

Millions of dots,
each with a unique story



Copy Number Variations in *CTNNA3* and *RBF1* Associate with Pediatric Food Allergy

This information is current as of July 30, 2015.

Jin Li, Irene Fung, Joseph T. Glessner, Rahul Pandey, Zhi Wei, Marina Bakay, Frank D. Mentch, Renata Pellegrino, Tiancheng Wang, Cecilia Kim, Cuiping Hou, Fengxiang Wang, Rosetta M. Chiavacci, Kelly A. Thomas, Jonathan M. Spergel, Hakon Hakonarson and Patrick M. A. Sleiman

J Immunol published online 17 July 2015
<http://www.jimmunol.org/content/early/2015/07/17/jimmunol.1402310>

-
- | | |
|-------------------------------|---|
| Supplementary Material | http://www.jimmunol.org/content/suppl/2015/07/17/jimmunol.1402310.DCSupplemental.html |
| Subscriptions | Information about subscribing to <i>The Journal of Immunology</i> is online at: http://jimmunol.org/subscriptions |
| Permissions | Submit copyright permission requests at: http://www.aai.org/ji/copyright.html |
| Email Alerts | Receive free email-alerts when new articles cite this article. Sign up at: http://jimmunol.org/cgi/alerts/etoc |



Copy Number Variations in *CTNNA3* and *RBFOX1* Associate with Pediatric Food Allergy

Jin Li,^{*,1} Irene Fung,^{†,1,2} Joseph T. Glessner,^{*} Rahul Pandey,^{*} Zhi Wei,[‡] Marina Bakay,^{*} Frank D. Mentch,^{*} Renata Pellegrino,^{*} Tiancheng Wang,^{*} Cecilia Kim,^{*} Cuiping Hou,^{*} Fengxiang Wang,^{*} Rosetta M. Chiavacci,^{*} Kelly A. Thomas,^{*} Jonathan M. Spergel,^{†,§,¶} Hakon Hakonarson,^{*,§,||} and Patrick M. A. Sleiman^{*,§,||}

Food allergy is a significant public health concern, especially among children. Previous candidate gene studies suggested a few susceptibility loci for food allergy, but no study investigated the contribution of copy number variations (CNVs) to food allergy on a genome-wide scale. To investigate the genetics of food allergy, we performed CNV assessment using high-resolution genome-wide single nucleotide polymorphism arrays. CNV calls from a total of 357 cases with confirmed food allergy and 3980 controls were analyzed within a discovery cohort, followed by a replication analysis composed of 167 cases and 1573 controls. We identified that CNVs in *CTNNA3* were significantly associated with food allergy in both the discovery cohort and the replication cohort. Of particular interest, *CTNNA3* CNVs hit exons or intron regions rich in histone marker H3K4Me1. CNVs in a second gene (*RBFOX1*) showed a significant association ($p = 7.35 \times 10^{-5}$) with food allergy at the genome-wide level in our meta-analysis of the European ancestry cohorts. The presence of these CNVs was confirmed by quantitative PCR. Furthermore, knockdown of *CTNNA3* resulted in upregulation of CD63 and CD203c in mononuclear cells upon PMA stimulation, suggesting a role in sensitization to allergen. We uncovered at least two plausible genes harboring CNV loci that are enriched in pediatric patients with food allergies. The novel gene candidates discovered in this study by genome-wide CNV analysis are compelling drug and diagnostic targets for food allergy. *The Journal of Immunology*, 2015, 195: 000–000.

Approximately 8% of children in the United States have a food allergy (1, 2), and the incidence is increasing (3, 4). As with other complex diseases, the etiology of a food allergy phenotype likely arises from a combination of multiple genetic factors and environmental triggers (5). Heritability studies suggest that genetic factors may play a significant role in predisposition to food allergy. An increased prevalence of peanut allergy was reported in siblings of affected children rela-

tive to the general population (6), and estimates of the heritability from a twin study are as high as 81.6% (7). Genetic susceptibility was reported previously through linkage studies (8, 9), and several food allergy genes of interest have been identified through candidate gene studies (reviewed in Ref. 10).

Copy number variations (CNVs) are defined as deletions, duplication, inversions, or complex rearrangements in genomic segments compared with the reference genome (5, 11). They are generated through errors in recombination or DNA replication and repair (reviewed in Ref. 12). When CNVs occur in functional areas of the genome, such as protein coding or regulatory regions, they can result in a phenotype through dysregulation of RNA levels (13) or by altering the structure of a gene and, hence, its protein product (14).

In clinical practice, the identification of common CNVs underlying intellectual disability, autism spectrum disorders, and multiple congenital anomalies has had a major impact (reviewed in Ref. 15). In the allergy field, CNVs within the filaggrin gene were shown to contribute to atopic dermatitis risk in a dose-dependent manner (16). No published study has examined CNVs in food allergy patients. In this study, our investigation of the association of CNVs in pediatric food allergy resulted in the identification of CNV enrichment in two genes that may contribute to food allergy susceptibility.

Materials and Methods

Research subjects

Individuals were selected from a population of research subjects from the Center for Applied Genomics. This population consisted of children recruited from the Children's Hospital of Philadelphia Health Care Network who permitted use of their health care records and genotyping data available at the center. Blood for genotyping was collected from phlebotomy collection sites within the hospital. Only individuals whose blood samples resulted in high-quality genome-wide genotyping data were included in our study.

^{*}Center for Applied Genomics, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, PA 19104; [†]Division of Allergy and Immunology, The Children's Hospital of Philadelphia, Philadelphia, PA 19104; [‡]Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 08540; [§]Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104; [¶]Center for Pediatric Eosinophilic Disorders, The Children's Hospital of Philadelphia, Philadelphia, PA 19104; and ^{||}Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA 19104

¹J.L. and I.F. contributed equally to this work.

²Current address: Irene Ngar Yee Fung Medicine Professional Corporation, Etobicoke, ON, Canada.

Received for publication September 9, 2014. Accepted for publication May 20, 2015.

This work was supported by a gift from the Kubert Family Estate, Pennsylvania Department of Health Grant 4100042728, U.S. Department of Defense Grant W81XWH-11-1-0507, and by an Institutional Development Fund to the Center for Applied Genomics from the Children's Hospital of Philadelphia.

Address correspondence and reprint requests to Dr. Hakon Hakonarson and Dr. Patrick M.A. Sleiman, Center for Applied Genomics, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, PA 19104. E-mail addresses: hakonarson@email.chop.edu (H.H.) and sleimanp@email.chop.edu (P.M.A.S.)

The online version of this article contains supplemental material.

Abbreviations used in this article: CN, copy number; CNV, copy number variation; CNVR, CNV region; GCWF, GC base pair wave factor; GWAS, genome-wide association study; KD, knockdown; LRR, log R ratio; PCA, principal component analysis; qPCR, quantitative PCR; siRNA, small interfering RNA; SNP, single nucleotide polymorphism.

Copyright © 2015 by The American Association of Immunologists, Inc. 0022-1767/15/\$25.00

Ethics statement

This study was approved by the institutional review board of the Children's Hospital of Philadelphia. All participants and/or their parents signed an informed consent allowing use of this information.

Food allergy population

Children with food allergy were first identified among research subjects if they had any of the following *International Classification of Diseases, Ninth Revision* codes of food allergy within their medical records: 693.1: food allergy, dermatitis due to food taken internally, or 995.7: other adverse food reactions NOS. Chart review was performed to verify candidacy. To be included as a case patient, an individual had to meet the following criteria: history consistent with food allergy documented by an allergy specialist [timing and symptoms after ingestion as per Sampson et al. (17)]; food allergy sensitivity (as shown by IgE > 0.35 kU_A/l or skin prick testing [> 3 mm wheal]); and food allergen identified as milk, egg, peanut, tree nut, soy, wheat, fish, or shellfish.

Healthy controls

To be a control subject in our study, individuals could not have *International Classification of Diseases, Ninth Revision* codes attached to their chart that identified food allergy, as listed above. We selected controls with average ages older than the cases to reduce the possibility of an individual developing the disease at a later stage.

Genotyping

Genotyping of all cases and controls was performed at the Center for Applied Genomics, Children's Hospital of Philadelphia. Genotyping of food allergy cases and controls of the discovery cohort was done on the Illumina HumanHap550 BeadChip or Illumina Human610-Quad BeadChip. The change to the Human610-Quad BeadChip for analysis occurred after production of the HumanHap550 BeadChip ceased. Because the two chip types have $> 500,000$ common single nucleotide polymorphisms (SNPs), data from the two chips could be combined and analyzed together. After production of the Human610-Quad BeadChip ceased, all subsequent samples were analyzed using the Illumina HumanOmniExpress-12 v1.0 chip. All food allergy cases and controls forming the replication cohort of our study were genotyped using this latest chip.

Principal component analysis

We used principal component analysis (PCA) to infer population structure among cases and controls, as well as to correct for potential population stratification. The principal components were generated on ancestry-informative markers using the EIGENSTRAT (18) software package.

CNV detection

For the discovery cohort, the $\sim 530,000$ SNPs common to the Illumina HumanHap550 BeadChip or the Illumina Human610-Quad BeadChip were used for CNV calling. The PennCNV (19) software package was used to call CNVs based on the intensity data from the Illumina arrays, including the log R ratio (LRR) and the B allele frequency for each SNP. Because we focused on markers on the autosomes, 522,879 autosomal SNPs were used for CNV calling. Similarly, 712,726 autosomal SNPs on the Illumina HumanOmniExpress-12 v1.0 chip were used for CNV calling for the replication cohort.

CNV quality control

As described previously (20), CNV metrics were reviewed, and samples with outlier values in the following categories were removed: call rate, cryptic relatedness between samples, intensity noise measured by the SD of LRR (LRR SD), intensity waviness measured by |GC base pair wave factor (GCWF)|, and high number of CNVs/sample. The distribution of the CNV metrics was plotted, the distribution plots typically display a linear phase and an exponential phase, and samples with CNV metrics falling into the exponential phase were removed. For both the discovery cohort and the replication cohort, samples with SNP call rate $< 98\%$ were excluded. One sample from each pair of duplicated or related samples, as defined by genome-wide identity-by-descent score (PI_HAT) > 0.30 , was removed. Because of the SNP coverage differences among the Illumina HumanHap550 BeadChip, the Illumina Human610-Quad BeadChip, and the Illumina HumanOmniExpress-12 v1.0 chip, the quality control plots of the CNV metrics were reviewed separately for samples in the discovery cohort and those in the replication cohort. In the discovery cohort, samples with LRR SD > 0.3 , |GCWF| > 0.05 , or CNV count > 100 were excluded

from further analysis. In the replication cohort, samples with LRR SD > 0.21 , |GCWF| > 0.035 , or CNV count > 100 were removed.

CNV association statistical analysis

We assessed the appropriate genome-wide multiple testing correction for our study. We focused on the association between phenotype and rare CNVs. The CNV association software ParseCNV (20) was used to detect the association between CNV region (CNVR) and the food allergy phenotype. ParseCNV used a genome-wide segment-based scoring approach. As described previously (20), the Fisher exact test was used to compare the CNV frequency among cases and controls at each SNP. Among the discovery set of 522,879 autosomal SNPs shared by the Illumina HumanHap550 BeadChip and the Illumina Human610-Quad BeadChip, 1232 (0.236%) showed deletion and 320 (0.0612%) showed duplication in at least two unrelated cases in the discovery cohort (frequency $\geq 0.560\%$). The reason why we selected the cutoff of two cases having a given CNV is that it is the minimal case frequency that can produce nominal statistical significance and reproducibility for a CNV in a given region when compared with controls using the Fisher exact test. Such SNP-based upfront exclusion is analogous to the inclusion threshold of 1% minor allele frequency in SNP-based genome-wide association studies (GWASs). Subsequently, SNPs were collapsed into CNVRs that constitute the genomic span of consecutive probes in proximity (< 1 MB) with comparable significance ($\pm 1 \log p$ value) in the Fisher exact test when comparing case and control status (20). In our study, SNPs were collapsed into 125 deletion and 32 duplication CNVRs, which means a total of 157 tests. Therefore, the genome-wide significant threshold is $p < 3.18 \times 10^{-4}$, considering multiple testing correction. It is similar to the conservative cutoff of $p = 5 \times 10^{-4}$ that we discussed previously (<http://parsecnv.sourceforge.net/#MultipleTestingCorrection>).

For logistic regression analysis, we first generated ped files for CNV status. In the deletion ped files, CNV genotype status was defined in the following way: 1 1 for CN = 0, 1 2 for CN = 1, and 2 2 for others. In the duplication ped file, CNV genotype status was defined as 1 1 for CN = 4, 1 2 for CN = 3, and 2 2 for others. This strategy of converting CNV status to "genotype" ped file was described in detail in our previous publication (20). It allows for logistic regression analysis including the first three principal components to adjust for population stratification.

Permutation tests

To test for batch effects, we randomized the cases and controls between the discovery and replication cohorts and conducted permutation tests. Because of the SNP coverage differences between platforms, intensity data from the discovery cohort and replication cohort cannot be combined directly prior to CNV calling. Instead, we coded the genotype of samples carrying the significant CNVR of *CTNNA3* as 1 2, because they are all heterozygous deletions, and genotype 2 2 for those normal diploids (CN = 2), as described above. Having converted the CNV calls to genotypes, we then combined all samples and conducted permutations by randomly shuffling case and control status. We similarly defined the *RBFOX1* "genotype" ped based on whether each sample in the European ancestry cohort carries the significant CNVRs. Finally, we conducted 10,000 permutations of the data and a Fisher exact test for each permutation. Empirical p values were derived from the permutation tests.

Meta-analysis

We conducted meta-analysis of the overlapping CNVR or overlapping genes using the sample size-based analytical strategy implemented in METAL software (21).

CNV validation

CNVs of interest were validated using TaqMan copy number probes (Life Technologies, Grand Island, NY), following the manufacturer's standard protocol.

Western blot analysis of *CTNNA3* deletion

EBV-immortalized cell lines were generated from PBMCs of one case carrying the *CTNNA3* CNVR and four samples without the deletion as controls. We followed the protocol as described (22). For Western blot analysis of *CTNNA3* expression, EBV cells were lysed with Nonidet P-40 lysis buffer (Invitrogen). Proteins were separated on 4–12% NuPAGE Bis-Tris gels in MOPS SDS running buffer and transferred overnight onto nitrocellulose membranes (Invitrogen). The membrane was blocked in 3% BSA and cut into two halves. The top half was incubated with mouse anti-*CTNNA3* mAb (Lifespan Biosciences), and the bottom half was incubated

Table I. Sample demographics

	Discovery Cohort		Replication Cohort	
	Cases	Controls	Cases	Controls
No. samples	357	3980	167	1573
Male (%)	67.5	48.5	63.5	51.0
Age (y; mean $\pm$ SD)	5.6 $\pm$ 4.4	10.2 $\pm$ 5.4	5.4 $\pm$ 3.9	12.6 $\pm$ 3.8
Ethnicity (n)				
European ancestry	222	2002	106	1414
African American	135	1978	61	159

with mouse anti- β -actin mAb (Santa Cruz Biotechnology). Subsequently, the membranes were washed, incubated with secondary Ab for 1 h, and washed again; bound Ab was detected with a WesternBright ECL chemiluminescence detection system (Advanta).

CD63 basophil test

Ficoll-isolated mononuclear cells (which also contain substantial amounts of basophils) were used as the primary source of cells. Cells were nucleofected with control small interfering RNA (siRNA) and *CTNNA3*-specific Trilencer-27 Human siRNA (OriGene) at 10 nM using an Amaxa nucleofector and program Y-001. Forty-eight hours after nucleofection, cells were incubated with CD63-allophycocyanin and CD203c-PE Ab and PMA (10 ng/ml) for 30 min. After incubation, cells were washed and analyzed by FACS for expression of CD63 and CD203c. *CTNNA3* knockdown (KD) was confirmed by immunoblot analysis. Lysates of 10^6 cells were separated by SDS-PAGE and evaluated for KD of *CTNNA3*. After transferring, the nitrocellulose membrane was cut into two halves. The top part was probed with mouse anti-*CTNNA3* mAb, and the bottom half was probed with mouse anti- β -actin mAb as a loading control.

Results

After quality control screening, the discovery cohort consisted of 357 cases with diagnosed food allergy and 3980 controls, and the replication cohort consisted of 167 cases and 1573 controls. The demographic features of the two cohorts are summarized in Table I. PCAs of cases and controls (Supplemental Fig. 1) indicate that they overlapped in terms of ethnicity based on SNP genotype data. The majority of the participants in our study are of European ancestry, with a smaller set of African Americans who are an admixed population of European ancestry and African ancestry. We first used the Fisher exact test to statistically analyze CNV association and then conducted logistic regression, including principal components, to correct for potential population stratification.

The CNV profile of cases and controls is presented in Table II. In the discovery cohort, 5234 deletions and 2042 duplications were found in the cases, with an average of 14.66 deletions and 5.72 duplications/individual. This was not significantly different from the number of CNVs detected in the control set, suggesting that the overall CNV burden was the same between the cases and controls. Similarly, in the replication cohort there was no significant

difference in the number of CNVs detected between cases and controls. The average length of deletions was similar between cases and controls in both the discovery cohort and the replication cohort. A lower average duplication size was observed in cases compared with controls in the discovery, but not the replication, cohort.

No large-scale unbiased genetic study has been reported in the food allergy research community, but there have been a few small-scale candidate gene studies; however, the results were not consistent with each other, as reviewed by Hong et al. (10). We examined the following candidate genes in our study: *HLA*, *CD14*, *FOXP3*, *STAT6*, *SPINK5*, *IL10*, and *IL13*. Taking a gene-based approach, we assessed whether the ratio of samples carrying CNVs within one or more exons in these genes was significantly different between cases and controls using the Fisher exact test implemented in ParseCNV (20). We did not observe any significant difference for any of these genes. Taking into account introns, *HLA-B* was the only gene for which the *p* value for rare CNVs of duplication was nominally significant in the replication cohort (*p* = 0.026) and marginally significant when the discovery cohort and replication cohort were combined (*p* = 0.063, two-sided). We observed two cases harboring duplications in the gene *HLA-B* at chr6:31300691-31304663 and three controls having duplications in this gene.

Next, we took a genome-wide segment-based scoring approach to identify CNVRs associated with food allergy. In our discovery cohort, we found 125 deletions and 32 duplications enriched in cases and 18 deletions and 10 duplications enriched in the controls.

We then examined which of the case-enriched CNVRs identified in our discovery cohort could also be identified in the independent replication cohort. Because the SNP coverage is different between the genotyping platforms used for the discovery cohort and for the replication cohort, and because of uncertainty about the CNVR boundaries, we expected that CNVRs that are commonly identified in both cohorts may only have partial overlap. We found five significant CNVRs in the discovery cohort that overlapped the CNVRs observed in the replication cohort (Table III). Meta-analysis showed that three of them had *p* values surpassing the

Table II. CNV profiles of cases and controls

	Discovery Cohort		Replication Cohort	
	Cases (n = 357)	Controls (n = 3,980)	Cases (n = 167)	Controls (n = 1,573)
Total no. CNVs	7,276	85,077	6,005	54,848
Deletions	5,234	59,044	2,359	28,567
Duplications	2,042	26,033	3,646	26,281
Average no. CNVs/subject	20.38	21.38	35.96	34.87
Deletions	14.66	14.84	14.13	18.16
Duplications	5.72	6.54	21.83	16.71
Average size of CNV (kb)	58.65	62.98	62.60	65.81
Deletions	50.56	45.02	69.28	69.86
Duplications	79.39	103.69	58.27	61.40

Table III. Overlapping significant CNVRs between discovery cohort and replication cohort

Overlapping Region (hg19)	Closest Gene	Discovery Cohort				Replication Cohort				Meta-Analysis					
		CNV	<i>p</i> Value	PCA-Corrected <i>p</i> Value	Cases (<i>n</i> [%])	Controls (<i>n</i> [%])	Type	CNV	<i>p</i> Value	PCA-Corrected <i>p</i> Value	Cases (<i>n</i> [%])	Controls (<i>n</i> [%])	Type	<i>p</i> Value	PCA-Corrected <i>p</i> Value
chr1:1246972-1264538	<i>CPSF3L</i> , <i>GLTPDI</i> , <i>PUSL1</i>	chr1:1176597-2450394	4.08×10^{-3}	3.66×10^{-4}	4 (1.12)	5 (0.13)	Del	chr1:1246972-1264538	0.0293	6.67×10^{-3}	4 (2.40)	9 (0.57)	Dup	3.27×10^{-4}	8.11×10^{-6}
chr4:172367540-172816200	<i>GALNTL6</i>	chr4:172367540-172816200	0.0191	0.0115	2 (0.56)	1 (0.03)	Del	chr4:171663744-172946287	3.23×10^{-3}	0.0106	3 (1.80)	1 (0.06)	Dup	3.62×10^{-4}	4.61×10^{-4}
chr4:189818663-189831920	<i>LOC401164</i>	chr4:189800486-190522201	0.0192	0.0126	2 (0.56)	1 (0.03)	Del	chr4:189818663-189831920	0.0257	6.52×10^{-3}	2 (1.20)	1 (0.06)	Dup	1.52×10^{-3}	3.66×10^{-4}
chr10:68282970-68284017	<i>CTNNA3</i>	chr10:68282970-68284017	0.0363	0.0184	2 (0.56)	2 (0.05)	Del	chr10:67318448-68879250	7.17×10^{-3}	0.0353	5 (2.99)	9 (0.57)	Dup	1.34×10^{-3}	1.82×10^{-3}
chr13:96550754-96558852	<i>UGCGL2</i>	chr13:96537274-96638636	0.0192	9.39×10^{-3}	2 (0.56)	1 (0.03)	Del	chr13:96550754-96558852	3.71×10^{-5}	1.41×10^{-6}	9 (5.39)	11 (0.70)	Dup	2.84×10^{-5}	1.79×10^{-6}

Del, deletion; Dup, duplication; PCA-corrected *p* value, *p* value from logistic regression, including the first three principal components as covariates.

Table IV. Overlapping genes between discovery cohort and replication cohort

Gene	Discovery Cohort					Replication Cohort					Meta-Analysis			
	CNV	Cases (<i>n</i> [%])	Controls (<i>n</i> [%])	<i>p</i> Value	PCA- Corrected <i>p</i> Value	Type	CNV	Cases <i>n</i> [%])	Controls <i>n</i> [%])	<i>p</i> Value	PCA- Corrected <i>p</i> Value	Type	<i>p</i> Value	PCA- Corrected <i>p</i> Value
<i>ODZ3</i>	chr4:183271349-183291465	2 (0.56)	1 (0.03)	0.0192	0.0116	Del	chr4:183559306-183565618	3 (1.80)	1 (0.06)	3.23×10^{-3}	0.0180	Dup	3.65×10^{-4}	6.79×10^{-4}
<i>CTNNA3</i>	chr10:68282970-68284017	2 (0.56)	2 (0.05)	0.0363	0.0184	Del	chr10:68383827-68407077	2 (1.20)	1 (0.06)	0.0257	0.0206	Del	2.99×10^{-3}	1.24×10^{-3}
<i>LUZP2</i>	chr11:24778961-24783183	2 (0.56)	2 (0.05)	0.0363	0.0226	Dup	chr11:24412621-24551109	2 (1.20)	1 (0.06)	0.0257	0.0153	Dup	2.99×10^{-3}	1.26×10^{-3}
<i>RBFOX1</i>	chr16:7126629-7196046	3 (0.84)	2 (0.05)	4.88×10^{-3}	4.72×10^{-3}	Del	chr16:6763216-6801846	2 (1.20)	0 (0.00)	9.16×10^{-3}	0.9989	Del	1.58×10^{-4}	0.0170
<i>MACROD2</i>	chr20:15104193-15126507	4 (1.12)	5 (0.13)	4.07×10^{-3}	3.37×10^{-3}	Del	chr20:14713890-14727386	3 (1.80)	3 (0.19)	0.0140	1.41×10^{-3}	Del	1.80×10^{-4}	2.85×10^{-5}

Del, deletion; Dup, duplication; PCA-corrected *p* value, *p* value from logistic regression, including the first three principal components as covariates.

genome-wide significant cutoff of $p < 5 \times 10^{-4}$. However, each of these five regions was a deletion CNV in the discovery cohort but a duplication CNV in the replication cohort.

We next investigated whether any genes had significant CNVRs both in the discovery cohort and in the replication cohort and found five such genes (Table IV). Among them, three genes reached genome-wide significance ($p < 5 \times 10^{-4}$) in the meta-analysis. The top one is *RBFOX1*, which overlapped with a significant CNVR in three cases and two controls in the discovery cohort (Table IV), as well as a significant CNVR in two cases and no controls in the replication cohort (Table IV). Meta-analysis yielded a p value of 1.58×10^{-4} . Another gene of interest is *CTNNA3*, because exon 11 of *CTNNA3* overlapped with the significant CNVR identified in 2 of 357 cases and 2 of 3980 controls in the discovery cohort (Fig. 1A, Table IV). CNVs in exons are more likely to have detrimental effects on gene products. Additionally, there is a clear enrichment of enhancer- and promoter-associated histone mark (H3K4Me1) in *CTNNA3* CNVRs in the replication cohort, which is also conserved, as shown in the UCSC genome browser (Fig. 1B), suggesting a loss of potential transcriptional regulation due to the deletion. This CNVR chr10:68383827-68407077 was found in 2 of 167 cases and 1 of 1573 controls (Table IV). The *CTNNA3* gene encodes a cadherin-associated protein. Conducting genomic annotation using the HaploReg (23) software package, we found that each of the above CNVRs harbors multiple SNPs overlapping with promoter/enhancer histone marks, protein binding sites, and transcription factor motifs (Supplemental Table I), suggesting potentially altered transcriptional regulation of *CTNNA3* and *RBFOX1* resulting from the CNVs in these regions.

To check whether any of the significant CNVR associations were confounded by population stratification, we conducted PCA and

performed logistic regression on the CNV status, including the first three principal components as covariates. All of the associations in Tables III and IV remained significant in both the discovery cohort and the replication cohort, with the exception of *RBFOX1* in the replication cohort (Tables III, IV). To tease out how population stratification affects the association of *RBFOX1*, we further ex-

motifs (Supplemental Table I), suggesting potentially altered transcriptional regulation of *CTNNA3* and *RBFOX1* resulting from the CNVs in these regions.

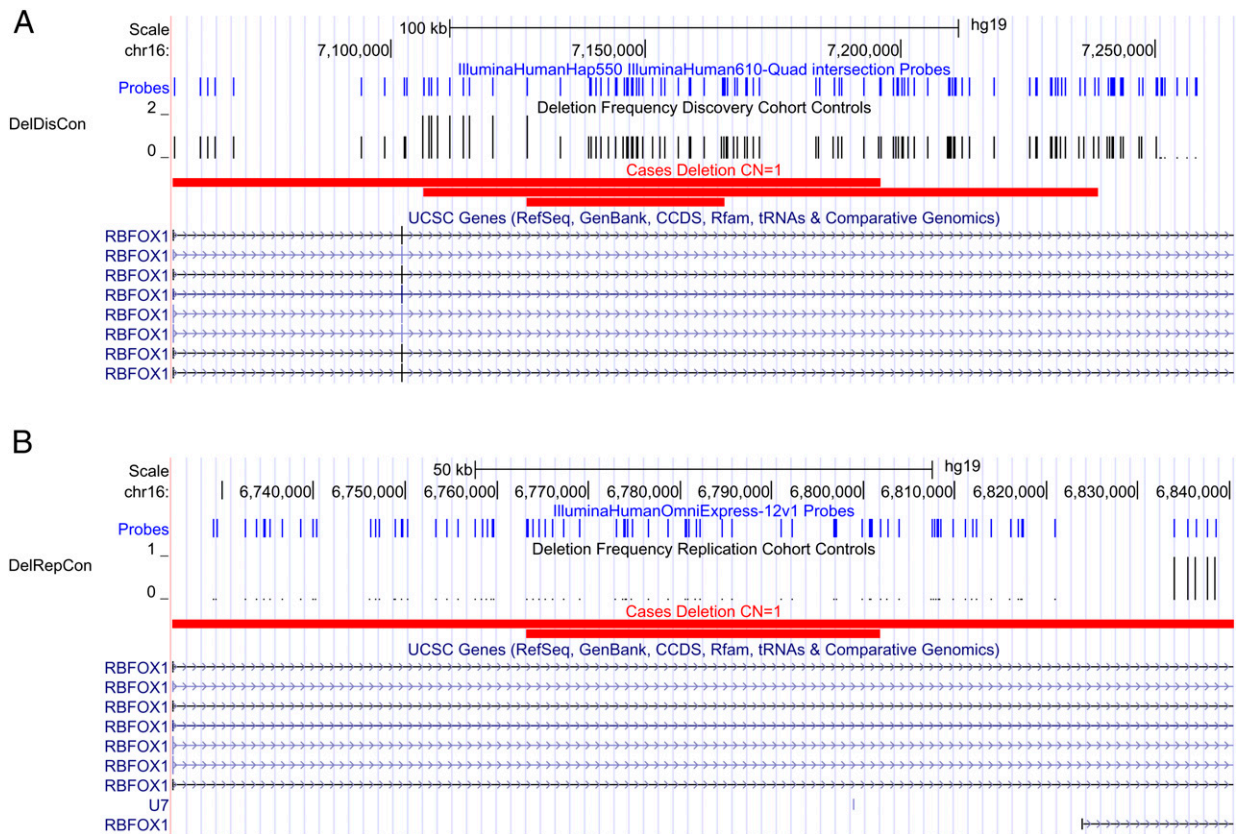


FIGURE 2. Deletions in gene *RBFOX1*. Case-enriched deletions in gene *RBFOX1* were detected in the European ancestry discovery cohort (A) and replication cohort (B). CCDS, Consensus Coding Sequence Project; DelDisCon, deletion frequency discovery cohort control; DelRepCon, deletion frequency replication cohort control; RefSeq, reference sequence database of the National Center for Biotechnology Information; Rfam, database containing information about non-coding RNA families and other structured RNA elements; tRNA, transfer RNA; Vert. Cons, vertebrate conservation.

Deletions spanning an exonic sequence can directly affect transcription and translation of gene products. The *CTNNA3* exonic deletion resulted in a frameshift mutation that is predicted to result in premature termination of transcription and subsequent protein degradation by the introduction of a downstream stop codon. We then examined the *CTNNA3* protein level in EBV cell lines established from one case sample carrying the exonic deletion CNVR (Fig. 1A) and four other samples without CNV at this region. We observed a reduced *CTNNA3* protein level in the deletion case sample on Western blot (Fig. 4).

To further investigate the functional significance and biological role of *CTNNA3* in the pathogenesis of food allergy, we conducted a CD63 basophil test in *CTNNA3*-KD mononuclear cells compared with controls. Changes in monocyte activation during the assay are reflective of IgE-dependent responses to allergen challenge. In our experiment, we used siRNA to mediate KD of *CTNNA3*, which resulted in a 70% reduction in *CTNNA3* protein level, as confirmed by Western blot (Fig. 5B). We then measured upregulation of CD63 and CD203c on cells upon PMA stimulation by FACS. Tetraspan Ag CD63 and CD203c are surface markers expressed on various myeloid cells, including basophils and monocytes, and they show an increased expression following exposure to allergen or anti-IgE Ab. In our experiment, both Ags were upregulated after activation with PMA (Fig. 5A). The upregulation was more pronounced in *CTNNA3*-KD cells in comparison with control siRNA cells activated with PMA. Therefore, we conclude that reduced *CTNNA3* expression leads to PMA-induced upregulation of CD63 and CD203C, suggesting increased sensitization to allergens.

Genes function collaboratively through genetic and physical interactions to contribute to disease etiology. Using Database for Annotation, Visualization, and Integrated Discovery (24), we tested whether the genes that are close to or harboring our significant CNVRs are enriched for certain functional categories. We used the genes that were close to significant overlapping CNVRs in Table III and those significant overlapping genes in Table IV as input for the analysis. The results showed that 7 of 10 input genes (*CPSF3L*, *RBFOX1*, *PUSL1*, *GALNTL6*, *MACROD2*, *LUZP2*, *CTNNA3*) fell into the functional categories of alternative splicing and splice variant, with nominal *p* values of 0.0035 and 0.0035, respectively. This is consistent with prevalent alternative splicing in the immune system (25, 26).

Discussion

Genome-wide CNV analysis is an emerging method of assessing the genetic underpinnings of clinical disease that eliminates preconceptions of any particular candidate genes. Using this method, we identified genes of interest in food allergy.

HLA-B is a class I MHC molecule that displays intracellular peptides to T cells. Although not previously identified as a gene of interest in food allergy, *HLA-B* differences are seen in drug hypersensitivity reactions (27) and chronic idiopathic urticaria (28). Moreover, region 6p21.33, in which *HLA-B* resides, is a susceptibility locus for subjects with environmental allergy (29).

CTNNA3 encodes a structural cadherin important for cell-cell adhesion (30). It is transcribed using multiple promoters, which produce alternate transcripts (31). The importance of *CTNNA3* in maintaining cell-cell adhesion may play a role in allergy.

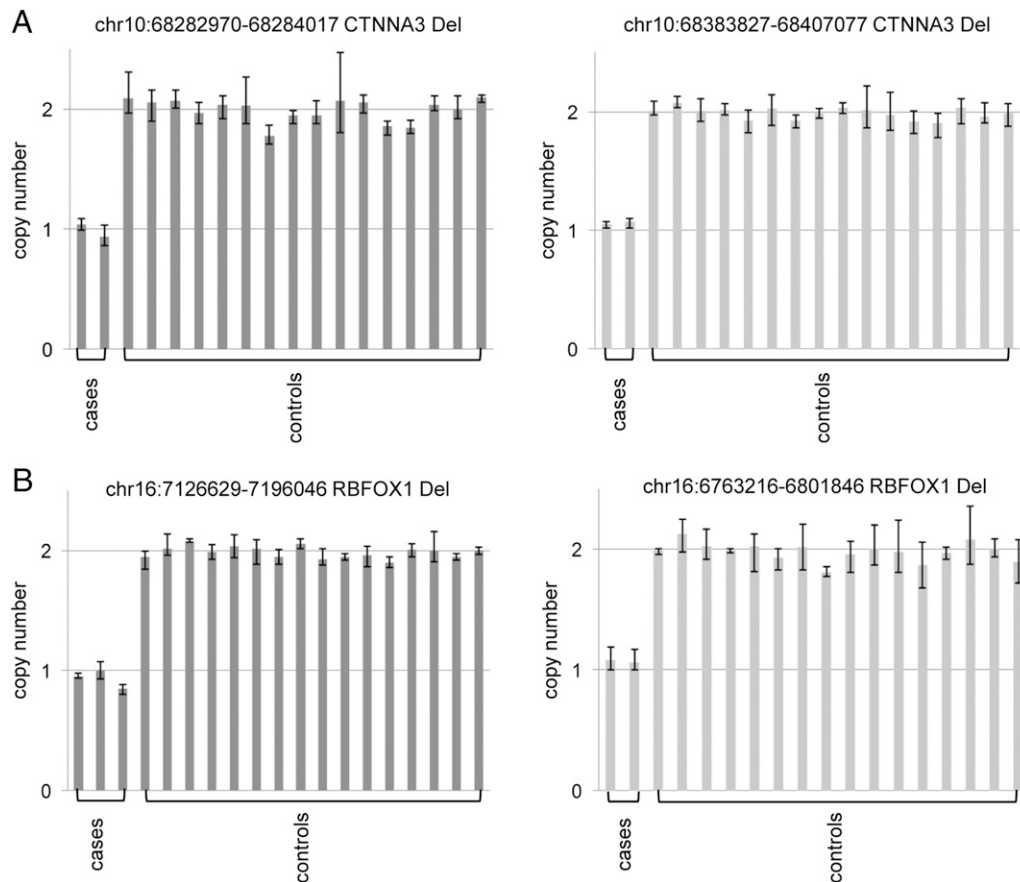


FIGURE 3. qPCR validation of CNV events detected by the Illumina genotyping array. Results represent the copy number calculated from triplicate runs by CopyCaller software. The minimal and maximal copy number calculated for each sample are shown with the copy number range bars. Cases are case samples carrying the corresponding CNV. Controls are randomly selected samples from each cohort without the corresponding CNV detected on array. Del, deletion.

Impaired skin and airway barrier is associated with atopic disease susceptibility (reviewed in Ref. 32). Electron microscope studies demonstrated impairment of this barrier in asthmatics compared with controls (33, 34). In addition, reduced CTNNA3 levels were found in the bronchial biopsies of asthmatics, and they correlated inversely with eosinophil numbers (35). Moreover, CTNNA3 SNPs were identified by a GWAS in Korean subjects with diisocyanate-induced occupational asthma (36). Two of the CTNNA3 SNPs identified in the Korean study were subsequently confirmed in a study of Canadian workers with occupational asthma (37). A recent GWAS demonstrated the association between SNP in CTNNA3 and response to glucocorticoid therapy in childhood asthma (38).

The RFXO1–ataxin-2 complex is considered important for RNA transport in neurons; furthermore, alterations lead to neuronal cytotoxicity (39). Both human and animal model studies demonstrated the importance of RFXO1 in neuronal development and regulation (40, 41). Although the connection between neurologic stimulation and atopic disease remains to be elucidated, it is noteworthy that stress can exacerbate atopic dermatitis via substance P-dependent neurogenic inflammation (42, 43).

Common variants have long been considered the major genetic determinants of common disease, playing a key role in disease etiology. However, as GWASs showed, even where large numbers of susceptibility loci have been identified, >100 genome-wide significant loci in the case of inflammatory bowel disease, the heritability explained by these loci is modest (44). Accumulating evidence suggests an important contribution of rare variants, including rare CNVs, to the etiology of complex disease. Significant associations with CNVs were reported across multiple diseases,

such as sporadic amyotrophic lateral sclerosis (45), autism (46), obesity (47), congenital heart disease (48), and endometriosis (49). Riggs et al. (50) recently reviewed the impact of CNVs on autism. They concluded that, although CNVs may be individually rare, each accounting for no more than 1% of cases, when combined the overall CNV burden explains a significant proportion of disease variance, estimated to be between 5 and 10%. These rare CNVs usually present a higher risk than most of the common variants. Therefore, it is important to identify and further characterize these rare structural variants to fully understand their clinical significance.

To the best of our knowledge, our analysis is the first to investigate CNVs associated with food allergy on a genome-wide scale. We used a hypothesis-free, unbiased approach. Genes identified in the discovery cohort were validated in an independent cohort genotyped on a different platform, strengthening the validity

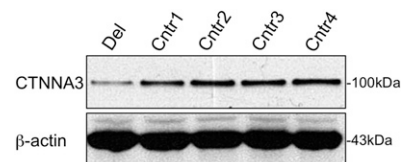
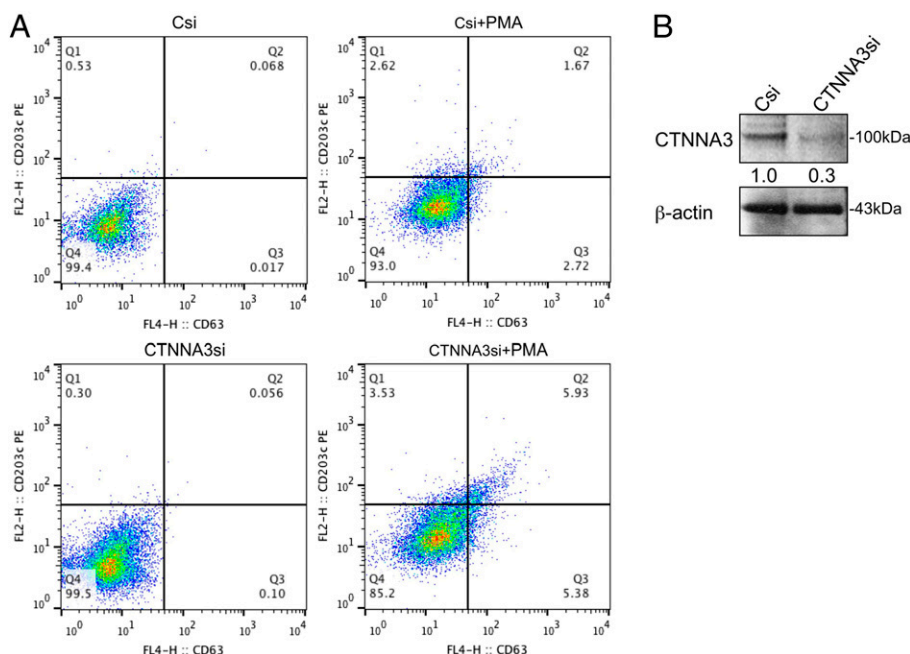


FIGURE 4. Western blot analysis of CTNNA3 deletion. Western blot analysis of CTNNA3 in EBV cell lines from one case sample carrying the CTNNA3 exonic deletion (lane 1) and four other subjects in our cohort without the CNV (copy number = 2) as controls (Cntr, lanes 2–5). Whole-cell lysates were separated and transferred, and membranes were probed with anti-CTNNA3 and anti- β -actin mouse mAb. Representative results of three independent experiments are shown. Del, deletion.

FIGURE 5. Comparison of PMA-induced monocyte activation in *CTNNA3*-KD cells and controls. **(A)** Monocytes were nucleofected with control siRNA (Csi) and *CTNNA3*-specific siRNA (*CTNNA3*si). Forty-eight hours after nucleofection, cells were incubated with CD63-allophycocyanin and CD203c-PE Ab and PMA (10 ng/ml) for 30 min. Expression of CD63 and monocyte marker CD203c was measured by FACS. Both Ags were upregulated after activation with PMA. The upregulation was more pronounced in *CTNNA3*-KD cells in comparison with control following PMA activation. **(B)** *CTNNA3* KD was validated by Western blot using mouse monoclonal *CTNNA3* Ab; β -actin was used as a loading control. The numbers represent densitometric analysis of *CTNNA3* expression in relation to β -actin.



of our discoveries. In addition, we showed the siRNA KD of *CTNNA3* results in increased degranulation, further demonstrating the biological importance of *CTNNA3* in allergy.

Although these results are important, there are certain limitations to our study. First, the number of cases analyzed was relatively small for a genome-wide study. Food allergy is a heterogeneous condition that requires stringent phenotyping criteria to reduce misdiagnosis. As a result, 47% of potential cases were excluded from our analysis following chart review. However, we included a relatively large number of controls, leading to a high control/case ratio of 10, which increases statistical power when the effect of exposure is large and when the prevalence of exposure among controls is low (51–53). In this study, the exposure is the presence of the CNV. Rare variants, including CNVs, usually have a greater effect size than do common variants, and the population frequency of rare variants is very low. Because of the moderate to large effect size of rare CNVs and the high control/case ratio, we were able to detect multiple significant associations in our cohort. However, we would miss the association of CNVs with a lower effect size (e.g., a type II error). With a larger cohort, more CNV areas of interest may have been identified. Given that we were able to demonstrate significant CNV frequency differences between cases and controls despite our sample size, we surmise that the *p* value attained in a larger sample size would be smaller. A second limitation is that we were unable to generate a sufficiently large cohort of food monoallergic patients to perform subgroup analyses. Third, false associations can be seen in CNV analysis, as discussed by Craddock et al. (54). In this regard, we note that the CNVs that we reported to be of clinical significance were experimentally validated. Finally, CNV association analysis is affected by choices of normalization and probe weighting, and optimal choice can vary (54).

In summary, we described a novel approach to assess for genetic susceptibility to food allergy. By working backward from genotype to phenotype, we identified two intriguing CNV variants that may play a role in food allergy. These variants are of unclear significance, and more functional studies are required to establish their mechanism of action. However, we also envision that more advanced computational algorithms may aid in determining their pathogenicity and identify other loci and gene networks of clinical interest.

Acknowledgments

We thank the patients and their families who participated in this study. We thank the Center for Applied Genomics recruitment and wet laboratory staff for their invaluable contributions. We also thank Megan Ott for help with data collection, as well as all participants of this genetics study.

Disclosures

The authors have no financial conflicts of interest.

References

- Gupta, R. S., E. E. Springston, M. R. Warrier, B. Smith, R. Kumar, J. Pongracic, and J. L. Holl. 2011. The prevalence, severity, and distribution of childhood food allergy in the United States. *Pediatrics* 128: e9–e17.
- Branum, A. M., and S. L. Lukacs. 2008. Food allergy among U.S. children: trends in prevalence and hospitalizations. *NCHS Data Brief* (10): 1–8.
- Rinaldi, M., L. Harnack, C. Oberg, P. Schreiner, J. St Sauver, and L. L. Travis. 2012. Peanut allergy diagnoses among children residing in Olmsted County, Minnesota. *J. Allergy Clin. Immunol.* 130: 945–950.
- Sicherer, S. H., A. Muñoz-Furlong, J. H. Godbold, and H. A. Sampson. 2010. US prevalence of self-reported peanut, tree nut, and sesame allergy: 11-year follow-up. *J. Allergy Clin. Immunol.* 125: 1322–1326.
- Almal, S. H., and H. Padh. 2012. Implications of gene copy-number variation in health and diseases. *J. Hum. Genet.* 57: 6–13.
- Houriha, J. O., T. P. Dean, and J. O. Warner. 1996. Peanut allergy in relation to heredity, maternal diet, and other atopic diseases: results of a questionnaire survey, skin prick testing, and food challenges. *BMJ* 313: 518–521.
- Sicherer, S. H., T. J. Furlong, H. H. Maes, R. J. Desnick, H. A. Sampson, and B. D. Gelb. 2000. Genetics of peanut allergy: a twin study. *J. Allergy Clin. Immunol.* 106: 53–56.
- Webb, B. T., E. van den Oord, A. Akkari, S. Wilton, T. Ly, R. Duff, K. C. Barnes, K. Carlsen, J. Gerritsen, W. Lenney, et al. 2007. Quantitative linkage genome scan for atopy in a large collection of Caucasian families. *Hum. Genet.* 121: 83–92.
- Venanzi, S., G. Malerba, R. Galavotti, M. C. Lauciello, E. Trabetti, G. Zanoni, L. Pescolliderung, L. C. Martinati, A. L. Boner, and P. F. Pignatti. 2001. Linkage to atopy on chromosome 19 in north-eastern Italian families with allergic asthma. *Clin. Exp. Allergy* 31: 1220–1224.
- Hong, X., H. J. Tsai, and X. Wang. 2009. Genetics of food allergy. *Curr. Opin. Pediatr.* 21: 770–776.
- Redon, R., S. Ishikawa, K. R. Fitch, L. Feuk, G. H. Perry, T. D. Andrews, H. Fiegler, M. H. Shaper, A. R. Carson, W. Chen, et al. 2006. Global variation in copy number in the human genome. *Nature* 444: 444–454.
- Gu, W., F. Zhang, and J. R. Lupski. 2008. Mechanisms for human genomic rearrangements. *Pathogenetics* 1: 4.
- Shianna, K. V., and H. F. Willard. 2006. Human genomics: in search of normality. *Nature* 444: 428–429.
- Hurles, M. E., E. T. Dermitzakis, and C. Tyler-Smith. 2008. The functional impact of structural variation in humans. *Trends Genet.* 24: 238–245.

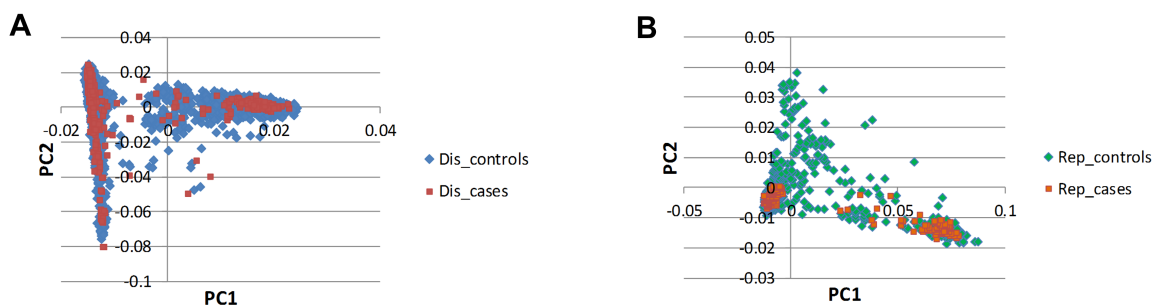
15. Coughlin, C. R., II, G. H. Scharer, and T. H. Shaikh. 2012. Clinical impact of copy number variation analysis using high-resolution microarray technologies: advantages, limitations and concerns. *Genome Med* 4: 80.
16. Brown, S. J., K. Kroboth, A. Sandilands, L. E. Campbell, E. Pohler, S. Kezic, H. J. Cordell, W. H. McLean, and A. D. Irvine. 2012. Intragenic copy number variation within filaggrin contributes to the risk of atopic dermatitis with a dose-dependent effect. *J. Invest. Dermatol.* 132: 98–104.
17. Sampson, H. A., A. Muñoz-Furlong, R. L. Campbell, N. F. Adkinson, Jr., S. A. Bock, A. Branum, S. G. Brown, C. A. Camargo, Jr., R. Cydulka, S. J. Galli, et al. 2006. Second symposium on the definition and management of anaphylaxis: summary report—second National Institute of Allergy and Infectious Disease/Food Allergy and Anaphylaxis Network symposium. *Ann. Emerg. Med.* 47: 373–380.
18. Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38: 904–909.
19. Wang, K., M. Li, D. Hadley, R. Liu, J. Glessner, S. F. Grant, H. Hakonarson, and M. Bucan. 2007. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* 17: 1665–1674.
20. Glessner, J. T., J. Li, and H. Hakonarson. 2013. ParseCNV integrative copy number variation association software with quality tracking. *Nucleic Acids Res.* 41: e64.
21. Willer, C. J., Y. Li, and G. R. Abecasis. 2010. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26: 2190–2191.
22. Tosato, G., and J. I. Cohen. 2007. Generation of Epstein-Barr Virus (EBV)-immortalized B cell lines. *Curr. Protoc. Immunol.* Chapter 7: Unit 7.22.
23. Ward, L. D., and M. Kellis. 2012. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* 40: D930–D934.
24. Huang, W., B. T. Sherman, and R. A. Lempicki. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4: 44–57.
25. Evsyukova, I., J. A. Somarelli, S. G. Gregory, and M. A. Garcia-Blanco. 2010. Alternative splicing in multiple sclerosis and other autoimmune diseases. *RNA Biol.* 7: 462–473.
26. Rodrigues, R., A. R. Grosso, and L. Moita. 2013. Genome-wide analysis of alternative splicing during dendritic cell response to a bacterial challenge. *PLoS ONE* 8: e61975.
27. Karlin, E., and E. Phillips. 2014. Genotyping for severe drug hypersensitivity. *Curr. Allergy Asthma Rep.* 14: 418.
28. Bozek, A., J. Krajewska, B. Filipowska, J. Polanska, R. Rachowska, A. Grzanka, and J. Jarzab. 2010. HLA status in patients with chronic spontaneous urticaria. *Int. Arch. Allergy Immunol.* 153: 419–423.
29. Hinds, D. A., G. McMahon, A. K. Kiefer, C. B. Do, N. Eriksson, D. M. Evans, B. St Pourcain, S. M. Ring, J. L. Mountain, U. Francke, et al. 2013. A genome-wide association meta-analysis of self-reported allergy identifies shared and allergy-specific susceptibility loci. *Nat. Genet.* 45: 907–911.
30. Janssens, B., S. Goossens, K. Staes, B. Gilbert, J. van Hengel, C. Colpaert, E. Bruyneel, M. Mareel, and F. van Roy. 2001. alphaT-catenin: a novel tissue-specific beta-catenin-binding protein mediating strong cell-cell adhesion. *J. Cell Sci.* 114: 3177–3188.
31. Smith, J. D., M. H. Meehan, J. Crean, and A. McCann. 2011. Alpha T-catenin (CTNNA3): a gene in the hand is worth two in the nest. *Cell. Mol. Life Sci.* 68: 2493–2498.
32. Takai, T., and S. Ikeda. 2011. Barrier dysfunction caused by environmental proteases in the pathogenesis of allergic diseases. *Allergol. Int.* 60: 25–35.
33. Jeffery, P. K., A. J. Wardlaw, F. C. Nelson, J. V. Collins, and A. B. Kay. 1989. Bronchial biopsies in asthma. An ultrastructural, quantitative study and correlation with hyperreactivity. *Am. Rev. Respir. Dis.* 140: 1745–1753.
34. Montefort, S., J. A. Roberts, R. Beasley, S. T. Holgate, and W. R. Roche. 1992. The site of disruption of the bronchial epithelium in asthmatic and non-asthmatic subjects. *Thorax* 47: 499–503.
35. de Boer, W. I., H. S. Sharma, S. M. Baelemans, H. C. Hoogsteden, B. N. Lambrecht, and G. J. Braunstahl. 2008. Altered expression of epithelial junctional proteins in atopic asthma: possible role in inflammation. *Can. J. Physiol. Pharmacol.* 86: 105–112.
36. Kim, S. H., B. Y. Cho, C. S. Park, E. S. Shin, E. Y. Cho, E. M. Yang, C. W. Kim, C. S. Hong, J. E. Lee, and H. S. Park. 2009. Alpha-T-catenin (CTNNA3) gene was identified as a risk variant for toluene diisocyanate-induced asthma by genome-wide association analysis. *Clin. Exp. Allergy* 39: 203–212.
37. Bernstein, D. I., M. Kashon, Z. L. Lummus, V. J. Johnson, K. Fluharty, D. Gaurin, J. L. Malo, A. Cartier, L. P. Boulet, J. Sastre, et al. 2013. CTNNA3 (α -catenin) gene variants are associated with diisocyanate asthma: a replication study in a Caucasian worker population. *Toxicol. Sci.* 131: 242–246.
38. Perin, P., and U. Potočník. 2014. Polymorphisms in recent GWA identified asthma genes CA10, SGK493, and CTNNA3 are associated with disease severity and treatment response in childhood asthma. *Immunogenetics* 66: 143–151.
39. Shibata, H., D. P. Huynh, and S. M. Pulst. 2000. A novel protein with RNA-binding motifs interacts with ataxin-2. *Hum. Mol. Genet.* 9: 1303–1313.
40. Fogel, B. L., E. Wexler, A. Wahnich, T. Friedrich, C. Vijayendran, F. Gao, N. Parikhshak, G. Konopka, and D. H. Geschwind. 2012. RBFOX1 regulates both splicing and transcriptional networks in human neuronal development. *Hum. Mol. Genet.* 21: 4171–4186.
41. Gehman, L. T., P. Stoilov, J. Maguire, A. Damianov, C. H. Lin, L. Shue, M. Ares, Jr., I. Mody, and D. L. Black. 2011. The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. *Nat. Genet.* 43: 706–711.
42. Pavlovic, S., M. Danilchenko, D. J. Tobin, E. Hagen, S. P. Hunt, B. F. Klapp, P. C. Arck, and E. M. Peters. 2008. Further exploring the brain-skin connection: stress worsens dermatitis via substance P-dependent neurogenic inflammation in mice. *J. Invest. Dermatol.* 128: 434–446.
43. Huang, C. H., I. C. Kuo, H. Xu, Y. S. Lee, and K. Y. Chua. 2003. Mite allergen induces allergic dermatitis with concomitant neurogenic inflammation in mouse. *J. Invest. Dermatol.* 121: 289–293.
44. Jostins, L., S. Ripke, R. K. Weersma, R. H. Duerr, D. P. McGovern, K. Y. Hui, J. C. Lee, L. P. Schumm, Y. Sharma, C. A. Anderson, et al; International IBD Genetics Consortium (IBDGC). 2012. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 491: 119–124.
45. Blauw, H. M., J. H. Veldink, M. A. van Es, P. W. van Vught, C. G. Saris, B. van der Zwaag, L. Franke, J. P. Burbach, J. H. Wokke, R. A. Ophoff, and L. H. van den Berg. 2008. Copy-number variation in sporadic amyotrophic lateral sclerosis: a genome-wide screen. *Lancet Neurol.* 7: 319–326.
46. Sebat, J., B. Lakshmi, D. Malhotra, J. Troge, C. Lese-Martin, T. Walsh, B. Yamrom, S. Yoon, A. Krasnitz, J. Kendall, et al. 2007. Strong association of de novo copy number mutations with autism. *Science* 316: 445–449.
47. Wang, K., W. D. Li, J. T. Glessner, S. F. Grant, H. Hakonarson, and R. A. Price. 2010. Large copy-number variations are enriched in cases with moderate to extreme obesity. *Diabetes* 59: 2690–2694.
48. Glessner, J. T., A. G. Bick, K. Ito, J. G. Homsy, L. Rodriguez-Murillo, M. Fromer, E. Mazaika, B. Vardarajan, M. Italia, J. Leipzig, et al. 2014. Increased frequency of de novo copy number variants in congenital heart disease by integrative analysis of single nucleotide polymorphism array and exome sequence data. *Circ. Res.* 115: 884–896.
49. Chettier, R., K. Ward, and H. M. Albersen. 2014. Endometriosis is associated with rare copy number variants. *PLoS ONE* 9: e103968.
50. Riggs, E. R., D. H. Ledbetter, and C. L. Martin. 2014. Genomic Variation: Lessons Learned from Whole-Genome CNV Analysis. *Curr. Genet. Med. Rep.* 2: 146–150.
51. Breslow, N. E., J. H. Lubin, P. Marek, and B. Langholz. 1983. Multiplicative models and cohort analysis. *J. Am. Stat. Assoc.* 78: 1–12.
52. Wacholder, S., D. T. Silverman, J. K. McLaughlin, and J. S. Mandel. 1992. Selection of controls in case-control studies. III. Design options. *Am. J. Epidemiol.* 135: 1042–1050.
53. Hennessy, S., W. B. Bilker, J. A. Berlin, and B. L. Strom. 1999. Factors influencing the optimal control-to-case ratio in matched case-control studies. *Am. J. Epidemiol.* 149: 195–197.
54. Craddock, N., M. E. Hurles, N. Cardin, R. D. Pearson, V. Plagnol, S. Robson, D. Vukcevic, C. Barnes, D. F. Conrad, E. Giannoulidou, et al; Wellcome Trust Case Control Consortium. 2010. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* 464: 713–720.

Copy number variations in CTNNA3 and RBFOX1 associate with pediatric food allergy

Jin Li, Irene Fung, Joseph T Glessner, Zhi Wei, Frank D Mentch, Renata Pellegrino,
Tiancheng Wang, Cecilia Kim, Cuiping Hou, Fengxiang Wang, Rosetta M Chiavacci, Kelly
A Thomas, Rahul Pandey, Marina Bakay, Jonathan M Spergel, Hakon Hakonarson,
Patrick MA Sleiman

Supplemental Data

Supplemental Figure 1. Principal components analysis of cases and controls in the discovery cohort (A) and the replication cohort (B).



Supplemental Table I: CNVRs in *CTNNA3* and *RBFOX1* genes contain SNPs overlapping with regulatory motifs. By Haploreg(1) search, we found CNVRs in *CTNNA3* and *RBFOX1* genes contain SNPs that are conserved, overlap with various regulatory motifs, or have proteins bound.

region	gene	Number of SNPs that overlap with					
		Conserved sites	Promoter histone markers	Enhancer histone markers	DNase	Proteins bound	Regulatory Motifs
chr10:68282970-68284017	<i>CTNNA3</i>	6					15
chr10:68383827-68407077	<i>CTNNA3</i>	40		163	67	23	357
chr16:7126629-7196046	<i>RBFOX1</i>	221	65	691	107	96	1861
chr16:6763216-6801846	<i>RBFOX1</i>	98	47	311	75	33	1082

Conserved sites=conserved sites by GERP (Genomic Evolutionary Rate Profiling)(2, 3) and/or SiPhy-omega (Site-specific PHYlogenetic analysis)(4) algorithm; Promoter histone markers= Promoter histone markers by ENCODE(5) and/or Roadmap(6); Enhancer histone markers= Enhancer histone markers by ENCODE(5) and/or Roadmap(6); DNase=DNase hypersensitive sites; Proteins bound=Proteins bound by ChIP (ENCODE(5)).

Supplemental Table II. Statistical results of CNVR in gene *RBFOX1* in the Caucasian cohort.

	Discovery cohort					Replication cohort					Meta-analysis
Gene	CNVR	Cases N (frequency)	Control N (frequency)	P-value	Type	CNVR	Cases N (frequency)	Controls N (frequency)	P-value	Type	P-value
<i>RBFOX1</i>	chr16:7126629-7196046	3 (1.35%)	1 (0.05%)	3.64×10^{-3}	Del	chr16:6763216-6801846	2 (1.89%)	0 (0.00%)	4.82×10^{-3}	Del	7.35×10^{-5}

Del=Deletion

References:

1. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* 2012;**40**(Database issue):D930-934.
2. Cooper GM, Stone EA, Asimenos G, Program NCS, Green ED, Batzoglou S, et al. Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res* 2005;**15**(7):901-913.
3. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol* 2010;**6**(12):e1001025.
4. Garber M, Guttman M, Clamp M, Zody MC, Friedman N, Xie X. Identifying novel constrained elements by exploiting biased substitution patterns. *Bioinformatics* 2009;**25**(12):i54-62.
5. ENCODE Project Consortium, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;**489**(7414):57-74.
6. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* 2010;**28**(10):1045-1048.