesis of food allergy**

The identification of C11orf30 as a genetic risk locus for food allergy opens further research possibilities into the pathways to food allergy. Mechanistically, it is possible that the C11orf30 region could be responsible for multiple phenotypes associated with the atopic march, including food allergy, eczema, and asthma, and other autoimmune conditions. There are several potential ways C11orf30 could exert its influence to result in a variety of clinical presentations and are consistent with an epigenetic connection to allergic disease. The protein complex formed by KDM5A, a histone demethylase, and EMSY, the protein encoded by C11orf30, appears to increase gene transcription. EMSY acts as a reader for trimethylation of lysine 4 on histone protein H3 (H3K4me3) through its recruitment with other members of the EMSY complex to H3K4me3 marked promoters, where it appears to be positively correlated with transcriptional activity of target genes and cell proliferation. Along with members of the EMSY protein complex, EMSY colocalizes with CBX1. Similarly to its highly conserved Drosophila homolog protein dHP1, CBX1 protein has a binding affinity for trimethylation of lysine 9 on histone protein H3 (H3K9me3), which has multiple SNPs identified in the CanPAR GWAS, has a known eQTL in CBX1, which encodes a protein that mediates gene silencing and alternative splicing, the product of which colocalizes with EMSY protein, as encoded by C11orf30.
histone marker H3K4me1 (monomethylation of lysine 4 on histone protein 3) in CTNNA3 CNVR. 

Identification of multiple genes (C11orf30, SKAP1, and CTNNA3) involved with histone-related proteins supports the hypothesis that epigenetic regulation is mechanistic in the development of allergy. Early exposure to peanut prevents development of PA,7 and maintenance of peanut consumption promotes continued tolerance of peanut8; the identified theme of histone-related loci reveals a potential biological mechanism through which this epigenetic phenomenon can occur. This finding could pave the way for potential therapies for those already affected by PA.

Several SNPs in or near genes related to vascular and endothelial cell factors were identified in this study and could be involved in the pathophysiology of PA and food allergy. For example, SNPs in or near genes related to vascular and endothelial cell factors were identified in this study and could be involved in the pathophysiology of PA and food allergy through 2 putative mechanisms: (1) an endothelial barrier defect

**FIG 2.** Quantile-quantile and Manhattan plots of related and unrelated analyses conditioned on the top HLA SNP (rs3134976). A, Quantile-quantile plot of the expected distribution of test statistics (x-axis) versus observed P values (y-axis) for related analysis conditioned on rs3134976. SNPs in complete linkage with rs3134976 were excluded. B, Manhattan plot: SNPs in 850 patients with PA and 926 hyper-control subjects for the related analysis conditioned on rs3134976. The x-axis denotes genomic location, and the y-axis denotes association level. The solid line indicates the threshold for genome-wide significance ($P < 3.60 \times 10^{-8}$), and the dashed line indicates the suggestive association significance threshold ($P < 1.49 \times 10^{-6}$).
promoting sensitization and (2) endothelial cells acting as antigen-presenting cells. Some evidence suggests that lymphatic endothelial cells present self-antigens and help regulate peripheral T-cell tolerance. Several identified loci from this study have connections to both allergy and vascular regulation (CTNNA3). Permeability (ANGPT4) or endothelial cell function (SKAPI and EXOC4). Their true role in the pathogenesis of allergic disease requires further research.

**Strengths and limitations of the study**

Many of the strengths of this study are directly tied to its limitations. The CanPAR study is large, with 850 cases after QC, and is the largest of the included case groups by a factor of 3, although it is still small for a GWAS. There are no similar PA case groups of equivalent size; this complicates the replication of novel hits. The 6 studies included in the meta-analysis only contribute an additional 732 patients with PA, which results in insufficient power to replicate CanPAR findings.

There are numerous sources of heterogeneity that increase the variance of effect size estimation and reduce power. These include study design, ascertainment criteria, case and control phenotype definition, ethnicity, age, and small sample size. Differences in ascertainment criteria (food allergy vs PA) can result in differences in the power of the studies to identify susceptibility genes for PA versus food allergy, even in the presence of equivalent sample sizes. The effect of ascertainment criteria on the power ultimately depends on the underlying genetic model and will differentially affect loci with peanut-specific susceptibility. The Dutch IMPACT and GENEVA samples are small and contribute point estimates for only 4 SNPs that were specifically genotyped for this study; the absence of genome-wide association data precludes the use of principal components to control for population stratification when generating the point estimates for the IMPACT and GENEVA studies.

The importance of phenotyping food allergy is evident by the sizable changes in $P$ values when a large cohort of self-reported GERA cohort subjects with food allergy was added to the meta-analysis, which highlights the potential difference in mechanism for C11orf30. The prevalence of food allergy in the GERA cohort, which is self-reported with no corroborating clinical history or diagnostic tests, is 17.58%, which is much higher than other population-based studies in North America (6.4% to 7.5%), indicating there is likely a high case misclassification rate. Misclassification is a known differential bias to the null and can result in false-negative results, which can be problematic for relatively rare phenotypes, such as PA and food allergy, because the misclassification rate can exceed the disease prevalence rate. This is particularly an issue in the use of universal controls, where genotyping of control subjects has been performed in a separate experiment and misclassification of cases and control subjects is common. Misclassification of an affected subject as a control subject is much less costly than misclassification of a control subject as an affected subject.

A strength of the CanPAR study is that the cases and control subjects were genotyped in the same experiment. Additionally, using hyper-control subjects with a comprehensive longitudinal lifetime history and in-depth phenotyping of asthma, atopy, skin prick tests, IgE, airway hyperresponsiveness, and eczema phenotypes virtually eliminates misclassification of control subjects in the CanPAR study.

Careful consideration of potential misclassification needs to be made when using self-reported food allergy phenotypes in the absence of food challenges, skin prick tests, and food-specific IgE measurements. This has important implications in today’s research environment, where investigators are continually striving to maximize the use of existing GWAS data. We also see this conundrum with the common adoption of general population controls, which might contribute to population stratification and can result in false-positive results, a differential bias away from the null; both the CanPAR and UFA studies used this approach. The use of hyper-control subjects in the CanPAR study increased the power to detect associations with PA but made evaluation of confounders, such as eczema or asthma, difficult because none of the control subjects expressed these phenotypes. Use of this study design with control subjects from a different population limits our ability to evaluate early-life environmental exposures.

Direct genotyping was not conducted for rs7936434 (imputed), the most significant association observed in the C11orf30/EMSY region. Although direct genotyping would ensure that the finding is not an imputation artifact, it is unlikely that the observed association is due to an artifact. The C11orf30/EMSY locus reached genome-wide significance because of the contributions not only of the CanPAR study but also 2 other studies (CFA study and GERA cohort). Additionally, the strong correlation of this locus with allergy and atopic conditions in multiple studies serves to alleviate concerns that this association might be an imputation artifact.

Despite all limitations, this study robustly identified new loci for PA and food allergy with genome-wide significance ($P = 7.50 \times 10^{-11}$) across populations in a meta-analysis of unprecedented size (1,582 patients with PA and 5,446 control subjects and 7,627 patients with food allergy and 29,084 food allergy control subjects). It is important to note the effect sizes for the loci identified are large, particularly for a complex disease, and that the finding of C11orf30/EMSY is robust because it demonstrates association across the studies, despite study diversity. eQTL data support these new loci and suggest new pathways in the pathogenesis of food allergy.

**Conclusion and future directions**

The results of this study identify novel genetic risk factors for PA and food allergy. New pathways identified by using this

---

**TABLE III. Top CanPAR SNPs associated with PA and eQTLs for each locus and tissue type**

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chromosome</th>
<th>Position</th>
<th>Nearest gene</th>
<th>Tissue</th>
<th>$P$ value</th>
<th>Gene symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs4491576</td>
<td>17</td>
<td>46,408,636</td>
<td>SKAPI</td>
<td>Whole blood</td>
<td>$1.05 \times 10^{-10}$</td>
<td>SNX11</td>
</tr>
<tr>
<td>rs16956501</td>
<td>17</td>
<td>46,497,274</td>
<td>SKAPI</td>
<td>Skin, sun-exposed lower leg</td>
<td>$4.48 \times 10^{-10}$</td>
<td>SNX11</td>
</tr>
<tr>
<td>rs139462954</td>
<td>17</td>
<td>46,523,678</td>
<td>LOC101927166</td>
<td>Skin, sun-exposed lower leg</td>
<td>$8.94 \times 10^{-10}$</td>
<td>SNX11</td>
</tr>
<tr>
<td>rs139462954</td>
<td>17</td>
<td>46,523,678</td>
<td>LOC101927166</td>
<td>Cells, transformed fibroblasts</td>
<td>$7.36 \times 10^{-7}$</td>
<td>CBX1</td>
</tr>
</tbody>
</table>
unbiased approach include C11orf30/EMSY in patients with PA and those with food allergy and the importance of epigenetic mechanisms. It is evident that further work will require larger sample sizes and international collaboration by using well-phenotyped subjects for both food allergies and other atopic conditions. Functional work, including studies of vascular and endothelial cell factors, might be valuable. Future studies will need to examine gene-environment interactions, including duration, timing, and mode of environmental exposure and the role of CNVs, methylation, and histone modification.

We thank the subjects and parents who participated in this study and in the CanPAR registry and their allergists for help in diagnostic documentation. We appreciate the help with subject recruitment from our partners: Food Allergy Canada (formerly Anaphylaxis Canada), Allergies Québec (formerly l’Association Québécoise des Allergies Alimentaires), and the Allergy/ Asthma Information Association. We acknowledge Basia Rogula and David Yin, who assisted in quality control; JinCheol Choi, who assisted with figure development; and Jessica Que, who assisted with SNP annotation. Whitney Steber, Chynnace Van Lambridge, Heather Waidhauser, Greg Shand, Elizabeth Turnbull, and Popi Panaritis assisted in data/DNA collection for CanPAR. We thank Peter Paré for his valuable comments and Reza Alizadehfar, Edmund Chan, and Peter Hull for their data collection and helpful discussions. The data used for the analyses described in this article were obtained from the GTEx Portal on January 1, 2017, to February 28, 2017.

Key messages

- The C11orf30 (EMSY) locus is a novel genetic risk factor for both PA and food allergy, reaching genome-wide significance for food allergy in a meta-analysis ($P = 7.50 \times 10^{-11}$).
- New loci associated with ANGPT4, MMP12/MMP13, and EXOC4 are suggestive of association with PA and food allergy.
- Epigenetic mechanisms might be a new pathway in the pathogenesis of PA.

REFERENCES

27. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559-75.


