

# Epidermal differentiation complex genetic variation in atopic dermatitis and peanut allergy



Michelle F. Huffaker, MD,<sup>a,\*</sup> Kanika Kanchan, PhD,<sup>b,\*</sup> Henry T. Bahnson, MS,<sup>c,d</sup> Ingo Ruczinski, PhD,<sup>e</sup> Gautam Shankar, MS,<sup>b</sup> Donald Y. M. Leung, MD, PhD,<sup>f</sup> Carolyn Baloh, MD,<sup>a,g</sup> George Du Toit, MBBCh,<sup>h</sup> Gideon Lack, MBBCh,<sup>h</sup> Gerald T. Nepom, MD, PhD,<sup>c,d</sup> and Rasika A. Mathias, ScD<sup>a</sup> *San Francisco, Calif; Baltimore, Md; Seattle, Wash; Denver, Colo; Boston, Mass; and London, United Kingdom*

**Background:** Deleterious variation in the epidermal differentiation complex (EDC) on chromosome 1 is a well-known genetic determinant of atopic dermatitis (AD) and has been associated with risk of peanut allergy (PA) in population-based studies.

**Objective:** Our aim was to determine the effect of genetic variation in the EDC on AD trajectory and risk of PA in early life.

**Methods:** Genome sequencing was used to measure genetic variation in the EDC in the Learning Early about Peanut Allergy (LEAP) study participants. Association tests were done to identify gene- and variant-level predicted deleterious variation associated with AD severity by using the Scoring Atopic Dermatitis (SCORAD) tool (n = 559) at baseline and each follow-up visit, as well as PA and food allergy in peanut avoiders (n = 275). Predicted deleterious variants included missense variants that were frameshift insertions, frameshift deletions, stop-gain mutations, or stop-loss mutations.

**Results:** The genes *FLG*, *FLG2*, *HRNR*, and *TCHH1* harbored the most predicted deleterious variation (30, 6, 3, and 1 variant, respectively). *FLG* variants were associated with SCORAD

score at all time points; 4 variants (R1798X, R501X, S126X, and S761fs) drove the association with SCORAD score at each time point, and higher variant load was associated with greater AD severity over time. There was an association between these variants and PA, which remained significant independent of baseline AD severity (odds ratio = 2.63 [95% CI = 1.11-6.01] [*P* = .02]).

**Conclusions:** Variation in *FLG* predicted to be deleterious is associated with AD severity at baseline and longitudinally and has an association with PA independent of baseline severity. (J Allergy Clin Immunol 2023;151:1137-42.)

**Key words:** Atopic dermatitis, eczema, peanut allergy, genetics, filaggrin, epidermal differentiation complex

## INTRODUCTION

The filaggrin gene (*FLG*) is a well-described genetic determinant of atopic dermatitis (AD),<sup>1</sup> and deleterious variants in *FLG* have been associated with early-onset severe disease.<sup>2-4</sup> *FLG* is located within the epidermal differentiation complex (EDC) on

From <sup>a</sup>the Immune Tolerance Network, San Francisco; <sup>b</sup>the Division of Allergy and Clinical Immunology, Department of Medicine, School of Medicine, and <sup>c</sup>the Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore; <sup>d</sup>the Immune Tolerance Network, Seattle; <sup>e</sup>the Benaroya Research Institute at Virginia Mason, Seattle; <sup>f</sup>the Division of Pediatric Allergy and Immunology, National Jewish Health, Denver; <sup>g</sup>the Division of Allergy and Clinical Immunology, Brigham and Women's Hospital, Boston; and <sup>h</sup>the Department of Pediatric Allergy, Division of Asthma, Allergy and Lung Biology, King's College London, and Guy's and St Thomas' NHS Foundation Trust, London.

\*These authors are co-first authors.

This research was performed as a project of the Immune Tolerance Network, an international clinical research consortium headquartered at the Benaroya Research Institute, and supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award no. UM1AI109565. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Disclosure of potential conflict of interest: H. T. Bahnson reports fees for contract work paid to his institution by Benaroya Research Institute, DBV Technologies, MYOR, King's College London, and Stanford University, as well as additional salary support paid by King's College London and Stanford University. D. Y. M. Leung reports consultant fees from Genentech, Sanofi-Genzyme, Leo Pharma, and Incyte Corporation. G. Du Toit reports grants from the National Institute of Allergy and Infectious Diseases (NIAID) of the National Institutes of Health (NIH), Food Allergy & Research Education (FARE), the Medical Research Council-Asthma (MRC-Asthma) UK Centre, UK Department of Health through the National Institute for Health Research

(NIHR), and the Action Medical Research and National Peanut Board; service as a member of the scientific advisory board of Aimmune; service as an investigator on pharma-sponsored allergy studies for Aimmune and DBV Technologies; and service as a scientific adviser to Aimmune, DBV, and Novartis. G. Lack reports grants from the National Institute of Allergy and Infectious Diseases (NIAID, NIH); other compensation from FARE, the MRC-Asthma UK Centre, the UK Department of Health through the NIHR, the National Peanut Board (NPB), and the Davis Foundation during the conduct of the study; ownership of shares in DBV Technologies, and Mighty Mission Me; and personal fees from Novartis, Sanofi-Genzyme, Regeneron, ALK-Abelló, and Lurie Children's Hospital outside the submitted work. The rest of the authors declare that they have no relevant conflicts of interest.

Received for publication July 17, 2022; revised November 9, 2022; accepted for publication November 11, 2022.

Available online November 17, 2022.

Corresponding author: Michelle F. Huffaker, MD, Immune Tolerance Network, University of California, San Francisco Diabetes Center, 513 Parnassus Ave, HSW 11, Box 0534, San Francisco, CA 94143-0534. E-mail: mhuffaker@immunetolerance.org.

The CrossMark symbol notifies online readers when updates have been made to the article such as errata or minor corrections

0091-6749

© 2022 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.2022.11.008>

**Abbreviations used**

|         |                                     |
|---------|-------------------------------------|
| AD:     | Atopic dermatitis                   |
| EDC:    | Epidermal differentiation complex   |
| FLG:    | Filaggrin                           |
| LEAP:   | Learning Early about Peanut Allergy |
| OR:     | Odds ratio                          |
| PA:     | Peanut allergy                      |
| SCORAD: | Scoring of Atopic Dermatitis        |
| WGS:    | Whole genome sequencing             |

chromosome 1, a cluster of genes critical for skin barrier function, and variants in this region outside of those in *FLG* have also been implicated in AD.<sup>5,6</sup> To our knowledge, no studies have examined the association of genetic variants across the EDC with the trajectory of AD over the first 5 years of life.

Deleterious variants in *FLG* have also been associated with peanut allergy (PA)<sup>7-10</sup>; however, there are gaps in our understanding of this relationship, particularly as to whether genetic risk is conferred in the absence of clinically apparent skin barrier dysfunction. A recent review examined prior work in detail,<sup>11</sup> and the association of *FLG* with food allergy in consideration of AD is unclear. Some studies demonstrate an association with food allergy independent of AD and/or AD severity,<sup>7,11</sup> whereas others do not find an independent association between the *FLG* variants and food allergy.<sup>11-13</sup> Therefore, a better understanding of the role of *FLG* in the first few years of life during the time of allergy development and a longitudinal evaluation of AD severity in this time frame are needed to clarify this relationship between genetics, AD, and food allergy.

The Learning Early about Peanut Allergy (LEAP) study enrolled infants aged 4 to 11 months with egg allergy and/or severe AD.<sup>14</sup> Participants were randomized to either peanut consumption or avoidance, and consumption proved to be highly protective against the development of PA. Whole genome sequencing (WGS) was previously performed on the LEAP study participants.<sup>15</sup> In this population enriched for AD, followed longitudinally over the first several years of life, and with challenge proven peanut allergic status, we have the unique opportunity to interrogate the relationship between genetic determinants of AD, AD severity, and PA. Here, we report the genetic predictors of AD in the LEAP study participants and their relationship with AD severity and the development of PA.

## RESULTS AND DISCUSSION

The landscape of exonic single-nucleotide variants and small frameshift insertion and deletions (INDELs) in the EDC in LEAP study participants is shown in Fig 1, A. The 4 genes with the greatest number of exonic variants are *FLG*, *FLG2*, *HRNR*, and *TCHH1*, and these are also the genes with the highest numbers of predicted deleterious variants (30, 6, 3, and 1 variant, respectively). The carrier rate and minor allele frequency of these predicted deleterious variants are presented in Table E1 (available in the Online Repository at [www.jacionline.org](http://www.jacionline.org)). The frequencies of the predicted deleterious variants are in general higher in the LEAP study than reported in the Exome Aggregation Consortium (ExAC) database (Fig 1, B and C); given the LEAP study

inclusion criteria of severe AD and the known importance of the EDC (and specifically *FLG*), this pattern is anticipated. The majority of the LEAP study participants (74%) were of European ancestry by self-report and genetic ancestry.<sup>15</sup>

Gene-based association tests were performed to identify predicted deleterious variants associated with AD severity by using the Scoring Atopic Dermatitis (SCORAD) tool at each study time point in both the peanut consumers and peanut avoiders (n = 559) for the 4 genes *FLG*, *FLG2*, *HRNR*, and *TCHH1* individually and in a combined test across all 4 genes (Fig 2). We found that predicted deleterious variants in the EDC were significantly associated with higher SCORAD scores across all time points ( $P < .001$ ) and that this association was driven by variation in *FLG* gene. A gene-based test including 30 *FLG* variants was associated with SCORAD score at all time points and with the SCORAD intercept (ie, the starting point of someone's AD severity at baseline) (Fig 2). No association was seen with the slope (ie, the general trajectory over the course [ $P = .29$ ]). To identify SCORAD score-driving variants in *FLG* at each study time point, we performed leave-1-out analysis and identified multiple *FLG* variants as driving the gene-based signal (Fig 3, A). Leave-1-out analysis for SCORAD intercept also revealed an association between the *FLG* variants and the SCORAD intercept ( $P = 2.58 \times 10^{-4}$ ) (Fig 3, B), thus implicating an association that begins within the first year of life. Further, to estimate the effect of *FLG*-predicted deleterious variants on SCORAD score, we performed single-variant association tests between *FLG* variants and SCORAD score at 30 months (Fig 3, A) and SCORAD intercept (Fig 3, B). For both SCORAD score at 30 months and SCORAD intercept, the variants that had the largest contributions to the gene-based signal (for SCORAD score at 30 months, R1798X, R501X, S126X, and S761fs; for SCORAD intercept, R2477X, R1798X, R501X, and S126X) also had significant association and effect sizes at the variant level.

Four *FLG* variants, namely, R1798X, R501X, S126X, and S761fs, were found to drive gene-based significance at all time points and for the intercept (Fig 3, C). These 4 variants are located in exon 3, the region of the gene comprising 10 to 12 tandem repeats of what becomes the FLG peptide (Fig 4<sup>16,17</sup>). R501X, S761fs, and R1798X, have previously been implicated in early onset of severe disease.<sup>4</sup> To our knowledge, this the first report of S126X. An association between variant load across these 4 variants and SCORAD score was tested by using a linear regression model. We found that 461 individuals had no deleterious alleles, 90 had 1 deleterious allele from any of the 4 variants, and 8 had 2; no individuals carried more than 2 deleterious alleles. Those with 2 variants had the highest SCORAD score at baseline, and although all study participants were noted to have an improvement in SCORAD score over the time line of the trial, those with 2 variant alleles maintained the highest severity across all time points (Fig 5).

Similar gene-based association tests using all predicted deleterious variants with PA in the peanut avoiders (n = 275) did not yield statistically significant associations (Fig 2). However, when we performed the gene-based test limited to only the 4 SCORAD score-driving variants in *FLG* (R1798X, R501X, S126X, and S761fs), we found a significant association with PA ( $P = .02$ ). To calculate an odds ratio (OR) for PA risk from these 4 variants, we performed logistic regression modeling for PA with

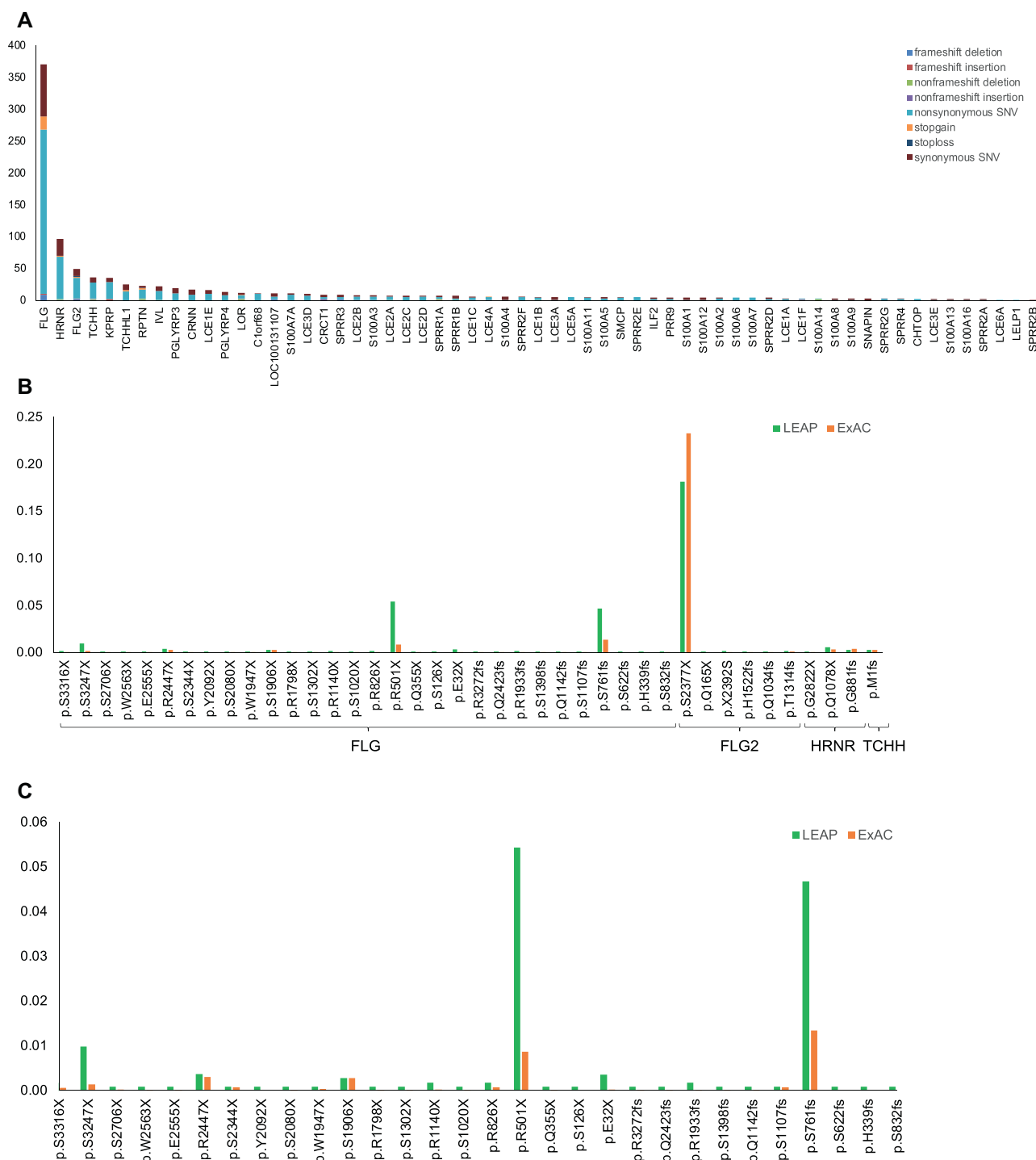


**FIG 1.** EDC region exonic variations and their frequencies in the LEAP study participants. **A**, Exonic variations in the EDC genes in the LEAP study participants. Genes are ordered on the total number of exonic variants observed, and bars show the representation of each class of variation. **B**, Frequencies of EDC deleterious variants in the LEAP study participants (green) and Exome Aggregation Consortium (ExAC) database (orange). **C**, FLG deleterious variant frequencies in the LEAP study participants (green) and ExAC database (orange).

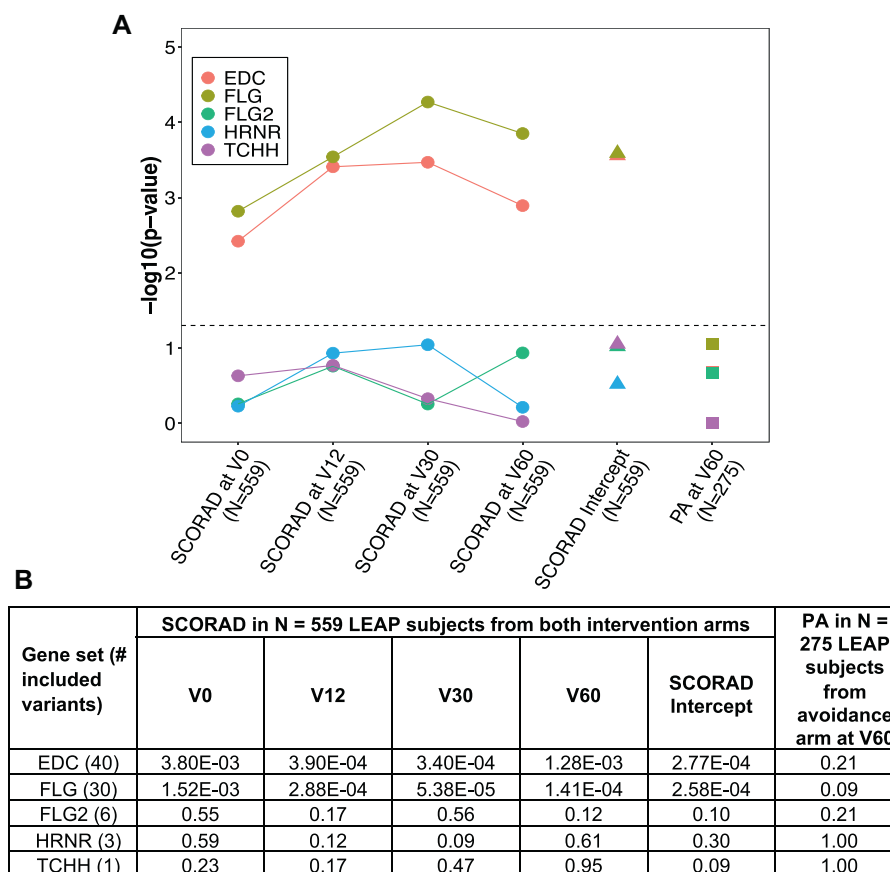


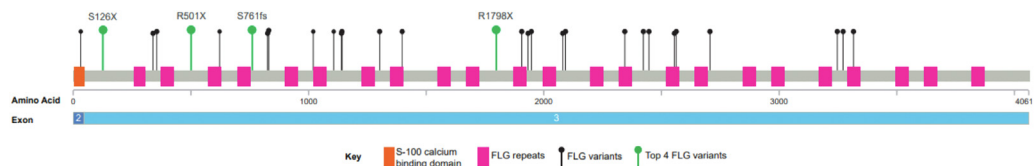
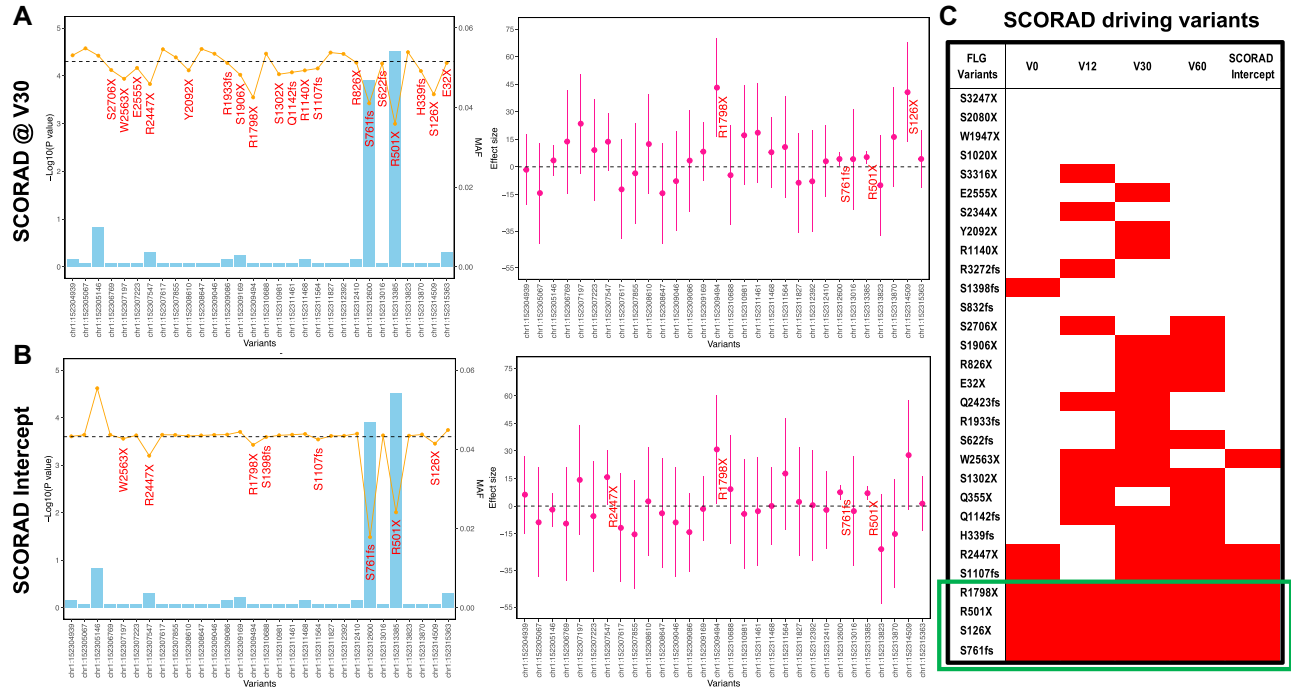
**FIG 2.** Association of deleterious variants in the EDC region with SCORAD score and PA. For SCORAD score, association was tested at all visits (at baseline and 12, 30, and 60 months of age) during the trial, and with SCORAD intercept (starting point of AD severity) in the full group of participants ( $n = 559$ ). For PA, the association was tested in the participants of the peanut avoidance group ( $n = 275$ ) at 60 months. **A**, Graphical illustration of the  $P$  values presented in **(B)**. The  $P$  values for SCORAD score at all visits are illustrated as points, for SCORAD intercept as triangles, and for PA at 60 months as squares. V0, Baseline visit; V12, 12-month visit; V30, 30-month visit; V60, 60-month visit.

carrier status defined as the presence of any of the 4 variants; we found an association with PA (OR = 3.13 [95% CI = 1.36-7.03] [ $P = .006$ ]). This association decreased but remained significant with adjustment for SCORAD score at baseline (OR = 2.63 [95% CI = 1.11-6.01] [ $P = .02$ ]). Similar to the findings by other groups, our findings do not indicate a residual effect of *FLG* on risk of PA development after adjustment for AD and/or AD severity at baseline entry into the study.<sup>7-9,11</sup>

There are several limitations to consider. First, the analysis was performed on variation in the EDC and did not consider variation outside this region. Second, there are genetic loci associated with AD outside the EDC, including variation in genes encoding for IL-4, IL-4 receptor, and IL-13 or others that could be considered in future analyses.<sup>18</sup> A third limitation is that the present work does not address environmental factors that may have contributed to the development and severity of AD and is instead limited to genetic analyses. Here, we are using genome sequencing at the standard depth of 30×, and although we have previously shown

the value of using WGS at this depth to study the EDC,<sup>19</sup> we recognize that we may have missed some EDC variants given the complexity of sequencing this region.<sup>20</sup> Another important limitation to consider is that it is difficult to tease apart the independent effects of the driving variants versus AD without these driving variants, as the population was selected for severe AD. An additional limitation is that although the variants are predicted to be deleterious, functional studies are needed to understand the impact on the skin barrier.

To our knowledge, this is the first report utilizing WGS to understand genetic risk on the trajectory of AD over the first several years of life. We found that participants carrying deleterious variation in the EDC, specifically *FLG*, had higher SCORAD scores across the first 5 years of life. The slope of improvement in AD severity over time, however, was similar in those with and without deleterious variants. This is also the first report utilizing WGS and well-characterized AD over the first 5 years of life to understand the relationship between *FLG* and



PA. The data suggest that there may be an association between *FLG* and PA independent of baseline AD severity. Identifying individuals with deleterious variation in *FLG* early may be important for mitigating PA risk through early oral introduction. Furthermore, future research is needed to understand the environmental factors that may contribute to AD development and severity, as well as the interplay of these environmental factors with underlying genetic risk. Whether the risk of food allergy in carriers of *FLG* variants may be reduced through early intervention and improvement of AD control remains unknown.

We thank Drs Daniel Rotrosen and Alkis Togias from the National Institute of Allergy and Infectious Diseases for their thoughtful review of the manuscript. We are grateful to all of the LEAP study participants who took part in the study.

**Clinical implications:** Variation predicted to be deleterious in EDC is associated with AD severity during the first 5 years of life. Deleterious variation in *FLG* is independently associated with PA.

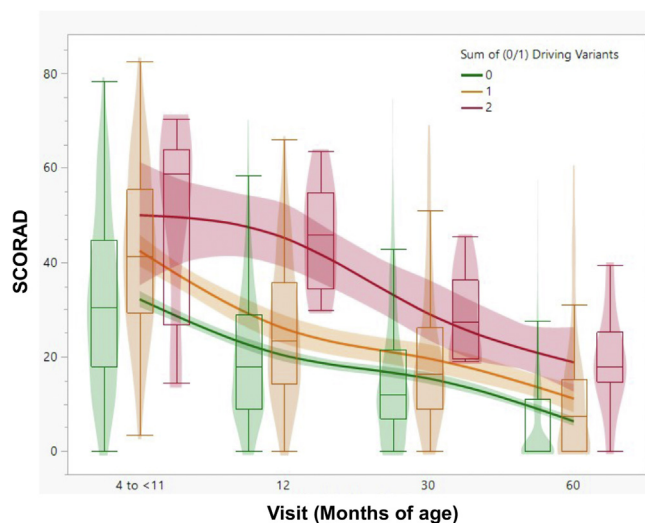


**FIG 5.** SCORAD score over time for LEAP study participants with 0, 1, or 2 driving variants. The unadjusted mean trajectory of the SCORAD score is shown by using a smoothed cubic spline with bootstrapped 95% CIs stratified by the number of driving variants. No study subjects were found to have more than 2 driving variants across the final 4 selected.

## REFERENCES

- Irvine AD, McLean WH, Leung DY. Filaggrin mutations associated with skin and allergic diseases. *N Engl J Med* 2011;365:1315-27.
- Barker JN, Palmer CN, Zhao Y, Liao H, Hull PR, Lee SP, et al. Null mutations in the filaggrin gene (FLG) determine major susceptibility to early-onset atopic dermatitis that persists into adulthood. *J Invest Dermatol* 2007;127:564-7.
- Hoyer A, Rehder EM, Fardig M, Asad S, Lodrup Carlsen KC, Endre KMA, et al. Filaggrin mutations in relation to skin barrier and atopic dermatitis in early infancy. *Br J Dermatol* 2022;186:544-52.
- Smieszek SP, Welsh S, Xiao C, Wang J, Polymeropoulos C, Birznies G, et al. Correlation of age-of-onset of atopic dermatitis with filaggrin loss-of-function variant status. *Sci Rep* 2020;10:2721.
- Mischke D, Korge BP, Marenholz I, Volz A, Ziegler A. Genes encoding structural proteins of epidermal cornification and S100 calcium-binding proteins form a gene complex ("epidermal differentiation complex") on human chromosome 1q21. *J Invest Dermatol* 1996;106:989-92.
- Paternoster L, Standl M, Chen CM, Ramasamy A, Bonnelykke K, Duijts L, et al. Meta-analysis of genome-wide association studies identifies three new risk loci for atopic dermatitis. *Nat Genet* 2011;44:187-92.
- Brown SJ, Asai Y, Cordell HJ, Campbell LE, Zhao Y, Liao H, et al. Loss-of-function variants in the filaggrin gene are a significant risk factor for peanut allergy. *J Allergy Clin Immunol* 2011;127:661-7.
- 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.
- Brough HA, Simpson A, Makinson K, Hankinson J, Brown S, Douiri A, et al. Peanut allergy: effect of environmental peanut exposure in children with filaggrin loss-of-function mutations. *J Allergy Clin Immunol* 2014;134:867-75.e1.
- Kalb B, Marenholz I, Jeanrenaud AC, Meixner L, Arnau-Soler A, Rosillo-Salazar OD, et al. Filaggrin loss-of-function mutations are associated with persistence of egg and milk allergy. *J Allergy Clin Immunol* 2022;150:1125-34.
- Kanchan K, Clay S, Irizar H, Bunyavanich S, Mathias RA. Current insights into the genetics of food allergy. *J Allergy Clin Immunol* 2021;147:15-28.
- Tan HT, Ellis JA, Koplin JJ, Matheson MC, Gurrin LC, Lowe AJ, et al. Filaggrin loss-of-function mutations do not predict food allergy over and above the risk of food sensitization among infants. *J Allergy Clin Immunol* 2012;130:1211-3.e3.
- Venkataraman D, Soto-Ramirez N, Kurukulaaratchy RJ, Holloway JW, Karmaus W, Ewart SL, et al. Filaggrin loss-of-function mutations are associated with food allergy in childhood and adolescence. *J Allergy Clin Immunol* 2014;134:876-82.e4.
- 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.
- Winters A, Bahnson HT, Ruczinski I, Boorgula MP, Malley C, Keramati AR, et al. The MALT1 locus and peanut avoidance in the risk for peanut allergy. *J Allergy Clin Immunol* 2019;143:2326-9.
- Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2012;2:401-4.
- Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 2013;6:pl1.
- Loiset M, Brown SJ, Saunes M, Hveem K. Genetics of atopic dermatitis: from DNA sequence to clinical relevance. *Dermatology* 2019;235:355-64.
- Bin L, Malley C, Taylor P, Preethi Boorgula M, Chavan S, Daya M, et al. Whole genome sequencing identifies novel genetic mutations in patients with eczema herpeticum. *Allergy* 2021;76:2510-23.
- Margolis DJ, Mitra N, Wubbenhorst B, Nathanson KL. Filaggrin sequencing and bioinformatics tools. *Arch Dermatol Res* 2020;312:155-8.

## METHODS

### Study participants and phenotype definitions

The detailed study design of the LEAP clinical trial was published previously.<sup>E1</sup> Briefly, participants were enrolled in the LEAP trial if they had egg allergy, severe AD, or both, and after enrolment they were randomly assigned to a dietary peanut avoidance or peanut consumption group.<sup>E1</sup> The data set for this genetics study includes a total of 559 participants from the LEAP trial who consented for genetics studies. Of these 559 LEAP study subjects, 542 completed the trial in accordance with their assigned group of peanut avoidance ( $n = 275$ ) or peanut consumption ( $n = 267$ ). The baseline characteristics of the LEAP study per-protocol participants were not different between the intervention groups.<sup>E2</sup>

In this study we focused on 2 phenotypes: AD severity and PA. AD severity was defined by SCORAD score by a clinician at the time of screening and at all study visits (when the participants were 12, 30, and 60 months of age), and PA was ascertained with a positive oral food challenge to peanuts at 60 months of age.

### WGS and quality control

WGS to a depth of 30× coverage was performed by using Illumina HiSeq X sequencer and HiSeq X HD SBS Kit reagents. Illumina HiSeq Control Software and Real-Time Analysis were used with the HiSeq X sequencers for real-time image analysis and base calling. Assembly of each individual genome was performed by using the Isaac aligner.<sup>E3</sup> The DRAGEN Germline Small Variant Caller was used to call both single-nucleotide variants (SNVs) and small insertions and deletions (INDELs), and to yield a genome variant file that includes variants along with quality metrics. We filtered the variants with GQX less than 30, DP less than 7, SNP hard quality less than 10.41, low-depth DP not exceeding 1, and ploidy conflict; variants not meeting thresholds for median base quality of alternate reads and likelihood. The analysis includes only biallelic SNVs and short INDELs.

### Ancestry deconvolution of LEAP genetics study participants

We used PC-AiR to perform a principal components analysis (PCA) on the WGS data for detection of population structure.<sup>E4</sup> Unlike standard PCA, PC-AiR accounts for relatedness in the sample to provide accurate ancestry inference that is not confounded by family structure. The gene-based association tests include the first 8 principal components (PCs) for ancestry.

### Gene variant annotation

Variant annotation was performed by using ANNOVAR<sup>E5</sup>; we were able to annotate a total of 5171 variants mapping to 61 genes in the EDC region (Fig 1, A). Annotated variants can be classified into 2 categories: exonic including frameshift deletion, frameshift insertion, nonframeshift deletion, nonframeshift insertion, nonsynonymous SNV, synonymous SNV, stop-gain, stop-loss, and noncoding. For our association analysis we focused on a set of deleterious variants that included only missense variants having exonic functional annotations of either frameshift insertion, frameshift deletion, stop-gain mutation, or stop-loss mutation. The predictions about the effect of these variants on protein structure and function and the information about the relationship of these variations and human health are shown in Table E1. We did not pre-filter variants on the basis of their allele frequency within LEAP or on their

comparative allele frequency to public catalogs such as ExAC; this was to allow for both protective and risk effects (eg, filtering to only those variants that were higher in LEAP would have limited our findings to presumed risk alleles) and to allow for common and rare variants to be included in the gene-based tests described in the following sections.

### Gene-based tests for association

Gene-based tests were performed by using the SKAT-O (SNP-set [Sequence] Kernel Association Test, Optimal) test from the R package SKAT.<sup>E6</sup> The Bonferroni multiple-comparison correction threshold of a  $P$  value less than .01 ( $P < .05/5$ , where 5 is the number of specific genes tests) was used to account for multiple testing. Tests for SCORAD score were performed in the full group of participants ( $N = 559$ ) by using SCORAD score at each visit. In addition to SCORAD score at each visit, we also used “slope” and “intercept” for each subject’s AD trajectory over the course of LEAP trial to fully model the longitudinal nature of the measured phenotype. Here, the intercept represents the starting point of someone’s AD severity at baseline and the slope represents the general trajectory over the course. The tests for PA were performed in the avoidance group participants ( $n = 275$ ). Age at baseline, sex, and 8 genetic PCs were included as covariates in all of these association test models.

We sought to determine which deleterious variants from the gene-based tests were driving the association by signal using a leave-1-out approach. This was accomplished by repeated gene-based tests using SKAT-O, each time leaving out a single variant and iterating over all variants. If a variant has a large contribution to the gene-based association signal, then a weakened association is represented as “dips” in the plots below the gene-based  $P$  value.

### Regression models

Given that the gene-based test does not output ORs, we tested the association between SCORAD score-driving variants and PA by using logistic regression modeling. We ran 2 regression models for PA: the first without adjustment for SCORAD score at baseline, and the second with adjustment for SCORAD scored at baseline. In both models, PA was included as a binary outcome, and age at baseline, sex, and 8 PCs were included as covariates.

### REFERENCES

- E1. 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.
- E2. Kanchan K, Grinek S, Bahnson HT, Ruczinski I, Shankar G, Larson D, et al. HLA alleles and sustained peanut consumption promote IgG4 responses in subjects protected from peanut allergy. *J Clin Invest* 2022;132:e152070.
- E3. Racz C, Petrovski R, Saunders CT, Chorny I, Kruglyak S, Margulies EH, et al. Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. *Bioinformatics* 2013;29:2041-3.
- E4. 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.
- E5. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;38:e164.
- E6. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet* 2012;91:224-37.

**TABLE E1.** Predicted deleterious variants of the genes *FLG*, *FLG2*, *HRNR*, and *TCHH*

| Annotation             |             |                 | Carrier rate and frequency |            |                 | Functional prediction |      |       | Clinical information |                                                           |                                                                                                                     |                                                      |                                              |
|------------------------|-------------|-----------------|----------------------------|------------|-----------------|-----------------------|------|-------|----------------------|-----------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------|------------------------------------------------------|----------------------------------------------|
| Variant                | rsID        | Exonic function | Carrier rate (LEAP)        | MAF (LEAP) | Freq (ExAC All) | SIFT                  | CADD | GERP  | Allele ID            | Condition/ disorder                                       | Disorder database identifier                                                                                        | Review status                                        | Clinical significance                        |
| <i>FLG</i><br>p.S3316X | rs149484917 | Stop-gain       | 0.36%                      | 0.18%      | 0.06%           | D                     | 49   | −1.03 | 263963               | Ichthyosis vulgaris/not provided                          | MONDO:MONDO:0024304, MedGen:C0079584, OMIM:146700/MedGen:CN517202                                                   | Criteria provided, multiple submitters, no conflicts | Pathogenic/likely pathogenic                 |
| p.S3247X               | rs150597413 | Stop-gain       | 1.97%                      | 0.99%      | 0.14%           | T                     | 48   | −1.39 | 65592                | AD/ichthyosis vulgaris/not provided                       | MONDO:MONDO:0011596, MedGen:C1853965, OMIM:605803/MONDO:MONDO:0024304, MedGen:C0079584, OMIM:146700/MedGen:CN517202 | Criteria provided, multiple submitters, no conflicts | Pathogenic/likely pathogenic                 |
| p.S2706X               | rs542799026 | Stop-gain       | 0.18%                      | 0.09%      | 0.03%           | T                     | 47   | 0.75  | 536220               | Not provided                                              | MedGen:CN517202                                                                                                     | Criteria provided, single submitter                  | Pathogenic                                   |
| p.W2563X               | rs767179116 | Stop-gain       | 0.18%                      | 0.09%      | 0.00%           | T                     | 45   | −2.38 | —                    | —                                                         | —                                                                                                                   | —                                                    | —                                            |
| p.E2555X               | rs760728737 | Stop-gain       | 0.18%                      | 0.09%      | 0.00%           | T                     | 44   | −2.54 | —                    | —                                                         | —                                                                                                                   | —                                                    | —                                            |
| p.R2447X               | rs138726443 | Stop-gain       | 0.73%                      | 0.37%      | 0.30%           | T                     | 45   | −3.61 | 65594                | AD/ichthyosis vulgaris/not provided                       | MONDO:MONDO:0011596, MedGen:C1853965, OMIM:605803/MONDO:MONDO:0024304, MedGen:C0079584, OMIM:146700/MedGen:CN517202 | Criteria provided, multiple submitters, no conflicts | Pathogenic                                   |
| p.S2344X               | rs372754256 | Stop-gain       | 0.18%                      | 0.09%      | 0.07%           | D                     | 45   | 0.85  | 511173               | Inborn genetic diseases/not provided                      | MeSH:D030342, MedGen:C0950123/MedGen:CN517202                                                                       | Criteria provided, multiple submitters, no conflicts | Pathogenic                                   |
| p.Y2092X               | —           | Stop-gain       | 0.18%                      | 0.09%      | —               | T                     | 43   | −2.46 | 611488               | Not provided                                              | MedGen:CN517202                                                                                                     | Criteria provided, single submitter                  | Pathogenic                                   |
| p.S2080X               | rs147145234 | Stop-gain       | 0.18%                      | 0.09%      | 0.01%           | T                     | 43   | −1.35 | 226865               | Inborn genetic diseases/not provided                      | MeSH:D030342, MedGen:C0950123/MedGen:CN517202                                                                       | Criteria provided, multiple submitters, no conflicts | Pathogenic                                   |
| p.W1947X               | rs560912365 | Stop-gain       | 0.18%                      | 0.09%      | 0.04%           | T                     | 40   | −2.44 | —                    | —                                                         | —                                                                                                                   | —                                                    | —                                            |
| p.S1906X               | rs141784184 | Stop-gain       | 0.56%                      | 0.28%      | 0.28%           | T                     | 41   | −0.71 | 227200               | Ichthyosis vulgaris                                       | MONDO:MONDO:0024304, MedGen:C0079584, OMIM:146700/MedGen:CN517202                                                   | Criteria provided, single submitter                  | Likely pathogenic                            |
| p.R1798X               | rs149105551 | Stop-gain       | 0.18%                      | 0.09%      | 0.01%           | T                     | 41   | −3.74 | 404915               | Not provided                                              | MedGen:CN517202                                                                                                     | Criteria provided, single submitter                  | Pathogenic                                   |
| p.S1302X               | rs754812742 | Stop-gain       | 0.18%                      | 0.09%      | 0.01%           | T                     | 38   | 1.62  | 227201               | Ichthyosis vulgaris/not provided                          | MONDO:MONDO:0024304, MedGen:C0079584, OMIM:146700/MedGen:CN517202                                                   | Criteria provided, multiple submitters, no conflicts | Pathogenic                                   |
| p.R1140X               | rs141375260 | Stop-gain       | 0.36%                      | 0.18%      | 0.02%           | T                     | 40   | −3.86 | 611490               | Not provided                                              | MedGen:CN517202                                                                                                     | Criteria provided, single submitter                  | Pathogenic                                   |
| p.S1020X               | rs200360684 | Stop-gain       | 0.18%                      | 0.09%      | 0.01%           | T                     | 37   | 0.58  | 611492               | Not provided                                              | MedGen:CN517202                                                                                                     | Criteria provided, single submitter                  | Pathogenic                                   |
| p.R826X                | rs115746363 | Stop-gain       | 0.36%                      | 0.18%      | 0.07%           | T                     | 27.3 | −3.38 | 259630               | FLG-related disorders/AD/ichthyosis vulgaris/not provided | MONDO:MONDO:0011596, MedGen:C1853965, OMIM:605803/MONDO:MONDO:0024304, MedGen:C0079584, OMIM:146700/MedGen:CN517202 | Criteria provided, conflicting interpretations       | Conflicting interpretations of pathogenicity |

(Continued)

TABLE E1. (Continued)

| Annotation |              |                     | Carrier rate and frequency |            |                 | Functional prediction |       |       | Clinical information |                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                         |                                                      |                                              |
|------------|--------------|---------------------|----------------------------|------------|-----------------|-----------------------|-------|-------|----------------------|---------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------|----------------------------------------------|
| Variant    | rsID         | Exonic function     | Carrier rate (LEAP)        | MAF (LEAP) | Freq (ExAC All) | SIFT                  | CADD  | GERP  | Allele ID            | Condition/ disorder                                                                   | Disorder database identifier                                                                                                                                                                                                                                                                                                                                                            | Review status                                        | Clinical significance                        |
| p.R501X    | rs61816761   | Stop-gain           | 10.13%                     | 5.43%      | 0.87%           | T                     | 28.5  | −4.14 | 31358                | FLG-related disorders/eczema/AD/ichthyosis vulgaris/susceptibility to AD/not provided | Human Phenotype Ontology:HP:0000964.Human Phenotype Ontology:HP:0001481.MedGen:C0013595/ Human Phenotype Ontology:HP:0001047. Human Phenotype Ontology:HP:0007533.Human Phenotype Ontology:HP:0007564.MedGen:C0011615, OMIM:603165, OMIM:PS603165/MONDO:MONDO:0011596, MedGen:C1853965, OMIM:605803/ MONDO:MONDO:0024304, MedGen:C0079584, OMIM:146700/MedGen:C2675432/ MedGen:CN517202 | Criteria provided, multiple submitters, no conflicts | Pathogenic/likely pathogenic                 |
| p.Q355X    | rs142991475  | Stop-gain           | 0.18%                      | 0.09%      | 0.00%           | T                     | 18.73 | −0.48 | 263942               | Not provided                                                                          | MedGen:CN517202                                                                                                                                                                                                                                                                                                                                                                         | Criteria provided, single submitter                  | Pathogenic                                   |
| p.S126X    | rs1405588197 | Stop-gain           | 0.18%                      | 0.09%      | —               | T                     | 12.88 | −2.14 | —                    | —                                                                                     | —                                                                                                                                                                                                                                                                                                                                                                                       | —                                                    | —                                            |
| p.E32X     | rs114733570  | Stop-gain           | 0.72%                      | 0.36%      | 0.02%           | D                     | 31    | 2.43  | 263950               | FLG-related disorders/ichthyosis vulgaris/not provided                                | MONDO:MONDO:0024304.MedGen:C0079584, OMIM:146700/ MedGen:CN517202                                                                                                                                                                                                                                                                                                                       | Criteria provided, multiple submitters, no conflicts | Pathogenic/likely pathogenic                 |
| p.R3272fs  | rs769826402  | Frameshift deletion | 0.18%                      | 0.09%      | 0.00%           | —                     | —     | −0.16 | 805161               | Not provided                                                                          | MedGen:CN517202                                                                                                                                                                                                                                                                                                                                                                         | Criteria provided, single submitter                  | Pathogenic                                   |
| p.Q2423fs  | rs760831749  | Frameshift deletion | 0.18%                      | 0.09%      | 0.01%           | —                     | —     | −0.33 | 259627               | Not provided                                                                          | MedGen:CN517202                                                                                                                                                                                                                                                                                                                                                                         | Criteria provided, single submitter                  | Pathogenic                                   |
| p.R1933fs  | rs753115579  | Frameshift deletion | 0.36%                      | 0.18%      | 0.00%           | —                     | —     | −2.70 | —                    | —                                                                                     | —                                                                                                                                                                                                                                                                                                                                                                                       | —                                                    | —                                            |
| p.S1398fs  | rs757562290  | Frameshift deletion | 0.18%                      | 0.09%      | 0.00%           | —                     | —     | 1.71  | —                    | —                                                                                     | —                                                                                                                                                                                                                                                                                                                                                                                       | —                                                    | —                                            |
| p.Q1142fs  | rs542661380  | Frameshift deletion | 0.18%                      | 0.09%      | 0.01%           | —                     | —     | 0.01  | —                    | —                                                                                     | —                                                                                                                                                                                                                                                                                                                                                                                       | —                                                    | —                                            |
| p.S1107fs  | rs200519781  | Frameshift deletion | 0.18%                      | 0.09%      | 0.07%           | —                     | —     | −3.43 | 404918               | Ichthyosis vulgaris/ susceptibility to AD/not provided                                | MONDO:MONDO:0024304.MedGen:C0079584,OMIM:146700/MedGen:C2675432/ MedGen:CN517202                                                                                                                                                                                                                                                                                                        | Criteria provided, single submitter                  | Pathogenic                                   |
| p.S761fs   | rs558269137  | Frameshift deletion | 8.63%                      | 4.68%      | 1.34%           | —                     | —     | 0.56  | 31359                | FLG-related disorder/eczema/AD/ichthyosis vulgaris/ susceptibility to AD/not provided | Human Phenotype Ontology:HP:0000964, Human Phenotype Ontology:HP:0001481, MedGen:C0013595/ MONDO:MONDO:0011596, MedGen:C1853965, OMIM:605803/MONDO:MONDO:0024304,MedGen:C0079584, OMIM:146700/MedGen:C2675432/MedGen:CN517202                                                                                                                                                           | Criteria provided, conflicting interpretations       | Conflicting interpretations of pathogenicity |
| p.S622fs   | —            | Frameshift deletion | 0.18%                      | 0.09%      | —               | —                     | —     | —     | —                    | —                                                                                     | —                                                                                                                                                                                                                                                                                                                                                                                       | —                                                    | —                                            |
| p.H339fs   | —            | Frameshift deletion | 0.18%                      | 0.09%      | —               | —                     | —     | —     | —                    | —                                                                                     | —                                                                                                                                                                                                                                                                                                                                                                                       | —                                                    | —                                            |

(Continued)

TABLE E1. (Continued)

| Annotation       |             |                      | Carrier rate and frequency |            |                 | Functional prediction |       |       | Clinical information |                    |                              |                                     |                       |
|------------------|-------------|----------------------|----------------------------|------------|-----------------|-----------------------|-------|-------|----------------------|--------------------|------------------------------|-------------------------------------|-----------------------|
| Variant          | rsID        | Exonic function      | Carrier rate (LEAP)        | MAF (LEAP) | Freq (ExAC All) | SIFT                  | CADD  | GERP  | Allele ID            | Condition/disorder | Disorder database identifier | Review status                       | Clinical significance |
| <b>p.S832fs</b>  | rs564976220 | Frameshift insertion | 0.18%                      | 0.09%      | 0.02%           | —                     | —     | 1.91  | —                    | —                  | —                            | —                                   | —                     |
| <b>FLG2</b>      |             |                      |                            |            |                 |                       |       |       |                      |                    |                              |                                     |                       |
| <b>p.S2377X</b>  | rs12568784  | Stop-gain            | 31.90%                     | 18.10%     | 23.23%          | T                     | 45    | −4.50 | —                    | —                  | —                            | —                                   | —                     |
| <b>p.Q165X</b>   | .           | Stop-gain            | 0.18%                      | 0.09%      | —               | T                     | 18.37 | —     | —                    | —                  | —                            | —                                   | —                     |
| <b>p.X2392S</b>  | rs150529054 | Stop-loss            | 0.36%                      | 0.18%      | 0.06%           | —                     | 2.88  | −2.66 | —                    | —                  | —                            | —                                   | —                     |
| <b>p.H1522fs</b> | rs528795062 | Frameshift deletion  | 0.18%                      | 0.09%      | 0.06%           | —                     | —     | −1.17 | —                    | —                  | —                            | —                                   | —                     |
| <b>p.Q1034fs</b> | rs556285121 | Frameshift deletion  | 0.18%                      | 0.09%      | 0.06%           | —                     | —     | 0.97  | —                    | —                  | —                            | —                                   | —                     |
| <b>p.T1314fs</b> | rs567184084 | Frameshift insertion | 0.36%                      | 0.18%      | 0.10%           | —                     | —     | −0.55 | —                    | —                  | —                            | —                                   | —                     |
| <b>HRNR</b>      |             |                      |                            |            |                 |                       |       |       |                      |                    |                              |                                     |                       |
| <b>p.G2822X</b>  | rs151033591 | Stop-gain            | 0.18%                      | 0.09%      | 0.08%           | T                     | 47    | 0.43  | —                    | —                  | —                            | —                                   | —                     |
| <b>p.Q1078X</b>  | rs141950802 | Stop-gain            | 1.08%                      | 0.54%      | 0.36%           | T                     | 37    | 1.06  | —                    | —                  | —                            | —                                   | —                     |
| <b>p.G881fs</b>  | rs762305427 | Frameshift deletion  | 0.54%                      | 0.27%      | 0.40%           | —                     | —     | 1.32  | —                    | —                  | —                            | —                                   | —                     |
| <b>TCHH</b>      |             |                      |                            |            |                 |                       |       |       |                      |                    |                              |                                     |                       |
| <b>p.M1fs</b>    | rs141946179 | Frameshift deletion  | 0.54%                      | 0.27%      | 0.29%           | —                     | —     | 2.18  | 718224               | Not provided       | MedGen:CN517202              | Criteria provided, single submitter | Likely benign         |

SIFT predicts whether an amino acid substitution is likely to affect protein function based on sequence homology and the physiochemical similarity between the alternate amino acids. CADD scores the predicted deleteriousness of SNVs and insertion/deletions by integrating multiple annotations including conservation and functional information into 1 metric. Phred-style CADD raw scores are displayed, with scores higher than 30 indicating "likely deleterious" and scores below indicating "likely benign." GERP identifies constrained loci in multiple sequence alignments by comparing the level of substitution observed with that expected if there was no functional constraint. Positive scores represent highly conserved positions, whereas negative scores represent highly variable positions. Allele ID is the identifier for each individual variant, which may be interpreted as a single variant or as part of a haplotype or genotype. Condition provides the condition(s) for which the variant has been interpreted. Disorder database identifier is the database identifier for the disorder or condition for which the variant has been interpreted. Review status consists of ClinVar reports for the level of review supporting the assertion of clinical significance for the variation as review status. The prediction of deleteriousness by SIFT equal to D or CADD greater than 30 or GERP greater than 0 are bolded.

CADD, Combined annotation-dependent depletion; D, deleterious; ExAC, Exome Aggregation Consortium; Freq, frequency; GERP, Genomic Evolutionary Rate Pro; ID, identifier; MAF, minor allele frequency; OMIM, Online Mendelian Inheritance in Man; rsID, reference single-nucleotide polymorphism identifier; SIFT, scale-invariant feature transform; T, tolerated.