View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Elsevier - Publisher Connector

 See related commentary on pg 1593

Association Screening in the Epidermal Differentiation Complex (EDC) Identifies an *SPRR3* Repeat Number Variant as a Risk Factor for Eczema

Ingo Marenholz[1,2,7], Vladimir A. Gimenez Rivera[1,2,7], Jorge Esparza-Gordillo[1,2], Anja Bauerfeind[2], Min-Ae Lee-Kirsch[3], Andrzej Ciechanowicz[4], Michael Kurek[5], Tereza Piskackova[6], Milan Macek[6] and Young-Ae Lee[1,2]

The genetically determined impairment of the skin barrier is a primary cause of eczema. As numerous genes essential for an intact epidermis reside within the epidermal differentiation complex (EDC), we screened the National Center for Biotechnology Information (NCBI) database for putatively functional polymorphisms in the EDC genes and tested them for association with eczema. We identified 20 polymorphisms with predicted major impact on protein function. Of these, 4 were validated in 94 eczema patients: a nonsense mutation in *FLG2* (rs12568784), a stop codon mutation in *LCE1D* (rs41268500), a 24-bp deletion in *SPRR3* (rs28989168), and a frameshift mutation in *S100A3* (rs11390146). The minor allele frequencies were 15.1, 6.1, 47.2, and 0.4%, respectively. Association testing of the validated polymorphisms in 555 eczema patients and 375 controls identified a significant effect of rs28989168 (*SPRR3*) on eczema. The association was replicated in another 1,314 cases and 1,322 controls, yielding an overall odds ratio of 1.30 (95% confidence interval 1.12–1.51; $P=0.00067$) for a dominant mode of inheritance. Small proline-rich proteins (SPRRs) are crossbridging proteins in the cornified cell envelope (CE), which provides the main barrier function of stratified squamous epithelia. The *SPRR3* variant associated with eczema carried an extra 24-bp repeat in the central domain, which may alter the physical properties of the CE.

Journal of Investigative Dermatology (2011) **131,** 1644–1649; doi:10.1038/jid.2011.90; published online 14 April 2011

INTRODUCTION

Eczema (atopic dermatitis) is a chronic inflammatory skin disease, which is characterized by a defect of the skin barrier and a dysregulated systemic immune response to common environmental allergens. In industrialized countries, up to 20% of the children are affected, and family and twin studies indicated that genetic factors contribute significantly to the disease (Lichtenstein and Svartengren, 1997; van Beijsterveldt and Boomsma, 2007). Although the molecular mechanisms underlying eczema are not fully understood, previous findings on loss-of-function mutations in the filaggrin gene (*FLG*) indicated that the impaired function of the epidermal barrier is a primary event in the development of eczema (Palmer *et al.*, 2006).

FLG is located within the epidermal differentiation complex (EDC), a region of 1.7 Mb on chromosome 1q21.3 comprising over 60 genes, most of which are expressed during terminal differentiation of the epidermis (Mischke *et al.*, 1996; Marenholz *et al.*, 2001). The EDC genes constitute several functionally, structurally, and evolutionarily related gene families, which are important for the maturation and cornification of the epidermis. They encode the precursor proteins of the cornified cell envelope (CE), such as small proline rich proteins (SPRRs) and late cornified envelope (LCE) proteins, both of which contain short tandem peptide repeats in the central domain (Hohl *et al.*, 1995; Marshall *et al.*, 2001). Moreover, they include the family of fused S100 genes that are characterized by an N-terminal S100-like calcium-binding domain fused to a large tandem repeat domain. Members of this family such as filaggrin and trichohyalin are crosslinked to the CE and serve as a matrix for the aggregation of keratin intermediate filaments in terminally differentiating keratinocytes (Presland and Dale,

[1]*Pediatric Pneumology and Immunology and Experimental and Clinical Research Center, Charité, Berlin, Germany;* [2]*Max-Delbrück-Centrum (MDC) for Molecular Medicine, Berlin, Germany;* [3]*Klinik fur Kinder- und Jugendmedizin, Technical University Dresden, Dresden, Germany;* [4]*Department of Laboratory Diagnostics and Molecular Medicine, Pomeranian Medical University, Szczecin, Poland;* [5]*Department of Clinical Allergology, Pomeranian Medical University, Szczecin, Poland and* [6]*Department of Biology and Medical Genetics, Charles University Prague, 2nd Medical School and Faculty, Hospital Motol, Prague, Czech Republic*

[7]*These authors contributed equally to this work.*

Correspondence: Young-Ae Lee, Pediatric Pneumology and Immunology and Experimental and Clinical Research Center, Charité, Campus Virchow Klinikum, Augustenburger Platz 1, D-13353 Berlin, Germany. E-mail: yo.lee@charite.de

Abbreviations: CE, cornified cell envelope; CI, confidence interval; EDC, epidermal differentiation complex; FLG, filaggrin gene; LCE, late cornified envelope; NCBI, National Center for Biotechnology Information; OR, odds ratio; RR, relative risk; SNP, single-nucleotide polymorphism; SPRR, small proline-rich protein

Received 8 December 2010; revised 25 February 2011; accepted 6 March 2011; published online 14 April 2011

 © 2011 The Society for Investigative Dermatology

2000; Listwan and Rothnagel, 2004). All these structural genes are flanked by 16 members of the S100 family, encoding small, calcium-binding S100 proteins, which display a wide range of inflammatory, immunological, and antimicrobial activities (Marenholz *et al.*, 2004).

The high degree of relationship among the genes of the EDC, their coordinated expression, as well as the orchestrated interaction of the encoded proteins in the formation of the epidermis suggest that genetic variants in EDC genes increase susceptibility to eczema similar to the well-established *FLG* mutations. Involvement of additional EDC genes in the pathogenesis of eczema is further supported by expression analyses that demonstrated dysregulated gene expression within the EDC region to be associated with the disease (Sugiura *et al.*, 2005; Esparza-Gordillo *et al.*, 2009; Guttman-Yassky *et al.*, 2009). Finally, a recent genome-wide association study has pointed to an eczema risk variant within the EDC apart from the mutations in *FLG* (Esparza-Gordillo *et al.*, 2009). We therefore screened the National Center for Biotechnology Information (NCBI) single-nucleotide polymorphism (SNP) database for polymorphisms in the EDC genes that were likely to alter protein function, including frameshift, nonsense, and stop codon mutations, as well as small deletions/insertions. Validated polymorphisms were then tested for association using a case–control study design. Finally, we demonstrated that a positive association result was not because of linkage disequilibrium with the *FLG* mutations.

RESULTS

Screening and validation of polymorphisms in the EDC genes

For the identification of polymorphisms within the EDC genes, we used the information present in the human NCBI SNP database (Build 36.3). We restricted our search to polymorphisms and mutations that are likely to have a major impact on protein function. Our systematic screening of the EDC region between *S100A10* (chr1:150.222 Mb) and *S100A1* (chr1:151.871 Mb) identified 13 frameshift mutations, 4 nonsense mutations, 2 stop codon mutations, and 1 deletion of 24 bp, which affected a total of 14 EDC genes (Supplementary Table S1 online). These were distributed over the entire region and affected members of all gene families of the EDC. Only for one of the polymorphisms, rs12568784, an allele frequency was reported in the NCBI SNP database. We then aimed to verify the polymorphisms in a set of 94 children with eczema from Germany. Only 4 out of 20 variants were validated by sequencing (Supplementary Table S1 online); the nonsense mutation c.7130A in the *filaggrin 2* gene (*FLG2*; rs12568784), the stop codon mutation p.X115G in *LCE1D*, the 24-bp deletion c.196_219del in *SPRR3* (rs28989168), which lacks one out of 14 tandem sequence repeats, and the frameshift mutation c.207delG in *S100A3* (rs11390146).

Testing the validated polymorphisms for association with eczema

Subsequently, the validated polymorphisms were tested for association in the discovery set, comprising 555 children with onset of eczema in the first 2 years of life and 375 unselected control individuals from Germany. For rs28989168 in *SPRR3*, we found an association of the insertion variant carrying 14 repeats with eczema yielding an odds ratio (OR) of 1.25 (95% confidence interval (CI) 1.04–1.50; $P=0.020$), whereas the polymorphisms in *FLG2*, *LCE1D*, and *S100A3* were not associated with eczema (Table 1). The allele frequencies among the controls were 47.2% for the risk allele in *SPRR3*, 15.1% for the nonsense mutation in *FLG2*, 6.1% for the stop codon mutation in *LCE1D*, and 0.4% for the frameshift mutation in *S100A3*.

Table 1. Association between the validated polymorphisms in the EDC genes and eczema in the discovery set

SNP ID	Gene	Minor allele	Allele frequency: Eczema	Allele frequency: Controls	OR (95% CI)	*P*-value
rs12568784	*FLG2*	A	13.4%	15.1%	0.87 (0.67–1.14)	0.33
rs41268500	*LCE1D*	G	6.0%	6.1%	0.98 (0.65–1.48)	0.93
rs28989168	*SPRR3*	14 repeats	52.7%	47.2%	1.25 (1.04–1.50)	0.021
rs11390146	*S100A3*	del (G)	0.5%	0.4%	1.34 (0.33–5.38)	0.68

Abbreviations: CI, confidence interval; EDC, epidermal differentiation complex; OR, odds ratio; SNP, single-nucleotide polymorphism.

Considering the distinct allele frequencies of the validated polymorphisms and a disease prevalence of 15%, we estimated the power of the discovery set to detect a potential effect of the risk variants. Power calculations were performed under an additive model with different relative risks (RRs) for the heterozygous (Aa) or homozygous (AA) carriers of the risk variants (Supplementary Table S2 online). Although the power of our sample was 97 and 75% to detect a moderate effect ($RR_{Aa}=1.5$ and $RR_{AA}=2.25$) of rs12568784 (*FLG2*) and rs41268500 (*LCE1D*), respectively, the power dropped to 11% for the frameshift mutation in *S100A3*, which was because of the low frequency of this allele.

Replicating the association of *SPRR3* with eczema

For the replication, we used 1,314 eczema cases with an age of onset <2 years and 1,322 population-based controls. We confirmed the association of rs28989168 with eczema, which we had identified in the discovery set, with a similar effect size (Table 2). For the combined sample we obtained an OR of 1.16 (95% CI 1.06–1.27, $P=0.0022$). In order to investigate the mode of inheritance in more detail, we then calculated the genotype ORs for heterozygous (OR_{het}) and homozygous (OR_{hom}) carriers of the risk allele separately. Heterozygotes and homozygotes yielded similar ORs compared with the noncarriers of the risk allele (Table 3). Accordingly, our data suggested a dominant mode of inheritance for the risk allele of *SPRR3* in eczema (OR 1.30, 95% CI 1.12–1.51, $P=0.00067$).

Table 2. Association between the *SPRR3* variant rs28989168 and eczema

	No. of individuals		Risk allele frequency			
Set	Eczema	Controls	Eczema	Controls	OR (95% CI)	*P*-value
Discovery	555	375	52.7%	47.2%	1.25 (1.04–1.50)	0.021
Replication	1,315	1,322	50.5%	47.5%	1.13 (1.01–1.25)	0.033
Combined	1,870	1,697	51.1%	47.5%	1.16 (1.06–1.27)	0.0020

Abbreviations: CI, confidence interval; OR, odds ratio; *SPRR3*, *small proline-rich protein 3*.

Table 3. Mode of inheritance of the *SPRR3* variant rs28989168 in eczema

	No. of individuals			
Genotype	Eczema	Controls	Genotype OR	
13/13 Repeats	442	486		
13/14 Repeats	944	812	OR_{het} (95% CI)	1.28 (1.09–1.50)
14/14 Repeats	484	399	OR_{hom} (95% CI)	1.33 (1.11–1.60)

Abbreviations: CI, confidence interval; OR_{het}, odds ratio for heterozygotes; OR_{hom}, odds ratio for homozygotes; *SPRR3*, *small proline-rich protein 3*.

Analyzing the *FLG* mutations as potential confounders on the eczema risk of rs28989168

Finally, we examined whether the mutations in *FLG*, which is also located within the EDC, had an influence on our results. The *FLG* mutations are the strongest genetic risk factors for eczema identified to date and a genome-wide association study on eczema had indicated strong linkage disequilibrium between the *FLG* mutations and polymorphisms located up to 500 kb away (Esparza-Gordillo *et al.*, 2009). We therefore included the combined *FLG* mutations as a cofactor in a logistic regression analysis. The effect of rs28989168 on eczema remained significant after adjusting for the combined *FLG* mutations (OR 1.23, 95% CI 1.05–1.44, $P=0.0086$).

DISCUSSION

The EDC in chromosomal region 1q21 harbors >60 genes that are expressed in the epidermis, and many of them are crucial for the formation of an intact skin (Mischke *et al.*, 1996; Marenholz *et al.*, 2001). Recent association and expression studies pointed to genetic risk factors for eczema within the EDC region, apart from the well-established *FLG* locus (Esparza-Gordillo *et al.*, 2009; Guttman-Yassky *et al.*, 2009). In a systematic screen for putatively deleterious mutations in the EDC, we have identified a repeat number variant in *SPRR3* as a, to our knowledge, previously unreported risk factor for eczema.

We used the information present in the NCBI SNP database in order to identify polymorphisms in the EDC genes, which might be disease relevant. Interestingly, out of 20 candidate polymorphisms, only a single one (rs12568784) had an allele frequency specified in any of the populations of the HapMap project. In addition, only 4 of the polymorphisms could be verified in 94 unrelated individuals, although the probability of finding a genetic variant with an allele frequency of 1% in this sample was 85%. We therefore conclude that the majority of polymorphisms in the databases without frequency information are either very rare variants, sequencing artifacts, or they occur only in discrete populations.

For rs12568784, leading to a frameshift in *FLG2*, the minor allele frequency in the discovery set was 15.1%, which was consistent with the frequency in the CEU HapMap population of Central European origin. Although the frameshift mutation in *S100A3* (rs11390146) was a rare variant (<1%), the stop codon mutation in *LCE1D* (rs41268500), the 24-bp deletion in *SPRR3* (rs28989168), as well as rs12568784 in *FLG2* were common polymorphisms with minor allele frequencies between 6 and 50%. According to the different allele frequencies, the power of the discovery set to detect a moderate RR of 1.5 for eczema ranged from 11% for rs11390146 to over 97% for rs12568784 and rs28989168. Consequently, the negative result for rs11390146 (*S100A3*) does not exclude an association of this SNP with eczema. In contrast, an effect of rs41268500 (*LCE1D*) or rs12568784 (*FLG2*) on eczema is unlikely. In particular for *FLG2*, this is a remarkable finding, as the closely related *FLG* is the strongest genetic risk factor for eczema identified to date (Marenholz *et al.*, 2006; Barker *et al.*, 2007). Although *FLG2* has the same structural organization as *FLG*, a similar expression pattern late in epidermal differentiation, and comparable keratin bundling properties (Listwan and Rothnagel, 2004; Wu *et al.*, 2009), the consequences of filaggrin 2 dysfunction seem to be different from filaggrin.

The association of the *SPRR3* risk allele with eczema was confirmed with a similar effect size in a large replication set comprising 1,314 children with early-onset eczema from Central Europe and 1,322 population-based controls, indicating a moderate effect on disease susceptibility. The rs28989168 polymorphism is located at position 195 of the coding region in exon 3, in the repeat domain of *SPRR3*. Four different alleles (-/C/T/CTGTACCAAGGTCCCTGAGCCAGG) were assigned to rs28989168 in the SNP database. Our sequencing results confirmed the presence of two alleles. One allele was identical to the NCBI Build 36.3 reference sequence comprising 14 tandem repeats of 24 bp in exon 3 of *SPRR3*. The other allele lacked one of the sequence repeats and corresponded to the alternative sequence HSA243667 (Figure 1). Accordingly, the encoded protein variant comprises 13 instead of 14 octapeptide repeats in the central domain. Obviously, the repeat structure of exon 3 and in particular identical tandem sequence repeats may lead to ambiguous SNP annotations in the database. Correspondingly, we did not find a frameshift mutation in *SPRR3*, which was reported for rs28989168 in the database.

The eczema-associated variant corresponds to a repeat number variant in *SPRR3* in which the central domain is expanded by one octapeptide repeat. Accordingly, one would expect an effect on protein structure and function

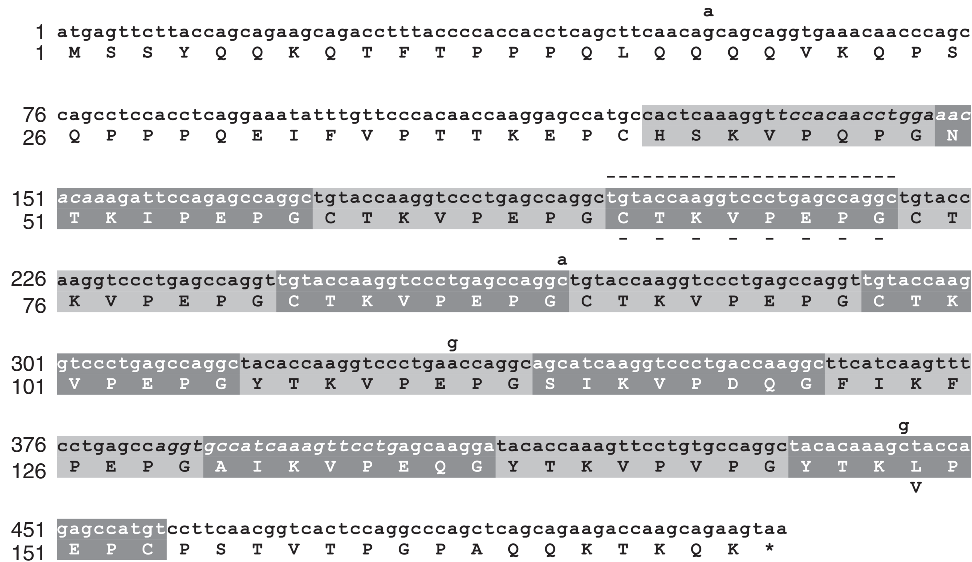

Figure 1. Sequence alignment of the two identified *small proline-rich protein 3* (*SPRR3*) variants. The complete coding sequences of the two *SPRR3* alleles and the predicted protein sequences are indicated. The numbering corresponds to the National Center for Biotechnology Information (NCBI) reference sequence and the reference mRNA NM_001097589, respectively. Tandem sequence repeats are shaded gray. Substitutions of nucleotides on the second *SPRR3* allele, which corresponds to the transcript variant HSA243667, are indicated above the reference sequence, substitutions of amino acids below the amino-acid sequence. The deletion is indicated by a dashed line. Sequences in italics were used to design primers for amplification of the two different alleles.

rather than on the gene expression level. Indeed, in a recent genome-wide mapping study of genetic variants influencing gene expression levels in normal human skin ($n=57$), Ding *et al.* (2010) did not find significant evidence for SNPs influencing *SPRR3* expression. *SPRR3* belongs to a multigene family, the members of which are expressed in terminally differentiating keratinocytes of stratified epithelia (Gibbs *et al.*, 1993; Hohl *et al.*, 1995). SPRRs are crossbridging proteins of the CE (Steinert *et al.*, 1998a) that provides the main barrier function in epithelial tissues. CEs of different epithelia have been shown to vary in their composition of SPRR proteins correlating with the presumed biomechanic requirements of a particular epithelium. Therefore, it was concluded that SPRR proteins serve as modifiers of the tissue-specific physical properties of the CEs (Steinert *et al.*, 1998b). *SPRR3* is expressed in a variety of stratified squamous epithelia, mainly in the mucosa of the oral cavity and esophagus (Hohl *et al.*, 1995). Moreover, it was detected in neoplastic diseases of the skin such as keratoacanthoma and squamous cell carcinoma (Heller-Milev *et al.*, 2000). In genome-wide expression studies on atopic eczema, no difference in *SPRR3* expression was observed between lesional skin and nonlesional or unaffected skin (Sugiura *et al.*, 2005; Esparza-Gordillo *et al.*, 2009; Guttman-Yassky *et al.*, 2009).

Repeat number variants have been described previously for other structural genes of the EDC like involucrin (*IVL*) and *FLG* (Gan *et al.*, 1990; Simon *et al.*, 1991). A phenotype was reported for filaggrin, the degradation products of which serve as natural moisturizing factors of the skin and in which a reduced number of repeats was associated with dry skin (Ginger *et al.*, 2005). The repeat number variation in *SPRR3* (Figure 1) has not been investigated previously, although it affected nearly 50% of the alleles in the study populations. *In vitro* studies have demonstrated that in contrast to the head and tail domains, the central repeat domain of *SPRR3* was not involved in the inter- and intra-molecular crosslinking (Steinert *et al.*, 1999). It was suggested that the central repeats of the SPRRs serve as highly flexible, crossbridging spacers in the CE (Steinert *et al.*, 1998a, 1999). Accordingly, the extra repeat in the eczema-associated variant of SPRR3 might disturb the physical properties of the CE and hence of the barrier.

In summary, we have identified a repeat number variant in *SPRR3* as a, to our knowledge, previously unreported risk factor for eczema and replicated this finding in a large case–control study population. The known function of *SPRR3* as a crossbridging protein of the CE in stratified squamous epithelia may point to an altered barrier function due to the disease-associated variant. Further studies are required in order to elucidate the mechanistic role of *SPRR3* in eczema.

MATERIALS AND METHODS

Study participants

The physician's diagnosis of eczema was made according to standard criteria in the presence of a chronic or chronically relapsing pruritic dermatitis with the typical morphology and distribution (Hanifin and Rajka, 1980). In all cases, the age of onset of eczema was <2 years. The discovery set comprised 555 eczema patients and 375 controls from Germany. The patients were recruited at Charité, Universitätsmedizin Berlin, and the controls were blood donors recruited at the Technical University Dresden. The replication set comprised 1,314 eczema patients from Germany, Poland, and the Czech Republic, and 1,322 geographically matched controls (Esparza-Gordillo *et al.*, 2009). The patients of the replication set were recruited for the ETAC (Early Treatment of the Atopic Child)

and EPAAC (Early Prevention of Asthma in Atopic Children) trials, two similar randomized, double-blind, placebo-controlled studies on the efficacy of levocetirizine and cetirizine, respectively, in the prevention of asthma (ETAC Study Group, 1998; Hill *et al.*, 2008). The controls were obtained from blood donation programs in Prague, Szczecin, and Dresden, respectively. A potential effect due to population stratification in the replication set has been analyzed and excluded previously (Esparza-Gordillo *et al.*, 2009).

The study was approved by the institutional ethics review boards of the participating centers and followed the Declaration of Helsinki Principles. Written, informed consent was obtained from all patients or their legal guardians.

Genotyping

Genomic DNA was prepared from whole blood using standard methods. Functional polymorphisms in the EDC genes identified in the database were genotyped in 94 randomly selected individuals with eczema of the discovery set. Genotyping was performed by sequencing with the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA), by fluorescence-based semi-automated allele-sizing using the 3730 DNA Sequence Analyzer (Applied Biosystems), or by size-separation of the amplified alleles using agarose gel electrophoresis. All identified variants were confirmed by sequencing. For each of the analyzed polymorphisms, primer sequences and annealing temperature for amplification by PCR as well as the detection method used are indicated in Supplementary Table S3 online. Validated polymorphisms were then genotyped in the whole discovery set and if there was association with eczema, additionally in the replication set. SNPs rs12568784 and rs41268500 created restriction fragment length polymorphisms, which were genotyped with the enzymes *Tru*1I and *Acu*I (Fermentas, St Leon-Rot, Germany), respectively. Genotyping of the four most common loss-of-function mutations in *FLG* (R501X, 2282del4, R2447X, and S3247X) was performed as described previously (Palmer *et al.*, 2006; Sandilands *et al.*, 2007). There was no deviation from Hardy–Weinberg equilibrium for any of the detected polymorphisms.

Statistical analysis

The Cochran–Armitage trend test was used to test for association between a polymorphism and eczema. In order to adjust for a potential effect of the *FLG* mutations, we performed a logistic regression analysis including the combined *FLG* mutations as a cofactor. The significance of the logistic model was expressed as the *P*-value of the likelihood ratio test for the ''full model'' (with risk factor and cofactor included) versus the ''null model'' (with cofactor only). A *P*-value of <0.05 was considered statistically significant. Power calculations were performed under an additive model in the discovery set with the Genetic Power Calculation software (Purcell *et al.*, 2003). We estimated the power for three different effect sizes, by using the allele frequencies identified in the discovery set, a disease prevalence of 15%, and a type I error rate (α) of 0.05.

CONFLICT OF INTEREST

The authors state no conflict of interest.

ACKNOWLEDGMENTS

We thank all the subjects with eczema and their families and physicians for their participation. We thank Iv Gwinner-Bühring, F Kussebi, C Ockeloen, T Schilling, and I Schulz for subject evaluation. We are grateful to C Flachmeier, S Kolberg, M Schwarz, and I Szangolies for technical assistance. We thank L Ghys and M De Longueville (UCB Pharma) and the following EPAAC and ETAC investigators for providing case samples: O Rybnícek (Brno, Czech Republic), J Chládková, T Chyba (Hradec Kralove, Czech Republic), F Kopriva (Olomouc, Czech Republic), O Škopková, V Balcárek (Ostrava, Czech Republic), H Honomichlova, P Honomichlová-Houdkova, P Honomichl (Plzen Lochotin, Czech Republic), M Špicáková, S Kynclova, V Petru, A Carbolova, K Kopecká, P Pohunek, T Svobodova, V Špicák, K Kopecka, M Maltulka (Praha, Czech Republic), M Kaczmarski, B Cudowska, J Wasilewska, E Matuszfewska (Bialystok, Poland), T Malaczynska, B Klajna-Kraluk (Gdansk, Poland), J Sokalska, E Masnica-Wasylkowska, I Biegun-Awramineko, M Chlon (Gliwice Poland), T Latos, E Pomaranska, G Gaszczyk, B Makuch, J Stanisz, J Bokiej (Karpacz, Poland), G Lis, E Cichocka-Jarosz, I Glodzik, T Szczerbinski (Kraków, Poland), D Chlebna-Sokol, A Stanczyk, J Wlazlowski, I Ligenza, B Kamer, K Pyziak, R Pasowska, J Zwaigzne-Raczynska (Lodz, Poland), A Emeryk, H Milanowska, G Zywicka, M Bartkowiak-Emeryk, E Chojna (Lublin, Poland), A Swiatly, A Szczawinska-Poplonyk, M Kulesza-Kazecka, J Alkiewicz, A Breborowicz (Poznan, Poland), R Kurzawa, A Wojcik, U Jedrys-Klucjasz, E Urbanek-Jozwik (Rabka Zdrój, Poland), L Dymek, A Dymek, A Bozek (Strzelce Opolskie, Poland), D Chmielewska-Szewczyk, J Lange, J Peradzynska, M Kulus, E Najberg, E Nowicka, A Chrupek (Warszawa, Poland), A Boznanski, E Willak-Janc, A Latkowska, E Sikorska, and MT Jarlinska (Wroclaw, Poland). The study was funded by the German Ministry of Education and Research (BMBF) through the Clinical Research Group for Allergy at Charité Berlin and the National Genome Research Network (NGFN). This work was supported by an EMBO long-term fellowship (522–2006) and by a Marie Curie EIF fellowship (039868) to JEG. This work was done in Berlin, Germany. This work was supported by MZ0FNM2005 to MM.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at http://www.nature.com/jid

REFERENCES

Barker JN, Palmer CN, Zhao Y *et al.* (2007) Null mutations in the filaggrin gene (FLG) determine major susceptibility to early-onset atopic dermatitis that persists into adulthood. *J Invest Dermatol* 127:564–7

Ding J, Gudjonsson JE, Liang L *et al.* (2010) Gene expression in skin and lymphoblastoid cells: refined statistical method reveals extensive overlap in cis-eQTL signals. *Am J Hum Genet* 87:779–89

Esparza-Gordillo J, Weidinger S, Folster-Holst R *et al.* (2009) A common variant on chromosome 11q13 is associated with atopic dermatitis. *Nat Genet* 41:596–601

ETAC Study Group (1998) Allergic factors associated with the development of asthma and the influence of cetirizine in a double-blind, randomised, placebo-controlled trial: first results of ETAC. Early Treatment of the Atopic Child. *Pediatr Allergy Immunol* 9:116–24

Gan SQ, McBride OW, Idler WW *et al.* (1990) Organization, structure, and polymorphisms of the human profilaggrin gene. *Biochemistry* 29: 9432–40

Gibbs S, Fijneman R, Wiegant J *et al.* (1993) Molecular characterization and evolution of the SPRR family of keratinocyte differentiation markers encoding small proline-rich proteins. *Genomics* 16:630–7

Ginger RS, Blachford S, Rowland J *et al.* (2005) Filaggrin repeat number polymorphism is associated with a dry skin phenotype. *Arch Dermatol Res* 297:235–41

Guttman-Yassky E, Suarez-Farinas M, Chiricozzi A *et al.* (2009) Broad defects in epidermal cornification in atopic dermatitis identified through genomic analysis. *J Allergy Clin Immunol* 124:1235–44

Hanifin JM, Rajka G (1980) Diagnostic features of atopic dermatitis. *Acta Derm (Stockholm)* 92(Suppl):44–7

Heller-Milev M, Huber M, Panizzon R *et al.* (2000) Expression of small proline rich proteins in neoplastic and inflammatory skin diseases. *Br J Dermatol* 143:733–40

Hill DJ, Hosking CS, de Benedictis FM *et al.* (2008) Confirmation of the association between high levels of immunoglobulin E food sensitization and eczema in infancy: an international study. *Clin Exp Allergy* 38:161–8

Hohl D, de Viragh PA, Amiguet-Barras F *et al.* (1995) The small proline-rich proteins constitute a multigene family of differentially regulated cornified cell envelope precursor proteins. *J Invest Dermatol* 104:902–9

Lichtenstein P, Svartengren M (1997) Genes, environments, and sex: factors of importance in atopic diseases in 7–9-year-old Swedish twins. *Allergy* 52:1079–86

Listwan P, Rothnagel JA (2004) Keratin bundling proteins. *Methods Cell Biol* 78:817–27

Marenholz I, Heizmann CW, Fritz G (2004) S100 proteins in mouse and man: from evolution to function and pathology (including an update of the nomenclature). *Biochem Biophys Res Commun* 322:1111–22

Marenholz I, Nickel R, Ruschendorf F *et al.* (2006) Filaggrin loss-of-function mutations predispose to phenotypes involved in the atopic march. *J Allergy Clin Immunol* 118:866–71

Marenholz I, Zirra M, Fischer DF *et al.* (2001) Identification of human epidermal differentiation complex (EDC)-encoded genes by subtractive hybridization of entire YACs to a gridded keratinocyte cDNA library. *Genome Res* 11:341–55

Marshall D, Hardman MJ, Nield KM *et al.* (2001) Differentially expressed late constituents of the epidermal cornified envelope. *Proc Natl Acad Sci USA* 98:13031–6

Mischke D, Korge BP, Marenholz I *et al.* (1996) 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* 106:989–92

Palmer CN, Irvine AD, Terron-Kwiatkowski A *et al.* (2006) Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. *Nat Genet* 38:441–6

Presland RB, Dale BA (2000) Epithelial structural proteins of the skin and oral cavity: function in health and disease. *Crit Rev Oral Biol Med* 11:383–408

Purcell S, Cherny SS, Sham PC (2003) Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 19:149–50

Sandilands A, Terron-Kwiatkowski A, Hull PR *et al.* (2007) Comprehensive analysis of the gene encoding filaggrin uncovers prevalent and rare mutations in ichthyosis vulgaris and atopic eczema. *Nat Genet* 39:650–4

Simon M, Phillips M, Green H (1991) Polymorphism due to variable number of repeats in the human involucrin gene. *Genomics* 9:576–80

Steinert PM, Candi E, Kartasova T *et al.* (1998a) Small proline-rich proteins are cross-bridging proteins in the cornified cell envelopes of stratified squamous epithelia. *J Struct Biol* 122:76–85

Steinert PM, Candi E, Tarcsa E *et al.* (1999) Transglutaminase crosslinking and structural studies of the human small proline rich 3 protein. *Cell Death Differ* 6:916–30

Steinert PM, Kartasova T, Marekov LN (1998b) Biochemical evidence that small proline-rich proteins and trichohyalin function in epithelia by modulation of the biomechanical properties of their cornified cell envelopes. *J Biol Chem* 273:11758–69

Sugiura H, Ebise H, Tazawa T *et al.* (2005) Large-scale DNA microarray analysis of atopic skin lesions shows overexpression of an epidermal differentiation gene cluster in the alternative pathway and lack of protective gene expression in the cornified envelope. *Br J Dermatol* 152:146–9

van Beijsterveldt CE, Boomsma DI (2007) Genetics of parentally reported asthma, eczema and rhinitis in 5-yr-old twins. *Eur Respir J* 29:516–21

Wu Z, Hansmann B, Meyer-Hoffert U *et al.* (2009) Molecular identification and expression analysis of filaggrin-2, a member of the S100 fused-type protein family. *PLoS One* 4:e5227