Analysis of the individual and aggregate genetic contributions of previously identified serine peptidase inhibitor Kazal type 5 *(SPINK5)*, kallikrein-related peptidase 7 *(KLK7)*, and filaggrin *(FLG)* polymorphisms to eczema risk

Stephan Weidinger, MD, PhD,[a,b]* Hansjörg Baurecht, MSc,[b,c]* Stefan Wagenpfeil, PhD,[c,d]* John Henderson, MD,[e] Natalija Novak, MD,[f] Aileen Sandilands, PhD,[g] Huijia Chen, BSc,[g] Elke Rodriguez, PhD,[b] Grainne M. O'Regan, MRCPI,[h] Rosemarie Watson, MD,[h] Haihui Liao, MD,[g] Yiwei Zhao, MD,[g] Jonathan N. W. N. Barker, MD,[i] Michael Allen, PhD,[i] Nick Reynolds, MD,[j] Simon Meggitt, MD,[j] Kate Northstone, PhD,[k] George D. Smith, DSc,[l] Carolin Strobl, MSc,[m] Caroline Stahl, MSc,[a,c] Thomas Kneib, PhD,[m] Norman Klopp, PhD,[n] Thomas Bieber, MD,[e] Heidrun Behrendt, MD,[b] Colin N. A. Palmer, PhD,[o] H.-Erich Wichmann, MD,[n] Johannes Ring, MD,[a,d] Thomas Illig, PhD,[n] W. H. Irwin McLean, PhD, DSc, FRSE,[g]* and Alan D. Irvine, MD[h,p]* *Munich and Bonn, Germany, Bristol, Dundee, London, and Newcastle, United Kingdom and Dublin, Ireland*

Background: Polymorphisms in the serine protease inhibitor gene serine peptidase inhibitor Kazal type 5 *(SPINK5)* and the serine protease kallikrein-related peptidase 7 gene *(KLK7)* appear to confer risk to eczema in some cohorts, but these findings have not been widely replicated. These genes encode proteins thought to be involved in the regulation of posttranslation processing of filaggrin *(FLG)*, the strongest identified genetic risk factor for eczema to date.

Objectives: We sought to clarify the individual risk of eczema conferred by the *SPINK5* polymorphism rs2303067 (Glu420Lys) and a previously described insertion in the 3′ untranslated region of *KLK7* and to examine potential epistatic effects between these variants and *FLG* mutations.

Methods: Initially, we examined the effects of these polymorphisms and *FLG* in 486 unrelated patients from a German family-based study, an additional 287 German patients, and 418 unrelated Irish/English patients with eczema (n for 3 genes studied = 1191 vs 4544 control subjects). We then additionally studied the *SPINK5* polymorphism and *FLG* mutations in 1583 patients with eczema from the Avon Longitudinal Study of Parents and Children cohort (sample size for 2 genes studied = 2774 vs 10,607 control subjects).

Results: No association was seen with the *SPINK5* or *KLK7* variants in the case-control analysis; however, a weaker effect was observed for the *SPINK5* variant with maternal transmission in the family-based study. No interactions were seen between the polymorphisms in *KLK7*, *SPINK5*, and *FLG*.

Conclusion: The *SPINK5* 420LysSer mutation confers a risk of eczema when maternally inherited but is not a major eczema

From [a]the Department of Dermatology and Allergy Biederstein, Technical University Munich; [b]the Division of Environmental Dermatology and Allergy, Helmholtz Zentrum Munich–German Research Center for Environmental Health and ZAUM–Center for Allergy and Environment, Technical University Munich; [c]the Institute for Medical Statistics and Epidemiology IMSE, Technical University Munich; [d]the Graduate School of Information Science in Health (GSISH), Technical University Munich; [e]the Department of Community-Based Medicine, University of Bristol; [f]the Department of Dermatology and Allergy, University of Bonn; [g]the Epithelial Genetics Group, Human Genetics Unit, Division of Pathology and Neuroscience, University of Dundee; [h]the Department of Paediatric Dermatology, Our Lady's Children's Hospital, Crumlin, Dublin; [i]the St John's Institute of Dermatology, Guys Hospital, London; [j]the Department of Dermatology, University of Newcastle; [k]the Department of Social Medicine, University of Bristol; [l]MRC Centre for Causal Analyses in Translational Epidemiology, University of Bristol; [m]the Department for Statistics, Ludwig-Maximilans University, Munich; [n]the Department of Epidemiology, Helmholtz Zentrum Munich–German Research Center for Environmental Health; [o]the Population Pharmacogenetics Group, Biomedical Research Centre, University of Dundee; and [p]the Department of Clinical Medicine, Trinity College Dublin.

*These authors contributed equally to this work.

Supported by the German Ministry of Education and Research (BMBF) as part of the National Genome Research Network (NGFN) with grant NUW-S31T05. S. Weidinger and S. Wagenpfeil are supported by research grants KKF-07/04 and KKF-27/05 of the University Hospital "Rechts der Isar," Technical University Munich. In addition, Dr Weidinger is supported by a grant from the "Wilhelm-Vaillant Stiftung." The UK Medical Research Council, the Wellcome Trust, and the University of Bristol provide core support for Avon Longitudinal Study of Parents and Children. Additional funding was obtained from the Department of Paediatric Dermatology, Our Lady's Children's Hospital, Dublin, and by a grant from Tenovus (Tayside) to C.N.P. C.N.P. is supported by the Chief Scientists Office of the Scottish Executive Generation Scotland Initiative. The McLean laboratory is supported by grants from British Skin Foundation/National Eczema Association, the Pachyonychia Congenita Project, the Dystrophic Epidermolysis Bullosa Research Association, the Medical Research Council (reference no. G0700314), and anonymous donations from atopic families in the Tayside region of Scotland.

Disclosure of potential conflict of interest: N. Reynolds has served as an investigator in an investigator-initiated study for Stiefel, UK, and as an expert witness for Newcastle University on antipsoriatic therapy in research. J. Ring has received research support from Novartis, Schering-Plough, Fujisawa, GlaxoSmithKline, Bencard, Stallergenes, ALK-Abelló, Allergopharma, Pharmacia, DPC Biermann, Aventis, Almirall, Leo, Galderma, and Switch Biotech and has served as president of Allergopharma. A. D. Irvine has received funding from the Children's Medical and Research Foundation. The rest of the authors have declared that they have no conflict of interest.

Received for publication March 13, 2008; revised April 28, 2008; accepted for publication May 23, 2008.

Reprint requests: Stephan Weidinger, MD, PhD, Department of Dermatology and Allergy Biederstein, Technical University Munich (TUM), Munich 80802, Germany (or Alan D. Irvine, MD, Paediatric Dermatology, Our Lady's Children's Hospital, Crumlin, Dublin 12, Ireland). E-mail: weidinger@lrz.tum.de or irvinea@tcd.ie.

0091-6749/$34.00

doi:10.1016/j.jaci.2008.05.050

risk factor. The *KLK7* insertion appears to confer no risk of eczema. We found no interaction between the *SPINK5* risk allele or the putative *KLK7* risk allele and *FLG* mutations. (J Allergy Clin Immunol 2008;122:560-8.)

Key words: *Eczema, atopy, skin barrier, stratum corneum, epistasis*

Eczema is a common chronic inflammatory skin disease with a complex, multifactorial cause and a strong genetic component.[1,2] Like other complex diseases, eczema is hypothesized to be determined by many genetic factors interacting with environmental components.[2] In the most commonly accepted paradigm for complex diseases, single genetic factors are considered to contribute only a modest amount to the total variation in the trait but are likely to exert additive or synergistic effects known as epistatic interactions.[3]

The identification of 2 common (R501X and 2282del4) and several rare mutations within the filaggrin gene *(FLG)* causing a deficiency of this key protein involved in skin barrier function delineated a major genetic risk for eczema.[4-6] Subsequently, an impressive series of replication studies[6-20] confirmed that these polymorphisms confer an exceptionally strong risk for eczema and subsequent allergen sensitization and that *FLG* is one of the strongest known genes for complex diseases in general.[21-24]

These observations suggest that the breakdown of the epidermal barrier represents one of the primary events in the development of eczema. This breach might then allow increased penetration of antigens, allergens, and irritants from the environment and thereby predispose to allergic sensitization and aberrant responses to microbial infection.[25-27]

Filaggrin is initially synthesized as biologically inactive profilaggrin, which is expressed as a highly phosphorylated insoluble protein in the granular layer of the epidermis. During the transition from granular cells to flattened squames, profilaggrin is processed to biologically active filaggrin monomers through several dephosphorylation and proteolytic steps,[28,29] the impairment of which might also impair skin barrier function. One of the proteases that has been suggested to be implicated in profilaggrin processing is the stratum corneum chymotryptic enzyme (SSCE),[30,31] which is possibly regulated by the serine protease inhibitor lymphoepithelial Kazal-type inhibitor (LEKTI), encoded by the serine peptidase inhibitor Kazal type 5 gene *(SPINK5)*.[31-33] Interestingly, an insertion in the 3′ untranslated region (UTR) of the kallikrein-related peptidase 7 gene *(KLK7)* encoding SCCE[34] has been reported to be associated with eczema. Early genome-wide linkage analysis of eczema family studies suggested a potential locus on 5q31, and after identification of 6 common polymorphisms in *SPINK5*, the variant Lys420Ser was associated with eczema in a cohort of British children.[35] This association has been replicated in 2 small Japanese studies,[36,37] but other studies have failed to replicate this association.

However, whereas *FLG* has been firmly established as a major gene for eczema, the reported effects of *KLK7* and *SPINK5* variants are rather weak and thus far lack robust confirmation in replication cohort studies. Therefore the aim of the present study was to address the existing literature and to clarify the role of these previously reported polymorphisms in *SPINK5* or *KLK7* in eczema. In addition, given their potential effects on posttranslational modification of filaggrin, we also sought to examine gene-gene interactions between *FLG*, *KLK7*, and *SPINK5*.

Abbreviations used

AIC:	Akaike Information Criterion
ALSPAC:	Avon Longitudinal Study of Parents and Children
FLG:	Filaggrin gene
KLK7:	Kallikrein-related peptidase 7 gene
KORA:	Co-operative Health Research in the Region of Augsburg
LEKTI:	Lymphoepithelial Kazal-type inhibitor
OR:	Odds ratio
SNP:	Single nucleotide polymorphism
SPINK5:	Serine peptidase inhibitor, Kazal type 5 gene
SSCE:	Stratum corneum chymotryptic enzyme
UK:	United Kingdom
UTR:	Untranslated region

METHODS

Study populations

SPINK5, *KLK7*, and *FLG* variants were typed in a cohort of 486 German parent–offspring trios for eczema, a collection of 418 English and Irish patients with eczema and 552 Irish blood donor control subjects, an additional series of 287 patients with eczema from Germany, and the population-based cross-sectional Co-operative Health Research in the Region of Augsburg (KORA) S4 cohort (n = 3992). In addition, the population-based Avon Longitudinal Study of Parents and Children (ALSPAC) cohort (n = 7646) was typed for the Glu420Lys *SPINK5* polymorphism and the 2 most common *FLG* mutations, R501X and 2282del4. Finally, to increase power, we performed a pooled analysis on all available data from all cohorts (for details on study populations, see Table E1 in this article's Online Repository at www.jacionline.org).

The study designs have been described in detail elsewhere.[9,18,38] Briefly, KORA S4 represents a sample of the general adult population of German nationality in the region of Augsburg recruited from October 1999 to April 2001. The survey comprised 4261 unrelated men and women between 25 and 74 years of age. All subjects had to complete a standardized questionnaire that, in addition to demographic data, included the basic allergy questions of the European Community Respiratory Health Survey on respiratory health.[39] All individuals received a skin examination by experienced senior dermatologists who had been additionally trained before the start of the study, according to the criteria of Hanifin and Rajka[40] and the United Kingdom (UK) diagnostic criteria for eczema.[41]

All German patients with eczema were unrelated and of white origin, with eczema diagnosed on the basis of a skin examination by experienced dermatologists using the UK diagnostic criteria.[41] In the family collection 10.4% of the parents had eczema (9.3% affected fathers and 11.1% affected mothers).

Patients with eczema from Ireland were recruited through attendance at a hospital-based clinic in Our Lady's Children's Hospital Crumlin, and the diagnosis was made according to the UK diagnostic guidelines by an experienced pediatric dermatologist (ADI, GO'R, or RW). The English eczema cohort was recruited from hospital-based clinics in London and Newcastle and has been described previously.[11] A summary of the demographics for all eczema study populations examined is presented in Table E1.

The ALSPAC is a longitudinal, population-based birth cohort study that recruited 14,541 unrelated pregnant women resident in Avon, UK, with expected dates of delivery between April 1, 1991, and December 31, 1992. There were 14,062 liveborn children. The study protocol has been described previously,[42,43] and further details are on the ALSPAC Web site (http://www.alspac.bris.ac.uk). At 6, 18, 30, and 42 months of age, the mothers were asked whether their children had skin rashes in the joints or creases of the body. As in previous studies, we defined individuals with eczema as those with reports of flexural dermatitis at 2 time points between 6 and 42 months.[25,42,43]

All study methods were approved by the relevant local authorities, and written informed consent that complies with all the Declaration of Helsinki Principles was obtained from all participants.

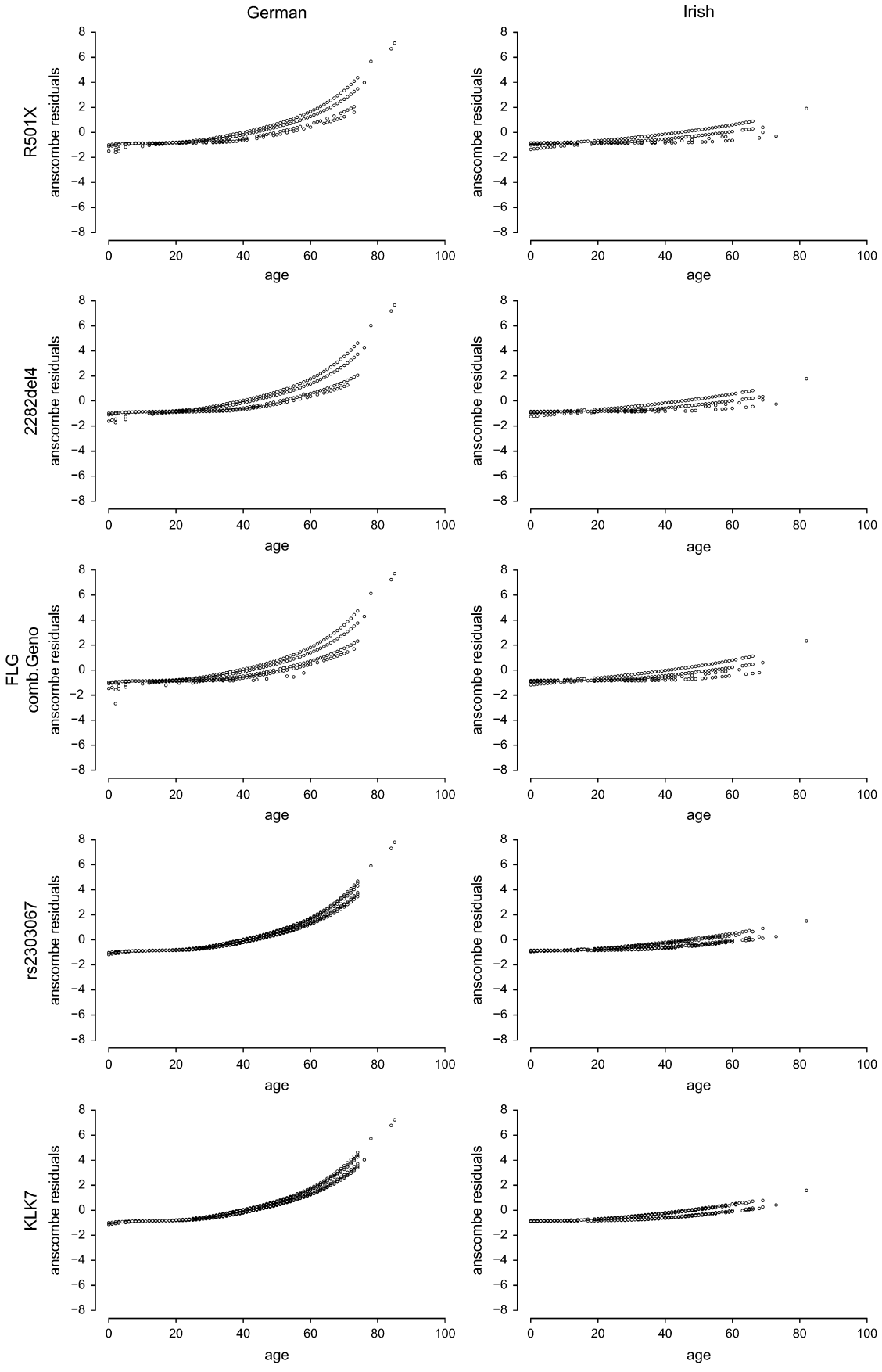

FIG 1. *(Continued).*

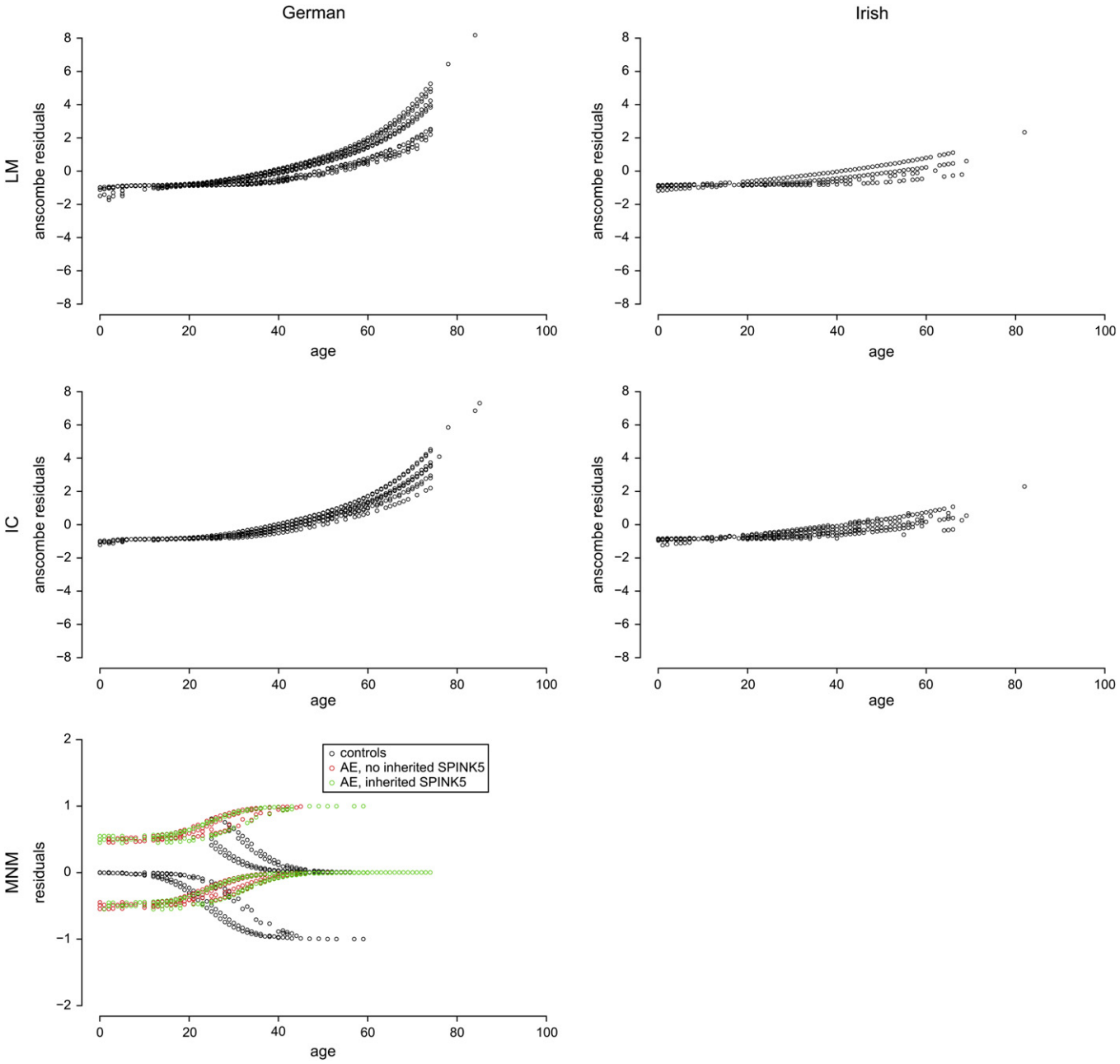

FIG 2. Residual plot of age after fitting a general linear model for SNP-SNP interaction analyses. The Anscombe residuals for LM and IC and the ordinary residuals for MNM show a functional form and do not spread randomly. Hence a general additive model is fitted to the data. The *columns* indicate the population's origin of the case-control studies, and the *rows* denote the applied methods. *LM*, Logistic regression model with product-interaction terms; *IC*, logistic regression model with interaction score; *MNM*, multinomial logistic model considering maternal inheritance.

Genotyping

Genotyping in German samples was performed with the MassARRAY system (Sequenom, San Diego, Calif), as described previously.[9] Genotyping calls were made in real time with MASSARRAY RT software (Sequenom). Primers, as well as allele frequencies, in the population-based KORA S4 cohort (n = 4198) are shown Table E2 (available in this article's Online Repository at www.jacionline.org).

The Glu420Lys polymorphism in *SPINK5* (rs2303067) was typed in the Irish and English eczema cohorts and control subjects by using a predesigned single nucleotide polymorphism (SNP) Taqman Genotyping Assay from Applied Biosystems (Foster City, Calif; product C_2000249_10) and run on a 7900HT Fast Real-Time PCR system using the manufacturer's recommended protocol. The 4-bp insertion in the 3′UTR of *KLK7* was typed in Irish and English eczema cohorts and control subjects by sizing of fluorescently labeled

FIG 1. Residual plot of age after fitting a general linear model for individual SNP analyses. The Anscombe residuals show a functional form and do not spread randomly. Hence a general additive model is fitted for the data. comb.Geno, Combined genotype.

TABLE I. ORs and 95% CIs for associations between polymorphisms and eczema in the family-based analysis

	Gene	Polymorphism	T:U	OR	95% CI		*P* value
German families (A)	*FLG*	r501x	41:12	3.42	1.84	6.70	1.2×10^{-4}
		2282del4	72:31	2.32	1.54	3.57	8.1×10^{-5}
		Combined genotype	110:40	2.75	1.93	3.98	1.8×10^{-8}
	SPINK5	rs2303067	259:207	1.25	1.04	1.50	.01815
	KLK7	AACC ins	183:177	1.03	0.84	1.27	.79215

T, Transmitted; *U*, untransmitted; *ins*, insertion.

TABLE II. Associations between polymorphisms and eczema in the case-control approaches

	Gene	Polymorphism	OR	95% CI		*P*-value
German (A, B, E)	*FLG*	Combined genotype (het vs wt)	4.15	3.10	5.56	1.3×10^{-21}
Cases: 773		Combined genotype (hom vs wt)	23.11	7.23	73.89	1.2×10^{-7}
Control subjects: 3992	*SPINK5*	rs2303067 (het vs wt)	1.14	0.87	1.49	.34285
Total: 4765		rs2303067 (hom vs wt)	1.22	0.89	1.67	.21252
	KLK7	AACC ins (het vs wt)	1.11	0.89	1.40	.35472
		AACC ins (hom vs wt)	0.95	0.64	1.42	.81033
Irish/English (C, F)	*FLG*	Combined genotype (het vs wt)	4.34	2.77	6.79	$1.3 \times 10-10$
Cases: 418		Combined genotype (hom vs wt)	2.6×10^{56}	0	∞	1.0
Control subjects: 552	*SPINK5*	rs2303067 (het vs wt)	0.78	0.52	1.18	.23945
Total: 970		rs2303067 (hom vs wt)	1.15	0.71	1.87	.57398
	KLK7	(het vs wt)	1.29	0.89	1.86	.17756
		(hom vs wt)	0.91	0.47	1.76	.76988
ALSPAC (D, G)	*FLG*	Combined genotype (het vs wt)	2.17	1.81	2.60	3.3×10^{6}
Cases: 1583		Combined genotype (hom vs wt)	3.3×106	0	∞	.94524
Control subjects: 6063	*SPINK5*	rs2303067 (het vs wt)	0.96	0.84	1.09	.52176
Total: 7646		rs2303067 (hom vs wt)	1.14	0.98	1.34	.09156
Pooled* (A, B, C, D, E, F, G)	*FLG*	Combined genotype (het vs wt)	3.04	2.68	3.44	$8.5 \times 10-68$
Cases: 2774		Combined genotype (hom vs wt)	49.38	19.72	123.61	$8.2 \times 10-17$
Control subjects: 10,607	*SPINK5*	rs2303067 (het vs wt)	0.97	0.87	1.07	.50850
Total: 13,381		rs2303067 (hom vs wt)	1.13	1.00	1.27	.04545

het, Heterozygous; *wt*, wild-type; *hom*, homozygous.
All logistic regression models are adjusted for age and sex. Every block refers to a single model in which only estimates of the genetic variables are displayed. No specific genetic model was assumed, and estimates are given for every genotype compared with the wild-type. Populations as designated in Table E1 are indicated in parentheses.
*For the pooled analyses, we estimated in all models a population effect as a confounder with *P* values of less than 10^{-4}.

PCR products on an Applied Biosystems 3130xl Genetic Analyser. Ten-milliliter PCR reactions were performed with 25 ng of genomic DNA with 400 nmol/L forward primer (5′ gtt tct tca agt gtg caa gtt cac caa 3′) and 400 nmol/L FAM-labeled reverse primer (5′ GAT TGG TTT ATC AAC AGG GC 3′) in AmpliTaq Gold Buffer containing 1.5 mmol/L $MgCl_2$, 10 nmol of each deoxyribonucleoside triphosphate, 4% vol/vol dimethyl sulfoxide, and 0.25 U of AmpliTaq Gold polymerase (Applied Biosystems). PCR reactions were amplified at an annealing temperature of 58°C. Diluted PCR products were sized against ROX-500 size standards (Applied Biosystems). Allele sizes were 201 and 205 bp (insertion).

For the ALSPAC cohort, the 2 most common *FLG* mutations were typed as previously described,[24] and the *SPINK5* Glu420Lys polymorphism was typed by means of Taqman assay. Because we discovered 3 negative associations for *KLK7* in the eczema cohorts, we did not perform *KLK7* analysis in this very large population cohort.

Statistical analyses

Descriptive statistics for quantitative and qualitative values are presented as means ± SD and relative frequencies or absolute numbers, respectively. Deviation from Hardy-Weinberg equilibrium was tested in parents for the family analyses and in control subjects for the case-control analyses. In the family setting we analyzed association of single SNPs with eczema using the classical transmission disequilibrium test. Parent-of-origin effects were investigated with the method proposed by Weinberg.[44]

Case-control analyses for single SNPs were performed with logistic regression models adjusted for age and sex. To not constrain the analyses to a specific genetic model, we modeled the 3 categorical genotypes by 2 dummy variables.

Gene-gene interaction analyses were performed after excluding individuals with 2 mutant *FLG* alleles because these individuals do not express the filaggrin protein, and therefore a biologic interaction with posttranslational modifications by SSCE and LEKTI is not plausible.

Four different approaches were carried out to estimate interaction effects between the single polymorphisms in *FLG*, *KLK7*, and *SPINK5*. In any of these approaches, we adjusted the models for the common covariates age and sex.

First, interaction was evaluated by using the logistic regression model with product-interaction terms. The Akaike Information Criterion (AIC) was used to select the appropriate model.[45]

Second, we defined an interaction score that counts the number of copies of the potentially disease-associated alleles,[46] which we used as a covariate in the logistic model.

Third, we modeled maternal effects observed for *SPINK5* using affected offspring from families only and estimated odds ratios (ORs) for 2 different affection statuses compared with control subjects in a multinomial regression model, which was carried out with BayesX 1.50.[47] An elaborate description of the statistical methods is given in this article's Methods section of the Online Repository (available at www.jacionline.org). For any of these regression approaches, the quantitative covariate age was modeled nonparametrically in a general additive model framework because residuals indicate structure in general linear models[48] (see Figs 1 and 2).

Finally, for further exploration, variable importance measures were computed by means of the random forest method. Random forests provide variable importance measures that can be used to detect variables relevant for

TABLE III. Associations between polymorphisms or gene-gene interaction terms and eczema

German						
LRM (A, B, E)	**Gene-gene interaction term**		**OR**	**95% CI**		***P* value**
	FLG combined genotype (het vs wt)		6.11	4.09	9.12	3.3×10^{-10}
	rs2303067 (het vs wt)		1.28	0.87	1.90	.11862
	rs2303067 (hom vs wt)		1.28	0.83	1.97	.18638
	FLG combined genotype (het) rs2303067 (het)		0.62	0.34	1.11	.17805
	FLG combined genotype (het) rs2303067 (hom)		0.61	0.06	6.20	.23915
IC (A, B, E)	Interaction score*	1 vs 0	1.08	0.72	1.61	.71881
		2 vs 0	1.18	0.80	1.76	.39974
		3 vs 0	1.67	1.08	2.57	.01982
		4 vs 0	2.75	1.53	4.92	.00067
		5 vs 0	1.45	0.14	14.70	.75107
MNM (A, E)	**Multinomial logit trait**	**Covariates**	**OR**	**95% CI**		***P* value**
	Response categories: 0: control subjects 1: eczema and *SPINK5* (not maternal) 2: eczema and *SPINK5* (maternal)	1: sex, F vs M	1.69	1.12	2.56	.01338
		1: *FLG* combined genotype	5.22	3.12	8.73	4.9 × 10−7
		2: sex, F vs M	1.32	0.87	2.00	.18653
		2: *FLG* combined genotype	6.16	3.70	10.26	1.1 × 10−7
	Effect of *FLG* combined genotype in response category 1 compared to category 2					.80071
Irish						
LRM (C, F)	**Gene-gene interaction term**		**OR**	**95% CI**		***P* value**
IC (C, F)	*FLG* combined genotype (het vs wt)		4.34	2.77	6.79	1.3×10^{-10}
	Interaction score*	1 vs 0	1.32	0.64	2.75	.45470
		2 vs 0	1.18	0.56	2.46	.66276
		3 vs 0	1.96	0.92	4.20	.08175
		4 vs 0	5.17	1.88	14.24	.00148
		5 vs 0	0.01	~0	2.6×10^{10}	.74599
ALSPAC						
LRM (D, G)	**Gene-gene interaction term**		**OR**	**95% CI**		***P* value**
	FLG combined genotype (het vs wt)		2.26	1.63	3.14	1.2×10^{-6}
	rs2303067 (het vs wt)		1.02	0.87	1.19	.83390
	rs2303067 (hom vs wt)		1.18	0.98	1.41	.07389
	FLG combined genotype (het) rs2303067 (het)		0.90	0.59	1.38	.63859
	FLG combined genotype (het) rs2303067 (hom)		1.04	0.62	1.74	.87865
IC (D, G)	Interaction score*	1 vs 0	1.06	0.92	1.23	.39907
		2 vs 0	1.29	1.09	1.51	.00245
		3 vs 0	2.72	1.84	4.02	5.4×10^{-7}
Pooled†						
LRM (A, B, C, D, E, F, G)	**Gene-gene interaction term**		**OR**	**95% CI**		***P* value**
	FLG combined genotype (het vs wt)		3.43	2.71	4.35	1.7×10^{-24}
	rs2303067 (het vs wt)		1.03	0.92	1.16	.58801
	rs2303067 (hom vs wt)		1.15	1.00	1.32	.04914
	FLG combined genotype (het) rs2303067 (het)		0.81	0.60	1.09	.16864
	FLG combined genotype (het) rs2303067 (hom)		0.96	0.67	1.37	.82037
IC (A, B, C, D, E, F, G)	Interaction score*	1 vs 0	1.10	0.99	1.23	.08116
		2 vs 0	1.37	1.21	1.55	6.5×10^{-7}
		3 vs 0	3.66	2.83	4.75	1.2×10^{-22}

LRM, Logistic regression model with product-interaction terms; *het*, heterozygous; *wt*, wild-type; *hom*, homozygous; *IC*, logistic regression model with interaction score; *MNM*, multinomial logistic model considering maternal inheritance; *F*, female; *M*, male.

All models are adjusted for age and sex. Every block refers to a single model. Populations as designated in Table E1 are indicated in parentheses.

*Score: number of copies of the "risk" alleles in the *FLG* combined genotype, *SPINK5* rs2303067, and *KLK7*. For ALSPAC and pooled analysis, the score is reduced to the number of copies of *FLG* and *SPINK5* alleles.

†For the pooled analyses in all models, we estimated a population effect as a confounder with *P* values of less than 10^{-4}.

TABLE IV. Results for random forests with 2 randomly preselected variables in each split

	German (A, B, E)		Irish (C, F)		ALSPAC (D, G)		Pooled (A, B, C, D, E, F, G)	
RF	**Mean − 2 SE**	**Mean − 2 SE**	**Mean − 2 SE**	**Mean + 2 SE**	**Mean − 2 SE**	**Mean + 2 SE**	**Mean −2 SE**	**Mean + 2 SE**
Age	0.12179	0.12241	0.19797	0.19891	NA	NA	NA	NA
Sex	−0.00037	−0.00032	0.01075	0.01106	−0.00004	−0.00003	−0.00022	−0.00021
FLG combined genotype	0.00465	0.00473	0.05204	0.05253	0.00001	0.00002	0.00468	0.00474
SPINK5	−0.00021	−0.00016	0.00040	0.00059	−0.00003	−0.00002	−0.00024	−0.00022
KLK7	−0.00015	−0.00011	−0.00112	−0.00097	NA	NA	NA	NA
Population	NA	NA	NA	NA	NA	NA	−0.00008	−0.00006

RF, Random forest; *NA*, not available.

The average permutation importance ± 2 SEM over 100 iterations is displayed for each variable. Results for random forests with 5 randomly preselected variables (ie, for bagging) were almost identical. High positive values indicate a high variable importance. Small positive or negative values indicate that a variable is irrelevant for predicting the response. Populations as designated in Table I are indicated in parentheses.

predicting the response. The most commonly used variable importance measure is the permutation importance.[49]

High positive values of the importance measure obtained indicate a high variable importance. Small positive or negative values indicate that a variable is irrelevant for predicting the response.

Additionally, we pooled all 3 study cohorts to increase the power of detecting any single SNP and interaction effect. Because British and German populations might be slightly different with regard to ethnicity,[17] we accounted for a population effect in every analysis approach by introducing a binary independent variable, which codes as follows: 1, British/Irish origin; 0, German origin. Thus we corrected the estimated genetic effect for potential population differences.

Power calculations[50] for the pooled single SNP analyses were performed with nQuery7.0, assuming a dominant model.

All statistical analyses were carried out with R 2.6.0 (R Foundation for Statistical Computing, Vienna, Austria),[51] unless otherwise stated.

RESULTS

Single-gene analyses

First we examined the effect of the individual polymorphisms in predisposition to eczema in our samples. Allele frequencies are presented in Table E3. The *SPINK5* and *KLK7* polymorphisms showed comparable allele frequencies across all study populations. *FLG* polymorphism results have been published for the German family study previously,[9,17] although for this study, further families and patients were tested to increase statistical power, in particular when looking for interaction between alleles. In the German family cohort *FLG* polymorphisms greatly increased the risk for eczema (OR, 2.75; 95% CI, 1.93-3.98; $P = 1.8 \times 10^{-8}$), whereas the *SPINK5* polymorphism rs2303067 showed only a slight overtransmission to eczema-affected offspring (OR, 1.25; 95% CI, 1.04-1.50; $P = .018$). No associations were seen for the *KLK7* 3′UTR insertion (Table I). In this family cohort the power to detect a proportion of 1.2 between transmission of the risk allele versus nontransmission (which corresponds to a difference in the proportions of 5% given the observed number of discordant pairs) was 35% for the *SPINK5* polymorphism and 44% for the *KLK7* insertion. The power to detect a difference in the proportion of 9% for *SPINK5* and of 8% for *KLK7* was 80%.

Because parent-of-origin effects have been reported for *SPINK5* variants,[36] we also tested for differences between maternal and paternal allele sharing, and we confirmed a stronger association for the maternally inherited rs23030067 A allele and a relative risk of transmission of maternal alleles compared with paternal alleles of 2.18 (95% CI, 1.38-3.43; $P = .0008$). The observed power to detect parent-of-origin effects, as described in Weinberg,[45] was greater than 90%. In the pooled German case-control cohort, the presence of at least 1 *FLG* variant greatly increased the risk for eczema (OR, 4.56; 95% CI, 3.44-6.05; $P = 5.9 \times 10^{-26}$). In contrast, neither the *SPINK5* nor the *KLK7* variant was associated with eczema (Table II).

In the Irish case-control cohort comparable results could be seen: presence of an *FLG*-null allele increased the risk for eczema about 5.88-fold (95% CI, 3.85-8.99; $P = 2.8 \times 10^{-16}$). No association between *SPINK5* or *KLK7* and eczema was detected (Table II).

The analyses of the ALSPAC cohort also showed a strong *FLG* effect (OR, 2.23; 95% CI, 1.87-2.67; $P = 1.2 \times 10^{-18}$). Consistent with our other cohorts, no association between the Lys420Ser *SPINK5* polymorphism and eczema was found (Table II).

For the pooled study in all models, a population effect was estimated as a confounder with P values of less than 10^{-4}. An exceptionally strong *FLG* effect on eczema was seen for the pooled analyses (OR, 3.36; 95% CI, 2.97-3.79; $P = 1.3 \times 10^{-84}$), with a power of greater than 99% for the observed proportion of *FLG* variants of 0.21 in the cases. No association was observed for the *SPINK5* variant. The power to detect an increased risk in OR of 1.2 for carriers of the *SPINK5* rare allele compared with noncarriers with an observed proportion of 0.73 in the cases was 95%, assuming a dominant genetic model. For *KLK7*, we had 78% power to detect an OR of 1.2 in the pooled analyses, with an observed minor allele frequency in the cases of 0.56.

Gene-gene interaction analyses

Gene-gene interactions were examined in all 3 cohorts separately and together in a pooled analysis.

The logistic regression model with product-interaction terms approach in the German case-control cohort was the best-fitting model on the basis of minimization of AIC with main effects of *FLG* and rs2303067 *(SPINK5)* and their product-interaction terms. In this model only *FLG* showed a significant genetic effect. In the Irish case-control cohort only *FLG*, along with the covariates age and sex, remained in the model according to the AIC (Table III).

In the interaction score approach, in addition to 1 mutant *FLG* allele, the A allele of rs2303067 *(SPINK5)* and the *KLK7* insertion were defined as "risk" variants. By deriving a score from the numbers of variant copies in the German case-control cohort, we observed a tendency for an increasing risk with the number of risk alleles an individual carried. The decrease in OR in the last category can probably be attributed to the low number of

observations, as reflected by the wide CI. Interestingly, the effect size increases with the number of risk alleles, probably because of the increased chance of *FLG* risk alleles in these cells.

For the Irish/British case-control cohort, similar results were observed, but the OR in the 4-variant group was more than twice as high as in the German case-control analyses. Using the multinomial regression model approach, we tried to account for the maternal parent-of-origin effect reported by Walley et al.[36] In our family collection we observed a tendency for an increasing risk of development of eczema caused by *FLG* mutations (regardless of inheritance of these *FLG* mutations) if the *SPINK5* SNP rs2303067 was inherited from the mother. There was an increased risk of eczema compared with paternal inheritance. However, the null hypothesis for equal *FLG* risk in both response categories could not be rejected.

In the random forest approach (Table IV), in the German sample only age and *FLG* status showed positive variable importance values. The average out-of-bag prediction accuracy was between 87.7% for the random forests and 88.3% for bagging. However, this is due to the fact that the average specificity was close to 100%, whereas the sensitivity was around 43% in the sample with approximately 21% cases. In the Irish case-control cohort only age and, to some extent, *FLG* and sex were suitable for predicting eczema. The average out-of-bag prediction accuracy was between 83.2% for the random forests and 84.4% for bagging.

The pooled analyses revealed results consistent with the previously estimated effects. In all models we estimated a population effect as a confounder with P values of less than 10^{-4}.

DISCUSSION

This large-scale study examined variants in 3 candidate genes that have previously been reported to be associated with eczema. All genes encode proteins that are involved in the highly organized process of epidermal differentiation and are important for the maintenance of skin barrier function. In addition, because of their biologic interactions, we hypothesized that there might also be gene-gene-interactions.

Filaggrin is a key protein for the development of the cornified envelope and the process of cornification.[52] Two common null mutations in the *FLG* gene (R501X and 2282del4) have been firmly established as strong risk factors for eczema.[22,23] *KLK7* encodes the protease SSCE, which has been suggested to be involved in the complex proteolytic processing of filaggrin. An insertion in the 3′UTR of the *KLK7* gene possibly influencing SSCE activity has been reported to be associated with eczema in a UK case-control study, but this association has not been replicated thus far.[14,35] *SPINK5* is the gene defective in Netherton syndrome and encodes the serine proteinase inhibitor LEKTI, which has been implicated in the regulation of SSCE activity. An association of a *SPINK5* SNP with eczema has previously been reported[36] but was not confirmed in a recent study.[53]

Using a large cohort of German families and an Irish/English case-control series, as well as a pooled and enlarged German case-control collection and the longitudinal ALSPAC cohort, we first examined the individual SNPs. Results from these analyses suggest that of the tested polymorphisms, only the *FLG* mutations represent important and replicable genetic determinants for eczema, whereas the *SPINK5* variant Glu420Lys appears to have a weaker effect and only when maternally inherited, and the *KLK7* insertion does not exert an effect. However, our data do not preclude that there are other variants in *SPINK5*, *KLK7*, or both of importance for eczema.

In a second step we aimed at elucidating potential gene-gene interaction. Using several statistical approaches, we found no evidence for an interaction between variants in these 3 genes. The fact that the random forests permutation importance is essentially zero for the *SPINK5* and *KLK7* variants confirms that these variants are not relevant for predicting eczema, neither individually nor in interactions with each other, because interactions would be captured by the random forest variable importance.[54] Because there was a considerable difference in age between patients and control subjects, we considered age as a covariate in all analyses and modeled it nonparametrically. However, this leads to little variability in the response data explained by *FLG* mutations in the random forest approach.

It is widely hypothesized that complex human diseases, such as eczema, result from an unknown number of genetic factors, each of which influences susceptibility through interactions with other genes and with environmental factors.[55,56] With whole-genome association studies with hundreds of thousands of measured genetic variations emerging, analyses of the complex molecular interactions on the DNA level are of utmost importance, and it will be necessary to develop innovative statistical methods. For eczema, this is the first study that directly addresses this issue by exploring the effect of potential interactions among genes encoding proteins in the filaggrin expression and processing pathways. Using diverse and complementary statistical approaches in this large sample, we did not find evidence for epistatic effects between *FLG* and *KLK7* variants that significantly predict eczema risk. Thus although our data underline the exceptional importance of *FLG* deficiency for eczema risk, it does not support the hypothesis that its effect is dependent on or modified by *KLK7*. The results of the pooled analyses and the family analyses give a hint that *SPINK5* might be a potential player (when maternally inherited) within the filaggrin cascade, but this association requires further exploration. However, it cannot be excluded that acquired alterations in filaggrin processing or variations in other genes in the same pathway, such as *KLK5*, might contribute to eczema susceptibility. Functional studies are needed to explore the individual roles of products of genes within the filaggrin pathway and their biologic interactions, and future large-scale studies with powerful statistical methods will aid in elucidating the relationship between combinations of polymorphisms for eczema susceptibility.

We thank all of the patients and families who took part in this study, the professionals who helped in recruiting them, and the KORA and ALSPAC teams, which include interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses.

Key messages

- **The *SPINK5* Lys420Ser polymorphism confers a risk of eczema when maternally inherited but is not a major genetic contributor to eczema risk.**
- **A previously reported association of a *KLK7* insertion and eczema could not be confirmed.**
- **There is no evidence for epistatic effects between *KLK7* or *SPINK5* variants and *FLG* mutations.**

REFERENCES

1. van Beijsterveldt CE, Boomsma DI. Genetics of parentally reported asthma, eczema and rhinitis in 5-yr-old twins. Eur Respir J 2007;29:516-21.
2. Morar N, Willis-Owen SA, Moffatt MF, Cookson WO. The genetics of atopic dermatitis. J Allergy Clin Immunol 2006;118:24-34; quiz 5-6.
3. Glazier AM, Nadeau JH, Aitman TJ. Finding genes that underlie complex traits. Science 2002;298:2345-9.
4. Smith FJ, Irvine AD, Terron-Kwiatkowski A, Sandilands A, Campbell LE, Zhao Y, et al. Loss-of-function mutations in the gene encoding filaggrin cause ichthyosis vulgaris. Nat Genet 2006;38:337-42.
5. Palmer CN, Irvine AD, Terron-Kwiatkowski A, Zhao Y, Liao H, Lee SP, et al. Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. Nat Genet 2006;38:441-6.
6. Sandilands A, O'Regan GM, Liao H, Zhao Y, Terron-Kwiatkowski A, Watson RM, et al. Prevalent and rare mutations in the gene encoding filaggrin cause ichthyosis vulgaris and predispose individuals to atopic dermatitis. J Invest Dermatol 2006;126:1770-5.
7. Ruether A, Stoll M, Schwarz T, Schreiber S, Folster-Holst R. Filaggrin loss-of-function variant contributes to atopic dermatitis risk in the population of Northern Germany. Br J Dermatol 2006;155:1093-4.
8. Marenholz I, Nickel R, Ruschendorf F, Schulz F, Esparza-Gordillo J, Kerscher T, et al. Filaggrin loss-of-function mutations predispose to phenotypes involved in the atopic march. J Allergy Clin Immunol 2006;118:866-71.
9. Weidinger S, Illig T, Baurecht H, Irvine AD, Rodriguez E, Diaz-Lacava A, et al. Loss-of-function variations within the filaggrin gene predispose for atopic dermatitis with allergic sensitizations. J Allergy Clin Immunol 2006;118:214-9.
10. Stemmler S, Parwez Q, Petrasch-Parwez E, Epplen JT, Hoffjan S. Two common loss-of-function mutations within the filaggrin gene predispose for early onset of atopic dermatitis. J Invest Dermatol 2007;127:722-4.
11. 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.
12. Sandilands A, Terron-Kwiatkowski A, Hull PR, O'Regan GM, Clayton TH, Watson RM, et al. Comprehensive analysis of the gene encoding filaggrin uncovers prevalent and rare mutations in ichthyosis vulgaris and atopic eczema. Nat Genet 2007;39:650-4.
13. Hubiche T, Ged C, Benard A, Leaute-Labreze C, McElreavey K, de Verneuil H, et al. Analysis of SPINK 5, KLK 7 and FLG genotypes in a French atopic dermatitis cohort. Acta Derm Venereol 2007;87:499-505.
14. Morar N, Cookson WO, Harper JI, Moffatt MF. Filaggrin mutations in children with severe atopic dermatitis. J Invest Dermatol 2007;127:1667-72.
15. Nomura T, Sandilands A, Akiyama M, Liao H, Evans AT, Sakai K, et al. Unique mutations in the filaggrin gene in Japanese patients with ichthyosis vulgaris and atopic dermatitis. J Allergy Clin Immunol 2007;119:434-40.
16. Rogers AJ, Celedon JC, Lasky-Su JA, Weiss ST, Raby BA. Filaggrin mutations confer susceptibility to atopic dermatitis but not to asthma. J Allergy Clin Immunol 2007;120:1332-7.
17. Weidinger S, Rodriguez E, Stahl C, Wagenpfeil S, Klopp N, Illig T, et al. Filaggrin mutations strongly predispose to early-onset and extrinsic atopic dermatitis. J Invest Dermatol 2007;127:724-6.
18. Nomura T, Akiyama M, Sandilands A, Nemoto-Hasebe I, Sakai K, Nagasaki A, et al. Specific filaggrin mutations cause ichthyosis vulgaris and are significantly associated with atopic dermatitis in Japan. J Invest Dermatol 2008; 128:1436-41.
19. Ekelund E, Lieden A, Link J, Lee SP, D'Amato M, Palmer CN, et al. Loss-of-function variants of the filaggrin gene are associated with atopic eczema and associated phenotypes in Swedish families. Acta Derm Venereol 2008;88:15-9.
20. Brown SJ, Relton CL, Liao H, Zhao Y, Sandilands A, Wilson IJ, et al. Filaggrin null mutations and childhood atopic eczema: a population-based case-control study. J Allergy Clin Immunol 2008;121:940-6.
21. Baurecht H, Irvine AD, Novak N, Illig T, Buhler B, Ring J, et al. Towards a major risk factor for atopic eczema: meta-analysis of filaggrin mutation data. J Allergy Clin Immunol 2007;120:1406-12.
22. Rodriguez E, Illig T, Weidinger S. Filaggrin loss-of-function mutations and association with allergic diseases. Pharmacogenomics 2008;9:399-413.
23. Weidinger S, O'Sullivan M, Illig T, Baurecht H, Depner M, Rodriguez E, et al. Filaggrin mutations, atopic eczema, hay fever, and asthma in children. J Allergy Clin Immunol 2008;121:1203-9.
24. Henderson J, Northstone K, Lee SP, Liao H, Zhao Y, Pembrey M, et al. The burden of disease associated with filaggrin mutations: a population-based, longitudinal birth cohort study. J Allergy Clin Immunol 2008;121:872-7.
25. Hudson TJ. Skin barrier function and allergic risk. Nat Genet 2006;38:399-400.
26. McLean WH, Hull PR. Breach delivery: increased solute uptake points to a defective skin barrier in atopic dermatitis. J Invest Dermatol 2007;127:8-10.
27. Sicherer SH, Leung DY. Advances in allergic skin disease, anaphylaxis, and hypersensitivity reactions to foods, drugs, and insects. J Allergy Clin Immunol 2007;119: 1462-9.
28. Resing KA, Walsh KA, Dale BA. Identification of two intermediates during processing of profilaggrin to filaggrin in neonatal mouse epidermis. J Cell Biol 1984;99:1372-8.
29. Resing KA, Walsh KA, Haugen-Scofield J, Dale BA. Identification of proteolytic cleavage sites in the conversion of profilaggrin to filaggrin in mammalian epidermis. J Biol Chem 1989;264:1837-45.
30. Resing KA, Thulin C, Whiting K, al-Alawi N, Mostad S. Characterization of profilaggrin endoproteinase 1. A regulated cytoplasmic endoproteinase of epidermis. J Biol Chem 1995;270:28193-8.
31. Descargues P, Deraison C, Bonnart C, Kreft M, Kishibe M, Ishida-Yamamoto A, et al. Spink5-deficient mice mimic Netherton syndrome through degradation of desmoglein 1 by epidermal protease hyperactivity. Nat Genet 2005;37:56-65.
32. Komatsu N, Takata M, Otsuki N, Ohka R, Amano O, Takehara K, et al. Elevated stratum corneum hydrolytic activity in Netherton syndrome suggests an inhibitory regulation of desquamation by SPINK5-derived peptides. J Invest Dermatol 2002;118:436-43.
33. Bitoun E, Micheloni A, Lamant L, Bonnart C, Tartaglia-Polcini A, Cobbold C, et al. LEKTI proteolytic processing in human primary keratinocytes, tissue distribution and defective expression in Netherton syndrome. Hum Mol Genet 2003;12:2417-30.
34. Vasilopoulos Y, Cork MJ, Murphy R, Williams HC, Robinson DA, Duff GW, et al. Genetic association between an AACC insertion in the 3′UTR of the stratum corneum chymotryptic enzyme gene and atopic dermatitis. J Invest Dermatol 2004;123:62-6.
35. Walley AJ, Chavanas S, Moffatt MF, Esnouf RM, Ubhi B, Lawrence R, et al. Gene polymorphism in Netherton and common atopic disease. Nat Genet 2001;29:175-8.
36. Kato A, Fukai K, Oiso N, Hosomi N, Murakami T, Ishii M. Association of SPINK5 gene polymorphisms with atopic dermatitis in the Japanese population. Br J Dermatol 2003;148:665-9.
37. Nishio Y, Noguchi E, Shibasaki M, Kamioka M, Ichikawa E, Ichikawa K, et al. Association between polymorphisms in the SPINK5 gene and atopic dermatitis in the Japanese. Genes Immun 2003;4:515-7.
38. Rathmann W, Haastert B, Icks A, Lowel H, Meisinger C, Holle R, et al. High prevalence of undiagnosed diabetes mellitus in Southern Germany: target populations for efficient screening. The KORA survey 2000. Diabetologia 2003;46:182-9.
39. Burney PG, Luczynska C, Chinn S, Jarvis D. The European Community Respiratory Health Survey. Eur Respir J 1994;7:954-60.
40. Hanifin JM, Rajka G. Diagnostic features of atopic eczema. Acta Derm Venereol 1980;92:44-7.
41. Williams HC, Burney PG, Hay RJ, Archer CB, Shipley MJ, Hunter JJ, et al. The U.K. Working Party's Diagnostic Criteria for Atopic Dermatitis. I. Derivation of a minimum set of discriminators for atopic dermatitis. Br J Dermatol 1994;131:383-96.
42. Golding J, Pembrey M, Jones R. ALSPAC—the Avon Longitudinal Study of Parents and Children. I. Study methodology. Paediatr Perinat Epidemiol 2001;15:74-87.
43. Jones RW, Ring S, Tyfield L, Hamvas R, Simmons H, Pembrey M, et al. A new human genetic resource: a DNA bank established as part of the Avon longitudinal study of pregnancy and childhood (ALSPAC). Eur J Hum Genet 2000;8:653-60.
44. Weinberg CR. Methods for detection of parent-of-origin effects in genetic studies of case-parents triads. Am J Hum Genet 1999;65:229-35.
45. Foster MR. Key concepts in model selection: Performance and generalizability. J Math Psychol 2000;44:205-31.
46. Weedon MN, McCarthy MI, Hitman G, Walker M, Groves CJ, Zeggini E, et al. Combining information from common type 2 diabetes risk polymorphisms improves disease prediction. PLoS Med 2006;3:e374.
47. Brezger A, Kneib T, Lang S. BayesX—software for Bayesian inference in structured additive regression models. Munich: Department for Statistics; 2007. Available at: http://www.stat.uni-muenchen.de/~bayesx/bayesx.html. Accessed July 25, 2008.
48. McCullagh P, Nelder JA. Generalized linear models. New York (NY): Chapman and Hall; 1989.
49. Strobl C, Boulesteix AL, Zeileis A, Hothorn T. Bias in random forest variable importance measures: illustrations, sources and a solution. BMC Bioinformatics 2007;8:25.
50. Fleiss JL, Tytun A, Ury HK. A simple approximation for calculating sample sizes for comparing independent proportions. Biometrics 1980;36:343-6.
51. Team RDC. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2007.
52. Candi E, Schmidt R, Melino G. The cornified envelope: a model of cell death in the skin. Nat Rev Mol Cell Biol 2005;6:328-40.
53. Folster-Holst R, Stoll M, Koch WA, Hampe J, Christophers E, Schreiber S. Lack of association of SPINK5 polymorphisms with nonsyndromic atopic dermatitis in the population of Northern Germany. Br J Dermatol 2005;152:1365-7.
54. Lunetta KL, Hayward LB, Segal J, Van Eerdewegh P. Screening large-scale association study data: exploiting interactions using random forests. BMC Genet 2004;5:32.
55. Moore JH. The ubiquitous nature of epistasis in determining susceptibility to common human diseases. Hum Hered 2003;56:73-82.
56. Moore JH. A global view of epistasis. Nat Genet 2005;37:13-4.

METHODS

Statistical methods

Descriptive statistics for quantitative and qualitative values are presented as means ± SD and relative frequencies or absolute numbers, respectively. Deviation from Hardy-Weinberg equilibrium was tested in parents for the family analyses and in control subjects for the case-control analyses. In the family setting we analyzed association of single SNPs with eczema using the classical transmission disequilibrium test. Parent-of-origin effects were investigated with the method proposed by Weinberg.[E1]

Case-control analyses for single SNPs were performed by using logistic regression models adjusted for age and sex. To not constrain the analyses to a specific genetic model, we modeled the 3 categorical genotypes by 2 dummy variables. Thus we estimated separate effects for heterozygotes and homozygotes compared with the wild-type. Because the English and Irish case series are ethnically related, were shown to have highly similar genotypes for *FLG*,[E2,E3] and here had almost identical minor allele frequencies for *KLK7* and *SPINK5* polymorphisms, we analyzed these cases together.

Gene-gene interaction analyses were performed after excluding individuals with 2 mutant *FLG* alleles because these individuals do not express the filaggrin protein, and therefore a biologic interaction with posttranslational modifications by SSCE and LEKTI is not plausible.

Four different approaches were carried out to estimate interaction effects between the single polymorphisms in *FLG*, *KLK7*, and *SPINK5*. In any of these approaches, we adjusted the models for the common covariates age and sex. First, interaction was evaluated by using the logistic regression model with product-interaction terms. We started with a sparse model of the known covariates age, sex, and *FLG*. Incrementally, we extended the model with additional SNPs in *SPINK5* and *KLK7* and SNP-SNP product-interaction terms. The AIC was used to select the appropriate model.[E4]

Second, we defined an interaction score, which counts the number of copies of the potentially disease-associated alleles,[E5] which we used as a covariate in the logistic model.

Third, we modeled maternal effects observed for *SPINK5* using affected offspring from families only. We constructed a 3-category trait: unaffected control subjects, affected offspring with no mutant allele inherited from the mother, and affected offspring with a mutant allele inherited from the mother. We then estimated ORs for both affected status compared with control subjects in a multinomial regression model.

For any of these regression approaches, the quantitative covariate age was modeled nonparametrically in a general additive model framework because of the functional structure of Anscombe residuals after applying analyses in the general linear model framework. For the multinomial model, we used a restricted maximum likelihood approach[E6] implemented in BayesX 1.50.[E7]

Finally, for further exploration, variable importance measures were computed by means of the random forest method. Random forests, and the related method of bagging, are an ensemble method in which a set of classification or regression trees is aggregated for prediction.[E8,E9] Random forests provide variable importance measures that can be used to detect variables relevant for predicting the response. The most commonly used variable importance measure is the permutation importance.[E10]

For variable selection purposes, the advantage of the random forest permutation accuracy importance measure compared with univariate screening methods is that it covers the effect of each predictor variable individually, as well as in multivariate interactions with other predictor variables. For example, Lunetta et al[E11] demonstrated that genetic markers relevant in interactions with other markers or environmental variables can be detected more efficiently by means of random forests than by means of univariate screening methods, such as the Fisher exact test. Here the random forest implementation cforest from the package party[E12,E13] in the R system for statistical computing[E14] is used because it guarantees unbiased variable selection for predictor variables of different scales of measurement.[E10] Predictor variables considered here were age, sex, and respective SNP variables.

One hundred random forests with 500 trees each were fitted with the configuration guaranteeing unbiased variable selection suggested by Strobl et al[E10] to assess the stability of the results. The random forests were built with either 2 randomly preselected variables in each split (argument mtry = 2) or 3 randomly preselected variables in each split (mtry = 3) for comparison. The latter approach is equivalent to bagging, which is contained in random forests as the special case in which the number of randomly preselected variables is equal to the number of available variables. High positive values of the importance measure obtained indicate a high variable importance. Small positive or negative values indicate that a variable is irrelevant for predicting the response.

Additionally, we pooled all 3 study cohorts to increase the power of detecting any single SNP and interaction effect. Because British and German populations might be slightly different with regard to ethnicity,[E15] we accounted for a population effect in every analysis approach by introducing a binary variable, which codes as follows: 1, British/Irish origin; 0, German origin. Thus we corrected the estimated genetic effect for potential population differences. Power calculations[E16] for the pooled single SNP analyses were performed with nQuery7.0, assuming a dominant model.

All statistical analyses were carried out with R 2.6.0,[E14] unless otherwise stated.

REFERENCES

E1. Weinberg CR. Methods for detection of parent-of-origin effects in genetic studies of case-parents triads. Am J Hum Genet 1999;65:229-35.
E2. Sandilands A, O'Regan GM, Liao H, Zhao Y, Terron-Kwiatkowski A, Watson RM, et al. Prevalent and rare mutations in the gene encoding filaggrin cause ichthyosis vulgaris and predispose individuals to atopic dermatitis. J Invest Dermatol 2006;126:1770-5.
E3. 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.
E4. Foster MR. Key concepts in model selection: performance and generalizability. J Math Psychol 2000;44:205-31.
E5. Weedon MN, McCarthy MI, Hitman G, Walker M, Groves CJ, Zeggini E, et al. Combining information from common type 2 diabetes risk polymorphisms improves disease prediction. PLoS Med 2006;3:e374.
E6. Kneib T, Baumgartner B, Steiner WJ. Semiparametric multinomial logit models for analysing consumer choice behaviour. AStA Adv Stat Anal 2007;91:225-44.
E7. Brezger A, Kneib T, Lang S. BayesX—software for Bayesian inference in structured additive regression models. Munich: Department for Statistics; 2007. Available at: http://www.stat.uni-muenchen.de/~bayesx/bayesx.html. Accessed July 25, 2008.
E8. Breiman L. Arcing classifiers. Ann Stat 1998;26:801-49.
E9. Breiman L. Random forests. Machine Learn 2001;45:5-32.
E10. Strobl C, Boulesteix AL, Zeileis A, Hothorn T. Bias in random forest variable importance measures: illustrations, sources and a solution. BMC Bioinformatics 2007;8:25.
E11. Lunetta KL, Hayward LB, Segal J, Van Eerdewegh P. Screening large-scale association study data: exploiting interactions using random forests. BMC Genet 2004;5:32.
E12. Hothorn T, Hornik K, Zeileis A. Unbiased recursive partitioning: a conditional inference framework. J Comp Graph Stat 2006;15:651-74.
E13. Hothorn T, Hornik K, Zeileis A. Party: a laboratory for recursive part(y)itioning package version 0.9-0. 2006.
E14. Team RDC. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2007.
E15. Rogers AJ, Celedon JC, Lasky-Su JA, Weiss ST, Raby BA. Filaggrin mutations confer susceptibility to atopic dermatitis but not to asthma. J Allergy Clin Immunol 2007;120:1332-7.
E16. Fleiss JL, Tytun A, Ury HK. A simple approximation for calculating sample sizes for comparing independent proportions. Biometrics 1980;36:343-6.

TABLE E1. Descriptive characterization of case and control populations

	n	Mean age (y [SD])	Male sex	Mean IgE (SD)	Country	Notation
German offspring	486	22.04 (10.64)	198 (40.7%)	990.4 (2472.7)	Germany	A
German cases	287	35.55 (16.15)	112 (39.0%)	1442.2 (2380.6)	Germany	B
Irish/English cases	418	19.42 (18.43)	199 (51.4%)	3008.0 (6170.0)	Ireland	C
ALSPAC cases*	1583	3.5	849 (53.6%)	286.0 (539.0)	England	D
Sum cases	2774					
KORA S4	3992	49.51 (13.90)	1971 (49.9%)	114.2 (1535.5)	Germany	E
Irish control subjects	552	35.71 (12.27)	170 (30.8%)	NA	Ireland	F
ALSPAC control subjects	6063	3.5*	3135 (51.7%)	200.4 (462.7)	England	G
Sum control subjects	10607					

*Eczema status was determined in all children at 42 months in the ALSPAC cohort.

TABLE E2. Genotyping details and minor allele frequencies in the KORA S4 population-based cohort

SNP ID	MAF	DIR	PCR primer	Extension primer
R501X	0.013	fwd	ACGTTGGATGCTGGAGGAAGACAAGGATCG	ATGCCTGGAGCTGTCTC
		rev	ACGTTGGATGATGGTGTCCTGACCCTCTTG	
2282del4	0.025	fwd	ACGTTGGATGTTGGTGGCTCTGCTGATGGT	GAAGACTCAGACACACAGT
		rev	ACGTTGGATGGTGAGGGACATTCAGAAGAC	
rs2303067	0.479	fwd	ACGTTGGATGCCATCCTTTTTTAGCCAAGC	GATTGTCTTTTGTTTCTTGATT
		rev	ACGTTGGATGCCTCAAAGGAAGCTGTACTC	
AACC ins	0.312	fwd	ACGTTGGATGTGATTGGTTTATCAACAGG	TTTCCTCAAAGATATATTTAAACC
		rev	ACGTTGGATGGACGCCGATGACCTATGAAG	

MAF, Minor allele frequency; *DIR*, direction; *fwd*, forward; *rev*, reverse.

TABLE E3. Allele frequencies across study populations

	GER		IRL		ALSPAC	
	Major	**Minor**	**Major**	**Minor**	**Major**	**Minor**
SPINK5	0.5181	0.48186	0.5227	0.4773	0.5254	0.47463
KLK7	0.6891	0.31089	0.6612	0.3388	NA	NA
FLG	0.9480	0.05201	0.8639	0.1361	0.9558	0.04418

NA, Not applicable.