

# 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,<sup>a,\*</sup> Aida Eslami, PhD,<sup>b,\*</sup> C. Dorien van Ginkel, BSc,<sup>c</sup> Loubna Akhabir, PhD,<sup>b</sup> Ming Wan, BSc,<sup>b</sup> George Ellis, BSc,<sup>b</sup> Moshe Ben-Shoshan, MD, MSc,<sup>d</sup> David Martino, PhD,<sup>e</sup> Manuel A. Ferreira, PhD,<sup>f</sup> Katrina Allen, MD, PhD,<sup>e</sup> Bruce Mazer, MD,<sup>d</sup> Hans de Groot, MD, PhD,<sup>g</sup> Nicolette W. de Jong, PhD,<sup>h</sup> Roy N. Gerth van Wijk, MD, PhD,<sup>h</sup> Anthony E. J. Dubois, MD, PhD,<sup>c</sup> Rick Chin, MSc,<sup>i</sup> Stephen Cheuk, MD,<sup>j</sup> Joshua Hoffman, PhD,<sup>k</sup> Eric Jorgensen, PhD,<sup>l</sup> John S. Witte, PhD,<sup>m</sup> Ronald B. Melles, MD,<sup>n</sup> Xiumei Hong, MD, PhD,<sup>o</sup> Xiaobin Wang, MD, MPH, ScD,<sup>o</sup> Jennie Hui, PhD,<sup>p</sup> Arthur W. (Bill) Musk, FRACP,<sup>q</sup> Michael Hunter, PhD,<sup>r</sup> Alan L. James, FRACP,<sup>s</sup> Gerard H. Koppelman, MD, PhD,<sup>c</sup> Andrew J. Sandford, PhD,<sup>b</sup> Ann E. Clarke, MD, MSc,<sup>l,†</sup> and Denise Daley, PhD<sup>b,‡</sup>  
*Kingston, Ontario, Montreal, Quebec, Vancouver, British Columbia, and Calgary, Alberta, Canada; Groningen, Delft, and Rotterdam, The Netherlands; Melbourne, Brisbane, and Perth, Australia; San Francisco, Oakland, and Redwood City, Calif; and Baltimore, Md*

**Background:** Peanut allergy (PA) is a complex disease with both environmental and genetic risk factors. Previously, PA loci were identified in *filaggrin* (*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 <sup>a</sup>the Division of Dermatology, Department of Medicine, Queen's University, Kingston, and the Division of Experimental Medicine, Department of Medicine, McGill University, Montreal; <sup>b</sup>the Centre for Heart Lung Innovation, University of British Columbia, Vancouver; <sup>c</sup>the University of Groningen, University Medical Center Groningen, Department of Pediatric Pulmonology and Pediatric Allergy and University of Groningen, University Medical Center Groningen, GRIAC Research Institute, Groningen; <sup>d</sup>the Division of Allergy and Immunology, Department of Pediatrics, Montreal Children's Hospital, Research Institute of the McGill University Health Centre, Montreal; <sup>e</sup>Murdoch Children's Research Institute, Royal Children's Hospital, University of Melbourne; <sup>f</sup>QIMR Berghofer Medical Research Institute, Brisbane; <sup>g</sup>the Department of Pediatric Allergy, Diaconessenhuis Voorburg, Reinier de Graaf Gasthuis, Delft; <sup>h</sup>Erasmus MC, Department of Allergy, Rotterdam; <sup>i</sup>the Division of Rheumatology, Department of Medicine, Cumming School of Medicine, University of Calgary; <sup>j</sup>Private Practice, Calgary, Alberta; <sup>k</sup>the Department of Epidemiology and Biostatistics, University of California, San Francisco; <sup>l</sup>the Division of Research, Kaiser Permanente Northern California, Oakland; <sup>m</sup>the Department of Epidemiology and Biostatistics, University of California, San Francisco; <sup>n</sup>the Department of Ophthalmology, Kaiser Permanente Northern California, Redwood City Medical Center; <sup>o</sup>the Department of Population, Family and Reproductive Health, Center on the Early Life Origins of Disease, Johns Hopkins University Bloomberg School of Public Health, Baltimore; <sup>p</sup>the School of Population Health, University of Western Australia, Australia and Pathology and Laboratory Medicine, University of Western Australia; <sup>q</sup>the Department of Respiratory Medicine, Sir Charles Gairdner Hospital, and School of Population Health, University of Western Australia, Perth; <sup>r</sup>the School of Population Health, University of Western Australia, Perth; and <sup>s</sup>the 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 (Y.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, Intersnack 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 Padiapharm. G. H. Koppelman has received grants from the Lung Foundation of the Netherlands, TEVA, UBBO 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@hli.ubc.ca](mailto:denise.daley@hli.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>

Two analyses were performed (related and unrelated) because examination of alleles determined to be identical by using state and KING<sup>24</sup> kinship coefficients identified related cases (siblings) and control subjects (first- to third-degree relatives). PC-AiR<sup>25</sup> and KING<sup>24</sup> (KING1.4; <http://people.virginia.edu/~wc9c/KING/>) were used to estimate principal components and kinship coefficients for the related analysis. Association analyses were conducted with Stata software,<sup>26</sup> 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).<sup>27</sup> 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  $3.60 \times 10^{-8}$  was considered the threshold for genome-wide significance (Bonferroni correction), with  $1.49 \times 10^{-6}$  being suggestive evidence for association. We chose  $1.49 \times 10^{-6}$  as our threshold based on significance levels presented in 2 previously published PA GWAS studies.<sup>20,21</sup>

## Conditioning on *HLA*

After identification of multiple SNPs in the *HLA* region,<sup>22</sup> 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<sup>20,21</sup> and unpublished 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]<sup>20</sup> and the Genetic Epidemiology Research on Aging [GERA] cohort [ $n = 29,053$ ; 5,108 self-reported food allergy]),<sup>28</sup> the Australian HealthNuts study ( $n = 221$ ; 73 patients with PA),<sup>21</sup> and the German Understanding Food Allergy (UFA) study ( $n = 2,592$ ; 205 patients with PA),<sup>21</sup> 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.<sup>29</sup> See Table E2 in this article's Online Repository at [www.jacionline.org](http://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 completed 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](http://gtexportal.org))<sup>30</sup> database was queried for novel regions.

## RESULTS GWAS

SNPs in *HLA*<sup>22</sup> and an imputed SNP on chromosome 2 close to integrin  $\alpha 6$  (*ITGA6*; rs115218289,  $P = 1.80 \times 10^{-8}$ ; Fig 1 and Table I and see Table E3 in this article's Online Repository at

[www.jacionline.org](http://www.jacionline.org)) reached genome-wide significance. Several SNPs with suggestive evidence for association were detected in novel loci (Table I 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;  $P = 2.90 \times 10^{-7}$ ), multiple SNPs within catenin  $\alpha 3$  (*CTNNA3*; chromosome 10), rs744597 near rho GTPase-activating protein 24 (*ARHGAP24*; chromosome 4,  $P = 3.98 \times 10^{-7}$ ), rs523865 in angiopoietin 4 (*ANGPT4*; chromosome 20,  $P = 4.42 \times 10^{-7}$ ), 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,  $P = 5.44 \times 10^{-7}$ ).

No significant difference in ORs for SNPs was noted between the unrelated and related analyses (Table I). For 2 imputed SNPs (rs115218289 and rs144897250) with low (approximately 2%) minor allele frequency (MAF), there were differences in the rank order between the related and unrelated analyses (see Table E3 in this article's Online Repository at [www.jacionline.org](http://www.jacionline.org)), likely because of the low MAF.

## Conditioning

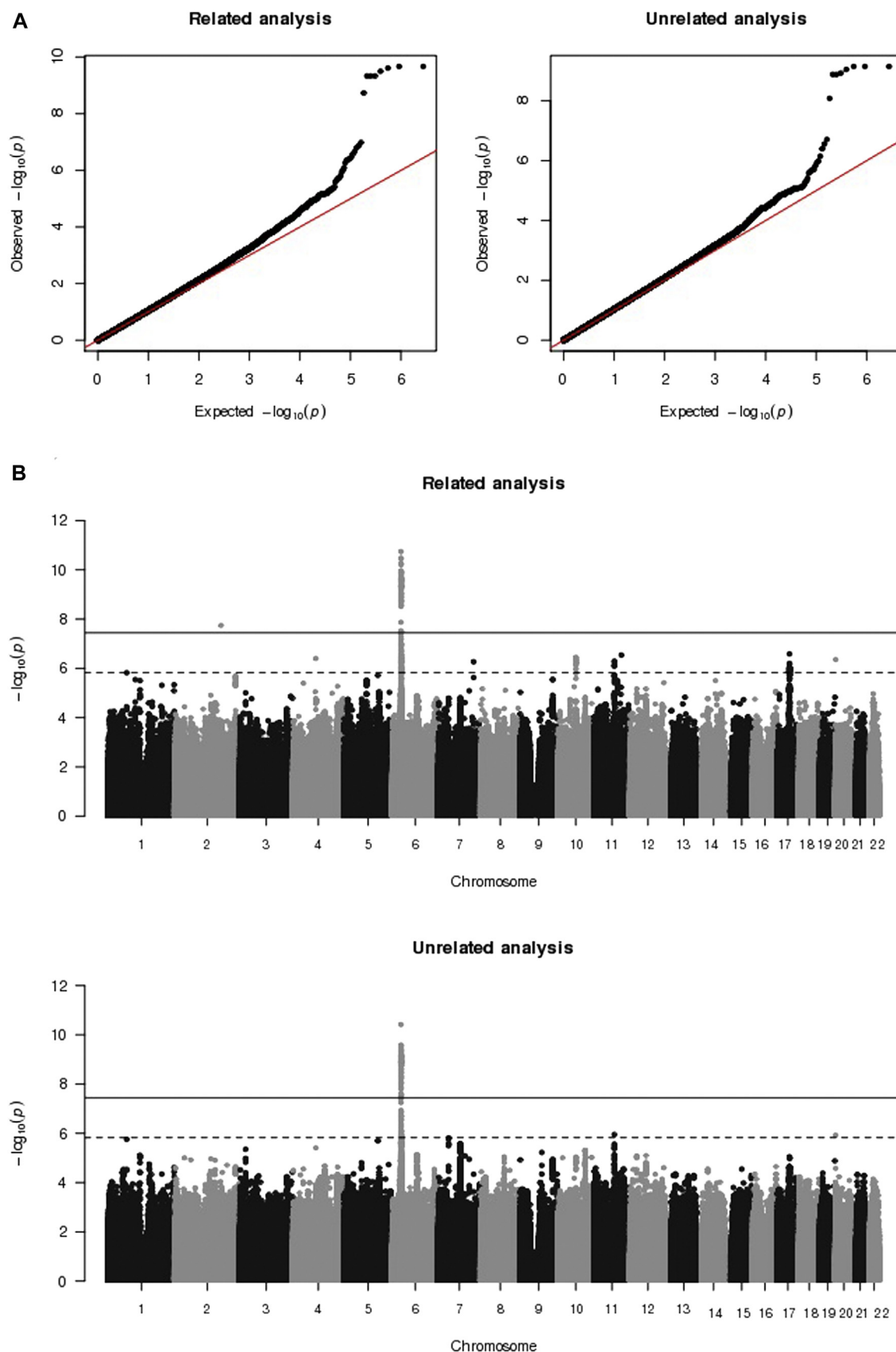
After conditioning on the top genotyped *HLA* SNP (rs3134976, Fig 2), deviation observed in the quantile-quantile plot was largely resolved (Figs 1, A, and 2, A); residual deviation is primarily due to the number of SNPs supporting *SKAP1*, *CTNNA3*, and *C11orf30/EMSY* associations. Conditioning identified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

## Meta-analysis for PA

We identified 85 SNPs in common between the CanPAR study and 1 or more of the previously reported PA GWASs.<sup>20,21</sup> 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 ( $P = 9.16 \times 10^{-8}$ ; Table II and see Table E5 and full results in Table E6 in this article's Online Repository at [www.jacionline.org](http://www.jacionline.org)). Loci in *ANGPT4* (rs523865,  $P = 1.54 \times 10^{-7}$ ) and intragenic SNPs (rs144897250,  $P = 2.94 \times 10^{-7}$ ) near *MMP12/MMP13*, *C11orf30* (rs7936434,  $P = 3.13 \times 10^{-7}$ ), and *EXOC4* (rs78048444,  $P = 3.73 \times 10^{-7}$ ) were suggestive of significance ( $P \leq 1.49 \times 10^{-6}$ ) 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* ( $P = 1.98 \times 10^{-8}$  and  $P = 7.50 \times 10^{-11}$  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 ( $P \leq 1.49 \times 10^{-6}$ ) 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 ( $P \leq 1.49 \times 10^{-6}$ ).



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

SNP	Chromosome	Position	Allele	MAF	Source of SNPs	Related analysis (850 cases and 926 control subjects)				Unrelated analysis (834 cases and 781 control subjects)				P value†	Gene/nearest gene
						OR	LCI	UCI	P value	OR	LCI	UCI	P value		
rs115218289	2	173265750	A/C	0.02	Imputed	0.18	0.10	0.32	$1.80 \times 10^{-8}$	0.20	0.09	0.46	$1.39 \times 10^{-4}$	$8.04 \times 10^{-1}$	(298 kb) <i>DLX2</i>   (26 kb) <i>ITGA6</i>
rs72827854	17	46460525	T/C	0.09	Imputed	2.16	1.61	2.90	$2.60 \times 10^{-7}$	2.08	1.50	2.87	$9.00 \times 10^{-6}$	$8.58 \times 10^{-1}$	<i>SKAP1</i>
rs144897250	11	102750264	A/C	0.02	Imputed	6.20	3.09	12.45	$2.90 \times 10^{-7}$	6.72	2.72	16.64	$3.79 \times 10^{-5}$	$8.90 \times 10^{-1}$	(5 kb) <i>MMP12</i>   (63 kb) <i>MMP13</i>
rs7475217	10	68444013	T/C	0.38	Genotyped	1.64	1.35	1.98	$3.58 \times 10^{-7}$	1.56	1.28	1.90	$9.19 \times 10^{-6}$	$7.37 \times 10^{-1}$	<i>CTNNA3</i>
rs744597	4	86337028	A/G	0.40	Genotyped	0.61	0.50	0.74	$3.98 \times 10^{-7}$	0.63	0.52	0.77	$3.91 \times 10^{-6}$	$8.01 \times 10^{-1}$	<i>ARHGAP24</i>
rs523865	20	894881	C/T	0.23	Genotyped	0.57	0.46	0.71	$4.42 \times 10^{-7}$	0.57	0.45	0.71	$1.19 \times 10^{-6}$	$9.49 \times 10^{-1}$	<i>ANGPT4</i>
rs7936434	11	76293805	C/G	0.49	Imputed	1.58	1.32	1.90	$5.17 \times 10^{-7}$	1.58	1.31	1.91	$2.73 \times 10^{-6}$	$9.85 \times 10^{-1}$	(30 kb) <i>C11orf30</i>   (43 kb) <i>LOC101928813</i>
rs78048444	7	132832218	C/T	0.02	Genotyped	0.22	0.12	0.39	$5.44 \times 10^{-7}$	0.23	0.11	0.46	$4.57 \times 10^{-5}$	$9.35 \times 10^{-1}$	(65 kb) <i>CHCHD3</i>   (106 kb) <i>EXOC4</i>
rs56151068	17	46381431	T/C	0.10	Imputed	2.06	1.54	2.76	$9.58 \times 10^{-7}$	1.97	1.44	2.70	$2.34 \times 10^{-5}$	$8.39 \times 10^{-1}$	<i>SKAP1</i> ; <i>LOC101927148</i>
rs139462954	17	46523678	A/AC	0.09	Imputed	2.06	1.54	2.76	$1.23 \times 10^{-6}$	1.97	1.43	2.71	$2.92 \times 10^{-5}$	$8.37 \times 10^{-1}$	<i>LOC101927166</i>

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 and see Table E5). SNPs in *EXOC4*, *ARHGAP24*, *SKAP1*, and *CTNNA3* were not suggestive of significance for food allergy.

### 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 (*SNX11*) 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  $1.0 \times 10^{-6}$  (Table III and see Table E7 in this article's Online Repository at [www.jacionline.org](http://www.jacionline.org)). Little is known about *SNX11*; it belongs to a family of retrograde transport molecules,<sup>31</sup> and its protein is involved in targeting cell-surface molecules to the lysosome.<sup>32</sup> *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.<sup>33,34</sup> 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 and PA meta-analysis 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 in both the PA GWAS and meta-analyses, the rank order of these 2 loci changed between the related and unrelated analyses, but no

significant changes in the ORs were observed. *SKAP1*, *CTNNA3*, and *ARHGAP24* were identified as suggestive for association with PA in the CanPAR GWAS but were not suggestive for association with PA in the meta-analysis. This is likely due to the small sample size and limited number of contributing studies, as evidenced by the minimal change in *P* values for these loci in CanPAR GWAS data compared with the meta-analysis.

Loci in *ANGPT4*, *C11orf30*, and *EXOC4* were suggestive of significance in a meta-analysis for the phenotype of PA but did not reach genome-wide significance, likely because of small PA sample sizes and heterogeneity in study designs (case-control vs family-based studies) and ascertainment criteria. Phenotype definitions of both cases and control subjects differed in each population: cases were defined by food challenge (IDEAL and GENEVA), food challenge and history with confirmatory testing (CanPAR, CFA, and HealthNuts studies), or self-report (GERA cohort), whereas control subjects ranged from hyper-control subjects (CanPAR study), control subjects without food allergy (IDEAL and GENEVA), control subjects without positive skin prick test responses to foods (HealthNuts study), population-based control subjects (CanPAR study, GERA cohort, and UFA study), and subjects with other food allergies (CFA study). Ethnicity was diverse across studies, with variable analytic methods used to control for it; age also differed because of the use of population-based control subjects. The Busselton cohort is a longitudinal data set, use of which is required for hyper-control subjects because they must have negative results for allergic phenotypes and inherently be older to ensure they will not have eczema, asthma, or other allergic phenotypes. Our efforts to evaluate effect sizes across studies and populations were additionally hindered by differences in GWAS chip and imputation reference panels. *C11orf30* is a prime example: it was identified in the CanPAR GWAS with PA, but it only reaches genome-wide significance ( $P = 7.50 \times 10^{-11}$ ) in the meta-analysis of food allergy. This is likely due to the small PA sample sizes in the other studies. This finding has important implications for other loci identified in the CanPAR GWAS: the lack of

**TABLE II.** Meta-analysis of Canadian, American, Australian, German, and Dutch populations for association with PA and food allergy phenotypes

SNP*	Chromosome	Allele	PA					P-Dutch GENEVA, family based <sup>f</sup>
			P-CanPAR <sup>†a</sup>	P-CFA <sup>b</sup>	P-HealthNuts <sup>†c</sup>	P-UFA <sup>†d</sup>	P-IDEAL/GENEVA, case-control <sup>e</sup>	
rs115218289 <sup>‡</sup>	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 <sup>§</sup>	20	C/T	$4.42 \times 10^{-7}$	NA	NA	NA	$8.33 \times 10^{-1}$	$1.63 \times 10^{-2}$
rs144897250 <sup>‡</sup>	11	A/C	$2.90 \times 10^{-7}$	NA	$3.84 \times 10^{-1}$	NA	NA	NA
rs7936434 <sup>‡</sup>	11	C/G	$5.17 \times 10^{-7}$	$3.66 \times 10^{-2}$	$1.43 \times 10^{-1}$	NA	NA	NA
rs78048444 <sup>§</sup>	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 <sup>§</sup>	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 <sup>‡</sup>	17	T/C	$2.60 \times 10^{-7}$	$1.36 \times 10^{-1}$	$3.55 \times 10^{-1}$	NA	NA	NA
rs55765969 <sup>‡</sup>	17	T/C	$1.23 \times 10^{-6}$	$1.97 \times 10^{-1}$	$2.81 \times 10^{-1}$	NA	NA	NA
rs56151068 <sup>‡</sup>	17	T/C	$9.58 \times 10^{-7}$	$2.51 \times 10^{-1}$	$3.57 \times 10^{-1}$	NA	NA	NA
rs71193762 <sup>‡</sup>	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 ( $P \leq 1.49 \times 10^{-6}$ ) in patients with PA.

<sup>a</sup>P value from CanPAR (n = 1,776).

<sup>b</sup>P value from the Chicago Food Allergy Study (n = 2,197).

<sup>c</sup>P value from the Australian HealthNuts study (n = 221).

<sup>d</sup>P value from the German Understanding of Food Allergy study (n = 2,592).

<sup>e</sup>Number of subjects = 226, 229, 227, and 217 for rs115218289, rs523865, rs78048444, and rs744597, respectively, corrected for atopic dermatitis, Asthma, and rhinoconjunctivitis.

<sup>f</sup>P value from the Dutch GENEVA family study; number of informative families = 20, 112, 21, and 112 for rs115218289, rs523865, rs78048444, and rs744597, respectively.

<sup>g</sup>P value from the Stouffer weighted z score meta-analysis method for PA.

<sup>h</sup>P value from the Chicago Food Allergy Study (n = 2,197).

<sup>i</sup>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, corrected for atopic dermatitis, Asthma, and rhinoconjunctivitis.

<sup>j</sup>P value from the Dutch GENEVA study; number of informative families for SNPs = 26, 196, 37, and 214 for rs115218289, rs523865, rs78048444, and rs744597, respectively.

<sup>k</sup>P value from the GERA food allergy study (n = 29,053).

<sup>l</sup>P value from the Stouffer weighted z score meta-analysis method for food allergy.

<sup>m</sup>P value from the Stouffer weighted z score meta-analysis method for food allergy without the GERA cohort.

\*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).

<sup>†</sup>Used for both PA and food allergy (FA).

<sup>‡</sup>Imputed SNP from CanPAR.

<sup>§</sup>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<sup>35</sup> and asthma.<sup>36</sup> Recently, *C11orf30* was identified as a risk factor for eosinophilic esophagitis (OR, 2.22;  $P = 5.38 \times 10^{-10}$ ),<sup>37</sup> a chronic allergic inflammatory disease of the esophagus that is mainly triggered by food proteins. Loci in this region have been significantly associated with atopic dermatitis<sup>38</sup> both with (rs2155219)<sup>39,40</sup> and without (rs7927894) any other disease-related phenotypes, such as asthma, allergic rhinoconjunctivitis, total serum IgE level, or family history of atopy.<sup>41</sup> These findings indicate that *C11orf30* is a risk factor for the atopic march, particularly those studies that investigated childhood eczema with later development of asthma.<sup>38</sup> *FLG* and *HLA-DQB1*, genes previously examined in CanPAR,<sup>18,19,22</sup> have similarly been found to be associated with childhood eczema and asthma.<sup>38</sup>

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.<sup>42</sup> 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.<sup>43</sup> *C11orf30* is associated with Crohn disease and ulcerative colitis,<sup>44,45</sup> 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.<sup>46,47</sup> 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

PA	Food allergy						Gene/nearest gene
<i>P</i> meta_PA <sup>g</sup>	<i>P</i> -CFA <sup>h</sup>	<i>P</i> -IDEAL/GENEVA, case-control <sup>i</sup>	<i>P</i> -GENEVA, family study <sup>j</sup>	<i>P</i> -GERA <sup>k</sup>	<i>P</i> meta <sup>l</sup>	<i>P</i> meta (without GERA) <sup>m</sup>	
$9.16 \times 10^{-8}$	NA	$1.84 \times 10^{-1}$	$1.90 \times 10^{-1}$	$5.25 \times 10^{-1}$	$2.91 \times 10^{-2}$	$2.38 \times 10^{-8}$	(298 kb) <i>DLX2</i>   (26 kb) <i>ITGA6</i>
$1.54 \times 10^{-7}$	NA	$2.03 \times 10^{-1}$	$2.60 \times 10^{-2}$	$2.66 \times 10^{-1}$	$9.29 \times 10^{-3}$	$4.09 \times 10^{-8}$	<i>ANGPT4</i>
$2.94 \times 10^{-7}$	NA	NA	NA	$5.89 \times 10^{-1}$	$6.83 \times 10^{-2}$	$2.94 \times 10^{-7}$	(5 kb) <i>MMP12</i>   (63 kb) <i>MMP13</i>
$3.13 \times 10^{-7}$	$5.89 \times 10^{-5}$	NA	NA	$4.13 \times 10^{-4}$	$1.98 \times 10^{-8}$	$7.50 \times 10^{-11}$	(30 kb) <i>C11orf30</i>   (43 kb) <i>LOC101928813</i>
$3.73 \times 10^{-7}$	NA	$8.29 \times 10^{-1}$	$7.87 \times 10^{-1}$	$6.97 \times 10^{-1}$	$8.88 \times 10^{-2}$	$2.53 \times 10^{-6}$	(65 kb) <i>CHCHD3</i>   (106 kb) <i>EXOC4</i>
$1.63 \times 10^{-6}$	$5.19 \times 10^{-1}$	$1.13 \times 10^{-1}$	$2.69 \times 10^{-1}$	$3.63 \times 10^{-1}$	$1.29 \times 10^{-2}$	$1.42 \times 10^{-5}$	<i>ARHGAP24</i>
$3.43 \times 10^{-6}$	$5.83 \times 10^{-1}$	NA	NA	$7.69 \times 10^{-1}$	$9.27 \times 10^{-2}$	$7.47 \times 10^{-5}$	<i>SKAP1</i>
$1.45 \times 10^{-5}$	$6.59 \times 10^{-1}$	NA	NA	$3.77 \times 10^{-1}$	$3.18 \times 10^{-2}$	$1.97 \times 10^{-4}$	<i>LOC101927166</i>
$2.33 \times 10^{-5}$	$7.33 \times 10^{-1}$	NA	NA	$4.45 \times 10^{-1}$	$4.45 \times 10^{-2}$	$2.66 \times 10^{-4}$	<i>SKAP1</i> ; <i>LOC101927148</i>
$2.73 \times 10^{-4}$	$6.87 \times 10^{-1}$	NA	NA	$7.06 \times 10^{-1}$	$8.83 \times 10^{-2}$	$1.41 \times 10^{-4}$	<i>CTNNA3</i>

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<sup>20,21,29</sup>; 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.

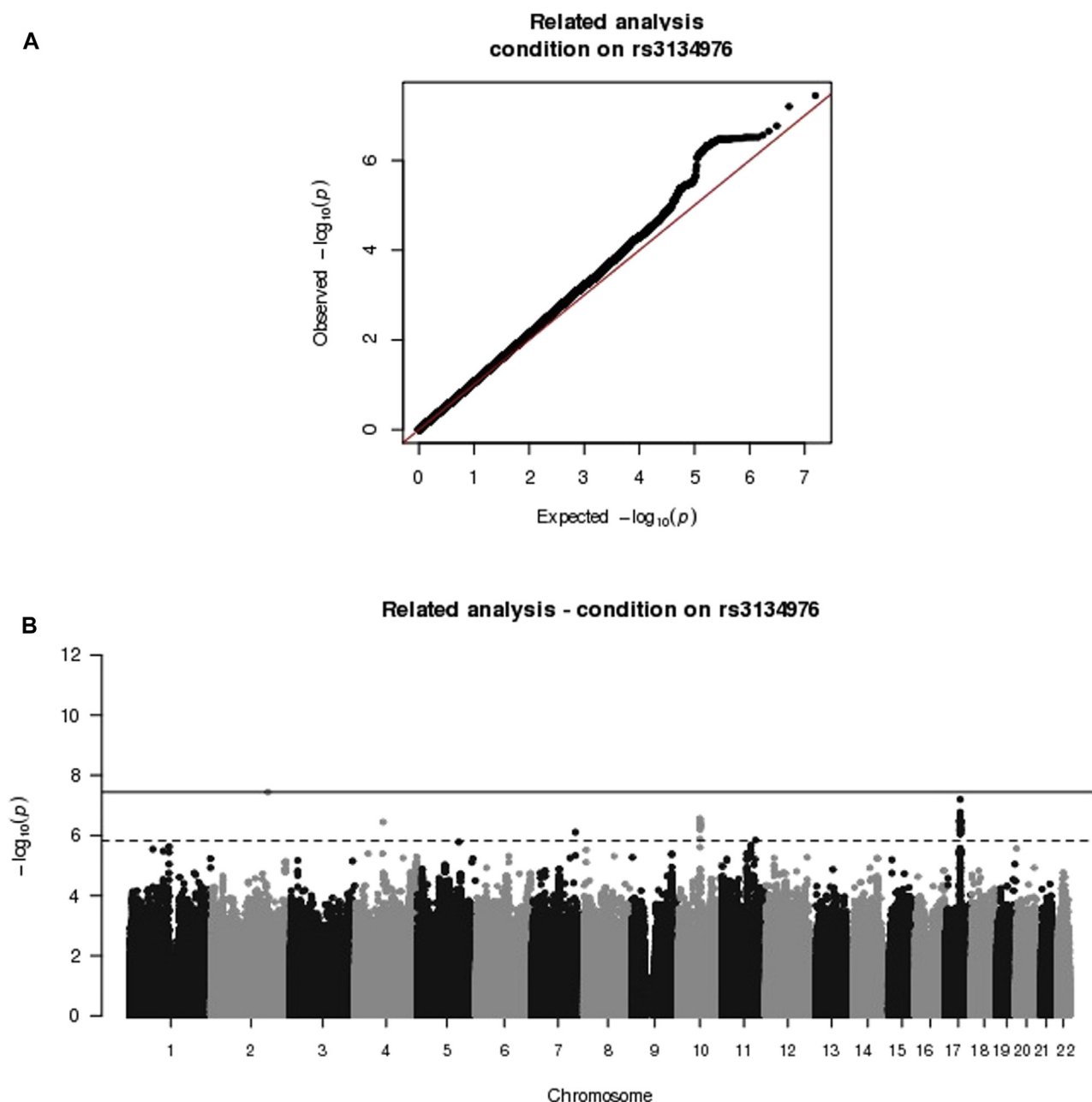
### Novel pathways in the pathogenesis of food allergy

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.<sup>20,48</sup> The protein complex formed by KDM5A, a histone demethylase, and EMSY, the protein encoded by *C11orf30*, appears to increase gene transcription.<sup>49</sup> EMSY acts

as a reader for trimethylation of lysine 4 on histone protein H3 (H3K4me3)<sup>50</sup> 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.<sup>49</sup>

Along with members of the EMSY protein complex, EMSY colocalizes with CBX1.<sup>33</sup> Similar to its highly conserved *Drosophila* homolog protein dHP1, CBX1 protein has a binding affinity for trimethylation of lysine 9 on histone protein H3 (H3K9me3),<sup>51</sup> modifications that are linked to transcription repression.<sup>52</sup> In this study *CBX1* was identified as being regulated by *SKAP1* because the same SNPs that are associated with PA also regulate the gene expression of *CBX1*, increasing the connections between the novel loci identified in the CanPAR GWAS. *SKAP1*, 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,<sup>33,34</sup> the product of which colocalizes with EMSY protein, as encoded by *C11orf30*.

*CTNNA3*, also identified as a PA locus in this study, has an identified copy number variant (CNV; CNVR 6828297068284017)<sup>53</sup> in pediatric food allergy. It is also tied to histone modification because there is an enrichment of enhancer- and promoter-associated



**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 ( $P \leq 3.60 \times 10^{-8}$ ), and the *dashed line* indicates the suggestive association significance threshold ( $P \leq 1.49 \times 10^{-6}$ ).

histone marker H3K4me1 (monomethylation of lysine 4 on histone protein 3) in *CTNNA3* CNVR.<sup>53</sup>

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,<sup>7</sup> and maintenance of peanut consumption promotes continued tolerance of peanut<sup>8</sup>; 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 through 2 putative mechanisms: (1) an endothelial barrier defect



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

SNP	Chromosome	Position	Nearest gene	Tissue	P value	Gene symbol
rs4491576	17	46,408,636	<i>SKAP1</i>	Whole blood	$1.05 \times 10^{-10}$	<i>SNX11</i>
rs16956501	17	46,497,274	<i>SKAP1</i>	Skin, sun-exposed lower leg	$4.48 \times 10^{-10}$	<i>SNX11</i>
rs139462954	17	46,523,678	LOC101927166	Skin, sun-exposed lower leg	$8.94 \times 10^{-10}$	<i>SNX11</i>
rs139462954	17	46,523,678	LOC101927166	Cells, transformed fibroblasts	$7.36 \times 10^{-7}$	<i>CBX1</i>

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.<sup>54</sup> Several identified loci from this study have connections to both allergy and vascular regulation (*CTNNA3*),<sup>55</sup> permeability (*ANGPT4*),<sup>56,57</sup> or endothelial cell function (*SKAP1*<sup>58</sup> and *EXOC4*<sup>59</sup>). 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%),<sup>2</sup> 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.<sup>60</sup> 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<sup>35-41</sup> 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

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 Québécois 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, Chynace Van Lambalgen, Heather Waldhauser, 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

1. Sicherer SH, Munoz-Furlong A, Godbold JH, Sampson HA. US prevalence of self-reported peanut, tree nut, and sesame allergy: 11-year follow-up. *J Allergy Clin Immunol* 2010;125:1322-6.
2. Soller L, Ben-Shoshan M, Harrington DW, Knoll M, Fragapane J, Joseph L, et al. Adjusting for nonresponse bias corrects overestimates of food allergy prevalence. *J Allergy Clin Immunol Pract* 2015;3:291-3.e2.
3. Osborne NJ, Koplin JJ, Martin PE, Gurrin LC, Lowe AJ, Matheson MC, et al. Prevalence of challenge-proven IgE-mediated food allergy using population-based sampling and predetermined challenge criteria in infants. *J Allergy Clin Immunol* 2011;127:668-76, e1-2.
4. Sicherer SH, Furlong TJ, Maes HH, Desnick RJ, Sampson HA, Gelb BD. Genetics of peanut allergy: a twin study. *J Allergy Clin Immunol* 2000;106:53-6.
5. Liem JJ, Huq S, Kozyskyj AL, Becker AB. Should younger siblings of peanut-allergic children be assessed by an allergist before being fed peanut? *Allergy Asthma Clin Immunol* 2008;4:144-9.
6. Hourihane JO, Dean TP, Warner JO. Peanut allergy in relation to heredity, maternal diet, and other atopic diseases: results of a questionnaire survey, skin prick testing, and food challenges. *BMJ* 1996;313:518-21.
7. Du Toit G, Roberts G, Sayre PH, Bahnson HT, Radulovic S, Santos AF, et al. Randomized trial of peanut consumption in infants at risk for peanut allergy. *N Engl J Med* 2015;372:803-13.
8. Du Toit G, Sayre PH, Roberts G, Sever ML, Lawson K, Bahnson HT, et al. Effect of avoidance on peanut allergy after early peanut consumption. *N Engl J Med* 2016;374:1435-43.
9. Silva R, Gomes E, Cunha L, Falcao H. Anaphylaxis in children: a nine years retrospective study (2001-2009). *Allergol Immunopathol (Madr)* 2012;40:31-6.
10. Hoffer V, Scheuerman O, Marcus N, Levy Y, Segal N, Lagovsky I, et al. Anaphylaxis in Israel: experience with 92 hospitalized children. *Pediatr Allergy Immunol* 2011;22:172-7.
11. Asero R, Antonicecchi L, Arena A, Bommario L, Caruso B, Colombo G, et al. Causes of food-induced anaphylaxis in Italian adults: a multi-centre study. *Int Arch Allergy Immunol* 2009;150:271-7.
12. Boehncke WH, Loeliger C, Kuehn P, Kalbacher H, Bohm BO, Gall H. Identification of HLA-DR and -DQ alleles conferring susceptibility to pollen allergy and pollen associated food allergy. *Clin Exp Allergy* 1998;28:434-41.
13. Howell WM, Turner SJ, Hourihane JO, Dean TP, Warner JO. HLA class II DRB1, DQB1 and DPB1 genotypic associations with peanut allergy: evidence from a family-based and case-control study. *Clin Exp Allergy* 1998;28:156-62.
14. Dreskin SC. Do HLA genes play a role in the genetics of peanut allergy? *Ann Allergy Asthma Immunol* 2006;96:766-8.
15. Madore AM, Vaillancourt VT, Asai Y, Alizadehfar R, Ben-Shoshan M, Michel DL, et al. HLA-DQB1\*02 and DQB1\*06:03P are associated with peanut allergy. *Eur J Hum Genet* 2013;21:1181-4.
16. Brough HA, Cousins DJ, Munteanu A, Wong YF, Sudra A, Makinson K, et al. IL-9 is a key component of memory TH cell peanut-specific responses from children with peanut allergy. *J Allergy Clin Immunol* 2014;134:1329-38.e10.
17. Dreskin SC, Ayars A, Jin Y, Atkins D, Leo HL, Song B. Association of genetic variants of CD14 with peanut allergy and elevated IgE levels in peanut allergic individuals. *Ann Allergy Asthma Immunol* 2011;106:170-2.
18. Brown SJ, Asai Y, Cordell HJ, Campbell LE, Zhao Y, Liao H, et al. Loss-of-function variants in the flaggrin gene are a significant risk factor for peanut allergy. *J Allergy Clin Immunol* 2011;127:661-7.
19. Asai Y, Greenwood C, Hull PR, Alizadehfar R, Ben-Shoshan M, Brown SJ, et al. Filaggrin gene mutation associations with peanut allergy persist despite variations in peanut allergy diagnostic criteria or asthma status. *J Allergy Clin Immunol* 2013;132:239-42.
20. Hong X, Hao K, Ladd-Acosta C, Hansen KD, Tsai HJ, Liu X, et al. Genome-wide association study identifies peanut allergy-specific loci and evidence of epigenetic mediation in US children. *Nat Commun* 2015;6:6304.
21. Martino DJ, Ashley S, Koplin J, Ellis J, Saffery R, Dharmage SC, et al. Genome-wide association study of peanut allergy reproduces association with amino acid polymorphisms in HLA-DRB1. *Clin Exp Allergy* 2017;47:217-23.
22. Asai Y, Eslami A, van Ginkel CD, Akhbari L, Wan M, Yin D, et al. Canadian genome-wide association study and meta-analysis confirm HLA as a risk factor for peanut allergy. *J Allergy Clin Immunol* 2017 [in press].
23. James AL, Knuiman MW, Divitini ML, Hui J, Hunter M, Palmer LJ, et al. Changes in the prevalence of asthma in adults since 1966: the Busselton health study. *Eur Respir J* 2010;35:273-8.
24. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinformatics* 2010;26:2867-73.
25. Conomos MP, Miller MB, Thornton TA. Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. *Genet Epidemiol* 2015;39:276-93.
26. StataCorp. Stata statistical software: release 11. College Station (TX): StataCorp LP; 2009.
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.
28. Hoffmann TJ, Kvale MN, Hesselson SE, Zhan Y, Aquino C, Cao Y, et al. Next generation genome-wide association tool: design and coverage of a high-throughput European-optimized SNP array. *Genomics* 2011;98:79-89.
29. van Ginkel CD, Flokstra-de Blok BM, Kollen BJ, Kukler J, Koppelman GH, Dubois AE. Loss-of-function variants of the flaggrin gene are associated with clinical reactivity to foods. *Allergy* 2015;70:461-4.
30. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 2013;45:580-5.
31. Xu J, Xu T, Wu B, Ye Y, You X, Shu X, et al. Structure of sorting nexin 11 (SNX11) reveals a novel extended phox homology (PX) domain critical for inhibition of SNX10-induced vacuolation. *J Biol Chem* 2013;288:16598-605.
32. Li C, Ma W, Yin S, Liang X, Shu X, Pei D, et al. Sorting nexin 11 regulates lysosomal degradation of plasma membrane TRPV3. *Traffic* 2016;17:500-14.
33. Hughes-Davies L, Huntsman D, Ruas M, Fuks F, Bye J, Chin SF, et al. EMSY links the BRCA2 pathway to sporadic breast and ovarian cancer. *Cell* 2003;115:523-35.
34. Lev Maor G, Yearim A, Ast G. The alternative role of DNA methylation in splicing regulation. *Trends Genet* 2015;31:274-80.

35. Li X, Ampleford EJ, Howard TD, Moore WC, Li H, Busse WW, et al. The C11orf30-LRRC32 region is associated with total serum IgE levels in asthmatic patients. *J Allergy Clin Immunol* 2012;129:575-8, e1-9.
36. Li X, Hastie AT, Hawkins GA, Moore WC, Ampleford EJ, Milosevic J, et al. eQTL of bronchial epithelial cells and bronchial alveolar lavage deciphers GWAS-identified asthma genes. *Allergy* 2015;70:1309-18.
37. Sleiman PM, Wang ML, Cianferoni A, Aceves S, Gonsalves N, Nadeau K, et al. GWAS identifies four novel eosinophilic esophagitis loci. *Nat Commun* 2014;5:5593.
38. Manz J, Rodriguez E, ElSharawy A, Oesau EM, Petersen BS, Baurecht H, et al. Targeted resequencing and functional testing identifies low-frequency missense variants in the gene encoding GARP as significant contributors to atopic dermatitis risk. *J Invest Dermatol* 2016;136:2380-6.
39. Marenholz I, Esparza-Gordillo J, Ruschendorf F, Bauerfeind A, Strachan DP, Spycher BD, et al. Meta-analysis identifies seven susceptibility loci involved in the atopic march. *Nat Commun* 2015;6:8804.
40. Weidinger S, Willis-Owen SA, Kamatani Y, Baurecht H, Morar N, Liang L, et al. A genome-wide association study of atopic dermatitis identifies loci with overlapping effects on asthma and psoriasis. *Hum Mol Genet* 2013;22:4841-56.
41. Greisenegger EK, Zimprich F, Zimprich A, Gleiss A, Kopp T. Association of the chromosome 11q13.5 variant with atopic dermatitis in Austrian patients. *Eur J Dermatol* 2013;23:142-5.
42. Rona RJ, Keil T, Summers C, Gislason D, Zuidmeer L, Sodergren E, et al. The prevalence of food allergy: a meta-analysis. *J Allergy Clin Immunol* 2007;120:638-46.
43. Amaral AF, Minelli C, Guerra S, Wjst M, Probst-Hensch N, Pin I, et al. The locus C11orf30 increases susceptibility to poly-sensitization. *Allergy* 2015;70:328-33.
44. Dubinsky MC, Kugathasan S, Kwon S, Haritunians T, Wrobel I, Wahbeh G, et al. Multidimensional prognostic risk assessment identifies association between IL12B variation and surgery in Crohn's disease. *Inflamm Bowel Dis* 2013;19:1662-70.
45. Waterman M, Xu W, Stempak JM, Milgrom R, Bernstein CN, Griffiths AM, et al. Distinct and overlapping genetic loci in Crohn's disease and ulcerative colitis: correlations with pathogenesis. *Inflamm Bowel Dis* 2011;17:1936-42.
46. Schmitt J, Schwarz K, Baurecht H, Hotze M, Folster-Holst R, Rodriguez E, et al. Atopic dermatitis is associated with an increased risk for rheumatoid arthritis and inflammatory bowel disease, and a decreased risk for type 1 diabetes. *J Allergy Clin Immunol* 2016;137:130-6.
47. Kim M, Choi KH, Hwang SW, Lee YB, Park HJ, Bae JM. Inflammatory bowel disease is associated with an increased risk of inflammatory skin diseases: A population-based cross-sectional study. *J Am Acad Dermatol* 2017;76:40-8.
48. Song Y, Liu C, Hui Y, Srivastava K, Zhou Z, Chen J, et al. Maternal allergy increases susceptibility to offspring allergy in association with TH2-biased epigenetic alterations in a mouse model of peanut allergy. *J Allergy Clin Immunol* 2014;134:1339-45.e7.
49. Varier RA, Carrillo de Santa Pau E, van der Groep P, Lindeboom RG, Matarese F, Mensinga A, et al. Recruitment of the mammalian histone-modifying EMSY complex to target genes is regulated by ZNF131. *J Biol Chem* 2016;291:7313-24.
50. Vermeulen M, Eberl HC, Matarese F, Marks H, Denissov S, Butter F, et al. Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell* 2010;142:967-80.
51. Kaustov L, Ouyang H, Amaya M, Lemak A, Nady N, Duan S, et al. Recognition and specificity determinants of the human cbx chromodomains. *J Biol Chem* 2011;286:521-9.
52. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. *Cell* 2007;129:823-37.
53. Li J, Fung I, Glessner JT, Pandey R, Wei Z, Bakay M, et al. Copy number variations in CTNNA3 and RBFOX1 associate with pediatric food allergy. *J Immunol* 2015;195:1599-607.
54. Hirose S, Vokali E, Raghavan VR, Rincon-Restrepo M, Lund AW, Cortes-Henrioud P, et al. Steady-state antigen scavenging, cross-presentation, and CD8+ T cell priming: a new role for lymphatic endothelial cells. *J Immunol* 2014;192:5002-11.
55. Folmsbee SS, Budinger GR, Bryce PJ, Gottardi CJ. The cardiomyocyte protein  $\alpha$ T-catenin contributes to asthma through regulating pulmonary vein inflammation. *J Allergy Clin Immunol* 2016;138:123-9.e2.
56. Kesler CT, Pereira ER, Cui CH, Nelson GM, Masuck DJ, Baish JW, et al. Angiopoietin-4 increases permeability of blood vessels and promotes lymphatic dilation. *FASEB J* 2015;29:3668-77.
57. Yao JH, Cui M, Li MT, Liu YN, He QH, Xiao JJ, et al. Angiopoietin1 inhibits mast cell activation and protects against anaphylaxis. *PLoS One* 2014;9:e89148.
58. Jiang L, Hu J, Feng J, Han D, Yang C. Substrate stiffness of endothelial cells directs LFA-1/ICAM-1 interaction: a physical trigger of immune-related diseases? *Clin Hemorheol Microcirc* 2016;61:633-43.
59. Barkefors I, Fuchs PF, Heldin J, Bergstrom T, Forsberg-Nilsson K, Kreuger J. Exocyst complex component 3-like 2 (EXOC3L2) associates with the exocyst complex and mediates directional migration of endothelial cells. *J Biol Chem* 2011;286:24189-99.
60. Edwards BJ, Haynes C, Levenstien MA, Finch SJ, Gordon D. Power and sample size calculations in the presence of phenotype errors for case/control genetic association studies. *BMC Genet* 2005;6:18.