Analysis of the individual and aggregate genetic contributions of previously identified *SPINK5* , *KLK7* and *FLG* polymorphisms to eczema risk

Stephan Weidinger MD[1,2*], Hansjörg Baurecht MSc[2,3*], Stefan Wagenpfeil PhD[3,4], John Henderson MD[5], Natalija Novak MD[6], Aileen Sandilands PhD[7], Huijia Chen BSc[7], Elke Rodriguez PhD[2], Grainne M. O'Regan MRCPI[8], Rosemarie Watson MD[8], Haihui Liao MD[7] , Yiwei Zhao MD[7], Jonathan NWN Barker MD[9], Michael Allen MPhil[9] , Nick Reynolds MD[10], Simon Meggit[10], Kate Northstone PhD[11], George Davy Smith*, Carolin Strobl MSc[12], Caroline Stahl[1,3], Thomas Kneib PhD[12], Norman Klopp PhD[13], Thomas Bieber MD[5], Heidrun Behrendt MD[2], Colin NA Palmer PhD[14], H.-Erich Wichmann MD[13], Johannes Ring MD[1,4], Thomas Illig PhD[13], WH. Irwin McLean PhD, DSc, FRSE[7*], Alan D. Irvine MD[8,15*]

1 Department of Dermatology and Allergy Biederstein, Technical University Munich (TUM), Munich, Germany
2 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, Munich, Germany
3 Institute for Medical Statistics and Epidemiology IMSE, Technical University Munich (TUM), Munich, Germany
4 Graduate School of Information Science in Health (GSISH), Technical University Munich (TUM), Munich, Germany
5 Department of Community-Based Medicine, University of Bristol, UK
6 Department of Dermatology and Allergy, University of Bonn
7 Epithelial Genetics Group, Human Genetics Unit, Division of Pathology and Neuroscience, University of Dundee, UK
8 Department of Paediatric Dermatology, Our Lady's Children's Hospital, Crumlin, Dublin, Ireland
9 St. John's Institute of Dermatology, Guys Hospital, London, UK
10 Department of Dermatology, University of Newcastle, UK
11 Department of Social Medicine, University of Bristol, UK
12 Institute for Statistics, Ludwig-Maximilans University (LMU), Munich, Germany
13 Department of Epidemiology, Helmholtz Zentrum Munich- German Research Center for Environmental Health
14 Population Pharmacogenetics Group, Biomedical Research Centre, University of Dundee, UK
15 Department of Clinical Medicine, Trinity College Dublin, Ireland
16 GDS affiliation: MRC Centre for Causal Analyses in Translational Epidemiology

* contributed equally

Correspondence to SW or ADI, as below:
Stephan Weidinger
Department of Dermatology and Allergy Biederstein,
Technical University Munich (TUM)
Munich 80802, Germany
Phone: 0049-89-4140-3396, Fax: 0049-89-4140-3127
Email: weidinger@lrz.tum.de
Alan D. Irvine
Paediatric Dermatology
Our Lady's Children's Hospital, Crumlin
Dublin 12, IRELAND
Phone: +3531 428 2532, Fax: +3531 428 2651
Email: irvinea@tcd.ie

ABSTRACT

Background

Polymorphisms in the serine protease inhibitor gene *SPINK5* and the serine protease *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 regulation of post-translation processing of filaggrin, the strongest identified genetic risk factor for eczema to date.

Objectives

To clarify the individual risk of eczema conferred by the *SPINK5* polymorphism-rs2303067 (Lys420Ser) and a previously described insertion in the 3'UTR 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 cases from a German family-based study, an additional 287 German cases, and 418 unrelated Irish/English eczema cases (n for 3 genes studied = 1191 vs. 4544 controls). We then additionally studied the *SPINK5* polymorphism and *FLG* mutations in 1583 eczema patients from the ALSPAC cohort (n for 2 genes studied = 2774 vs. 10607 controls).

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 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.

Abstract word count: 254

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

Capsule Summary

Previously reported polymorphisms in *SPINK5*, *KLK7* and *FLG* were studied in 2774 eczema cases and 10607 controls. No association with eczema was seen with the *KLK7* insertion, a weak maternal effect was seen with the *SPINK5* Lys420Ser polymorphism; neither polymorphism had an epistatic effect with *FLG*.

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

Abbreviations:

AIC	Akaike Information Criterion
ALSPAC	Avon Longitudinal Study of Parents and Children
CI	Confidence Interval
FLG	Filaggrin
IC	Logistic regression model with an interaction score
KLK7	Kallikrein-related peptidase 7
KORA	Co-operative Health Research in the Region of Augsburg
LEKTI	Lympho-epithelial Kazal type inhibitor
LRM	Logistic regression model with product interaction terms
MAF	Minor Allele Frequency
MNM	Multinomial regression model
OR	Odds Ratio
RF	Random forest
SE	Standard error
SNP	Single nucleotide polymorphism
SPINK5	Serine peptidase inhibitor, Kazal type 5
SSCE	Stratum corneum chymotryptic enzyme
TDT	Transmission disequilibrium test

DECLARATION OF ALL SOURCES OF FUNDING:

This work was funded by the German Ministry of Education and Research (BMBF) as part of the National Genome Research Network (NGFN) with grant NUW-S31T05. S.W. and S.W. are supported by research grants KKF-07/04 and KKF-27/05 of the University Hospital “Rechts der Isar”, Technical University Munich. In addition S.W. (first author) is supported by a grant from the “Wilhelm-Vaillant Stiftung”. We are extremely grateful to all 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. The UK Medical Research Council, the Wellcome Trust and the University of Bristol provide core support for ALSPAC. 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 CP. CP 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 number G0700314) and anonymous donations from atopy families in the Tayside region of Scotland.

CONFLICT OF INTEREST STATEMENT

WHIM holds patents related to diagnosis and therapeutic correction of the filaggrin gene. All other authors declare that they have no competing financial or other interests.

Introduction

Eczema is a common chronic inflammatory skin disease with a complex, multifactorial aetiology 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 two common (R501X and 2282del4) and several rare mutations within the *filaggrin* (*FLG*) gene 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[6-21] 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[22-25].

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[26-28].

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 by several dephosphorylation and proteolytic steps[29, 30], 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)[31, 32], which is possibly regulated by the

serine protease inhibitor LETKI, encoded by *SPINK5*[32-34]. Interestingly, an insertion in the 3′ untranslated region of the *kallikrein 7* gene (*KLK7*) encoding SCCE[35] 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,[36] this association has been replicated in 2 small Japanese studies, [37, 38] but other studies have failed to replicate this association.{REFS HERE} 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 so 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 post translational modification of filaggrin, we also sought to examine gene-gene interactions between *FLG*, *KLK7* and *SPINK5.*

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 eczema cases, and 552 Irish blood donor controls, an additional series of 287 eczema cases from Germany and the population-based cross-sectional KORA S4 cohort (n=3992). In addition, the population-based ALSPAC cohort (n=7646) was typed for the Lys420Ser *SPINK5* polymorphism and the two 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 supplementary table 1).

The study designs have been described in detail elsewhere[9, 18, 39]. 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 standardised questionnaire, that, in addition to demographic data included the basic allergy questions of the European Community Respiratory Health Survey (ECRHS) on respiratory health[40]. 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 & Rajka[41] and the UK diagnostic criteria for eczema[42].

All German eczema cases were unrelated and of white origin with eczema diagnosed on the basis of a skin examination by experienced dermatologists using the UK diagnostic criteria[42]. In the family collection, 10.4% of the parents suffered from eczema (9.3% affected fathers and 11.1% affected mothers).

Eczema cases 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 paediatric 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 in the online repository.

The Avon Longitudinal Study of Parents and Children (ALSPAC) is a longitudinal, population-based birth cohort study that recruited 14541 unrelated pregnant women resident in Avon, UK with expected dates of delivery between 1st April 1991 and 31st December 1992. There were 14,062 liveborn children. The study protocol has been described previously[43, 44] and further details are on the ALSPAC website:

http://www.alspac.bris.ac.uk. At 6, 18, 30 and 42 months of age the mothers were asked whether their child 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, 43, 44]

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

Genotyping:

Genotyping in German samples was performed using the MassARRAY system (Sequenom, San Diego, USA) 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 in the online repository table E2.

The Lys420Ser polymorphism in *SPINK5* (rs2303067) was typed in the Irish and English eczema cohorts and controls using a predesigned SNP Taqman Genotyping Assay from Applied Biosystems (product C_2000249_10) and run on a 7900HT Fast Real-Time PCR system using the manufacturers recommended protocol. The 4bp insertion in the 3'UTR of *KLK7* was typed in Irish and English eczema cohorts and controls by sizing of fluorescently labelled PCR products on an Applied Biosystems 3130xl Genetic Analyser. 10ml PCR reactions were performed using 25ng of genomic DNA with 400nM forward primer (5' gtt tct tca agt gtg caa gtt cac caa 3') and 400nM FAM-labelled reverse primer (5' gat tgg ttt atc aac agg gc 3') in AmpliTaq Gold Buffer containing 1.5mM MgCl2, 10nmol of each dNTP, 4% v/v DMSO and 0.25U AmpliTaq Gold polymerase (Applied Biosystems). PCR reactions were amplified using an

annealing temperature of 58°C. Diluted PCR products were sized against ROX-500 size standards (Applied Biosystems). Allele sizes were 201bp and 205bp (insertion).

For the ALSPAC cohort the 2 commonest *FLG* mutations were typed as previously described [25] and the *SPINK5* Lys420Ser polymorphism was typed by Taqman assay. As we discovered three 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 given by mean ± standard deviation (SD) and relative frequencies or absolute numbers, respectively. Deviation from Hardy-Weinberg equilibrium was tested in parents for the family-analyses and in controls for the case-control analyses. In the family-setting we analysed association of single SNPs with eczema using the classical transmission disequilibrium test (TDT). Parent-of-origin effects were investigated with the method proposed by Weinberg[45].

Case-control analyses for single SNPs was performed using logistic regression models adjusted for age and gender. In order not to constrain the analyses to a specific genetic model we modelled the three categorical genotypes by two dummy variables.

Gene-gene interaction analyses was performed after excluding individuals with two mutant *FLG* alleles since these individuals do not express the filaggrin protein and therefore a biological interaction with post-translational modifications by SSCE and LEKTI is not plausible.

In order to estimate interaction effects between the single polymorphisms in *FLG*, *SSCE* and *LEKTI,* four different approaches were carried out. In any of these approaches we adjusted the models for the common covariates age and gender. Firstly,

277 interaction was evaluated using the logistic regression model with product interaction
278 terms (LRM). The Akaike Information Criterion (AIC) was used to select the appropriate
279 model[46]Secondly, we defined an interaction score (IC) which counts the number of
280 copies of the potentially disease-associated alleles[47], which we used as a covariate in
281 the logistic model. Thirdly, we modelled maternal effects observed for *SPINK5* using
282 affected offspring from families only and estimated ORs for two different affection
283 status.. compared with controls in a multinomial regression model (MRM), which was
284 carried out with BayesX 1.50[48]. An elaborate description of the statistical methods are
285 given in the online repository.
286 For any of these regression approaches the quantitative covariate age was modelled
287 non-parametrically in a general additive model framework (GAM)
288 Finally, for further exploration, variable importance measures were computed by means
289 of the random forest method. Random forests provide variable importance measures
290 that can be employed to detect variables relevant for predicting the response. The most
291 commonly used variable importance measure is the permutation importance[49].
292 High positive values of the importance measure obtained indicate a high variable
293 importance. Small positive or negative values indicate that a variable is irrelevant for
294 predicting the response.
295 Additionally, we pooled all three study cohorts to increase the power of detecting any
296 single SNP and interaction effect. As British and German populations might be slightly
297 different with regard to ethnicity[17] we accounted for a population effect in every analysis
298 approach by introducing a binary independent variable which codes 1="British/Irish
299 origin" and 0="German origin". Thus we corrected the estimated genetic effect for
300 potential population differences.
301 Power calculations[50] for the pooled single SNP analyses were performed with
302 nQuery7.0 assuming a dominant model.

All statistical analyses were carried out with R 2.6.0[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 supplementary table 1. 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, 52], although for this study further families and cases 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 over-transmission to eczema-affected offspring (OR=1.25, 95%CI=1.04-1.50, p=0.018). No associations were seen for the *KLK7* 3'UTR insertion (Table 1). In this family cohort, the power to detect a proportion of 1.2 between transmission of the risk allele vs. non transmission (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%.

Since 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 to paternal alleles of 2.18 (95% CI 1.38-3.43, p=0.0008). The observed power to detect parent-of-origin effects as described in [45] was greater than 90%. In the pooled German case-control cohort the presence of at least one *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 were associated with eczema (Table 2).

In the Irish case-control cohort comparable results could be seen: presence of a *FLG* null allele increased the risk for eczema about 5.88 fold (95%CI=3.85-8.99, $p=2.8x10^{-16}$). No association between *SPINK5* or *KLK7* and eczema was detected (Table 2).

The analyses of ALSPAC also showed a strong *FLG* effect (OR=2.23, 95%CI=1.87-2.67, $1.2x10^{-18}$). Consistent with our other cohorts, no association between the Lys420Ser *SPINK5* polymorphism and eczema was found (Table 2).

For the pooled study in all models a population effect was estimated as a confounder with p-values$<10^{-4}$. An exceptionally strong *FLG* effect on eczema was seen for the pooled analyses (OR=3.36, 95%CI=2.97-3.79, $p=1.3x10^{-84}$) with a power of >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=1.2 for carriers of the *SPINK5* rare allele compared to non-carriers 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 MAF in the cases of 0.56.

Gene-gene interaction analyses:

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

The LRM 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 gender remained in the model according to the AIC (Table 3).

In the IC approach, in addition to one mutant *FLG* allele the A allele of rs2303067 (*SPINK5*) as well as 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 fall 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 risk *FLG* alleles in these cells.

For the Irish/British case-control cohort similar results were observed, but the OR in the four variant group was more than twice as high as in the German case-control analyses. Using the MNM approach,we tried to account for the maternal parent-of-origin effect reported by Walley et al. 2001[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 RF approach (Table 4), 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%, while 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 gender 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 consistent results to the previously estimated effects. In all models we estimated a population effect as a confounder with p-values$<10^{-4}$.

Discussion

This large-scale study examined variants in three candidate genes, which 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 the skin barrier function. In addition, due to their biological 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[53]. Two common null mutations in the *FLG* gene (R501X, 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 so 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[54].

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 Lys420Ser appears to have a weaker effect, and only when maternally inherited, and the *KLK7* insertion does not exert an effect. However, our data does not preclude that there are other variants in *SPINK5* and/or *KLK7* 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 three 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, as interactions would be captured by the random forest variable importance[55]. Since there was a considerable difference in age between cases and controls, we considered age as a covariate in all analyses and modelled it non-parametrically. However, this leads to little variability in the response data explained by *FLG* mutations in the RF 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[56, 57]. With whole-genome association studies with hundreds of thousands of measured genetic variations emerging, analyses of the complex molecular interactions on the DNA level is 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, while our data underlines the exceptional importance of filaggrin 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* may 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 using powerful statistical methods will aid in elucidating the relationship between combinations of polymorphisms for eczema susceptibility.

Table 1: ORs and 95%CIs for associations between polymorphisms and eczema in the family-based analysis. T, transmitted; U, untransmitted; comb., combined.

German families (A)	Gene	Polymorphism	T:U	OR	95% CI	P-value
	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}
		comb. genotype	110: 40	2.75	1.93 3.98	1.8×10^{-8}
	SPINK5	rs2303067	259: 207	1.25	1.04 1.50	0.01815
	KLK7	AACC ins	183: 177	1.03	0.84 1.27	0.79215

Table 2: Associations between polymorphisms and eczema in the case-control approaches. All logistic regression models are adjusted for age and gender. Every block refers to a single model where only estimates of the genetic variables are displayed. No specific genetic model was assumed and estimates are given for every genotype compared to the wildtype. comb., combined; het, heterozygous; hom, homozygous. Populations as designated in table 1 are indicated in brackets.

	Gene	Polymorphism	OR	95% CI		P-value
German *(A,B,E)* *cases: 773 controls: 3992 total: 4765*	*FLG*	comb. genotype (het. vs. wt)	4.15	3.10	5.56	1.3×10^{-21}
		comb. genotype (hom. vs. wt)	23.11	7.23	73.89	1.2×10^{-7}
	SPINK5	rs2303067(het. vs. wt)	1.14	0.87	1.49	0.34285
		rs2303067(hom. vs. wt)	1.22	0.89	1.67	0.21252
	KLK7	AACC ins (het. vs. wt)	1.11	0.89	1.40	0.35472
		AACC ins (hom. vs. wt)	0.95	0.64	1.42	0.81033
Irish/English ***(C,F)*** ***cases: 418 controls: 552 total: 970***	*FLG*	comb. genotype (het. vs. wt)	4.34	2.77	6.79	1.3×10^{-10}
		comb. genotype (hom. vs. wt)	2.6×10^{56}	0	∞	1.0
	SPINK5	rs2303067(het. vs. wt)	0.78	0.52	1.18	0.23945
		rs2303067(hom. vs. wt)	1.15	0.71	1.87	0.57398
	KLK7	(het. vs. wt)	1.29	0.89	1.86	0.17756
		(hom. vs. wt)	0.91	0.47	1.76	0.76988
ALSPAC *(D,G)* *cases: 1583 controls: 6063 total: 7646*	*FLG*	comb. genotype (het. vs. wt)	2.17	1.81	2.60	3.5×10^{-17}
		comb. genotype (hom. vs. wt)	3.3×10^{6}	0	∞	0.94524
	SPINK5	rs2303067(het. vs. wt)	0.96	0.84	1.09	0.52176
		rs2303067(hom. vs. wt)	1.14	0.98	1.34	0.09156
Pooled* *(A,B,C,D,E,F,G)* *cases: 2774 controls: 10607 total: 13381*	*FLG*	comb. genotype (het. vs. wt)	3.04	2.68	3.44	8.5×10^{-68}
		comb. genotype (hom. vs. wt)	49.38	19.72	123.61	8.2×10^{-17}
	SPINK5	rs2303067(het. vs. wt)	0.97	0.87	1.07	0.50850
		rs2303067(hom. vs. wt)	1.13	1.00	1.27	0.04545

* for the pooled analyses we estimated in all models a population effect as confounder with p-values<10^{-4}

Table 3: Associations between polymorphisms or gene-gene interaction terms and eczema. All models are adjusted for age and gender. Every block refers to a single model. LRM, logistic regression model with product interaction terms; IC, logistic regression model with interaction score; MNM, multinomial logistic model considering maternal inheritance. Populations as designated in table 1 are indicated in brackets.

German						
	gene-gene interaction term		**OR**	**95% CI**		**P-value**
LRM *(A,B,E)*	*FLG* comb. genotype (het. vs. wt.)		6.11	4.09	9.12	$3.3x10^{-10}$
	rs2303067 (het. vs. wt.)		1.28	0.87	1.90	0.11862
	rs2303067 (hom. vs. wt.)		1.28	0.83	1.97	0.18638
	FLG comb. genotype (het.) rs2303067 (het.)		0.62	0.34	1.11	0.17805
	FLG comb. genotype (het.) rs2303067 (hom.)		0.61	0.06	6.20	0.23915
IC *(A,B,E)*	*interaction score**	1 vs. 0	1.08	0.72	1.61	0.71881
		2 vs. 0	1.18	0.80	1.76	0.39974
		3 vs. 0	1.67	1.08	2.57	0.01982
		4 vs. 0	2.75	1.53	4.92	0.00067
		5 vs. 0	1.45	0.14	14.70	0.75107
MNM *(A,E)*	**Multinomial logit trait**	**covariates**	**OR**	**95% CI**		**P-value**
	0: controls	1:gender f vs. m	1.69	1.12	2.56	0.01338
	1: eczema & *SPINK5* (not maternal)	1: *FLG* comb. genotype	5.22	3.12	8.73	$4.9x10^{-7}$
		2:gender f vs. m	1.32	0.87	2.00	0.18653
	2: eczema & *SPINK5* (maternal)	2: *FLG* comb. genotype	6.16	3.70	10.26	$1.1x10^{-7}$
	test $ß_{1:combGeno}=ß_{2:combGeno}$					0.80071
Irish						
	gene-gene interaction term		**OR**	**95% CI**		**P-value**
LRM *(C,F)*	*FLG* comb. genotype (het. vs. wt.)		4.34	2.77	6.79	1.3×10^{-10}
IC *(C,F)*	*interaction score**	1 vs. 0	1.32	0.64	2.75	0.45470
		2 vs. 0	1.18	0.56	2.46	0.66276
		3 vs. 0	1.96	0.92	4.20	0.08175
		4 vs. 0	5.17	1.88	14.24	0.00148
		5 vs. 0	0.01	~0	$2.6x10^{10}$	0.74599
ALSPAC						
	gene-gene interaction term		**OR**	**95% CI**		**P-value**
LRM *(D,G)*	*FLG* comb. genotype (het. vs. wt.)		2.26	1.63	3.14	1.2×10^{-6}
	rs2303067 (het. vs. wt.)		1.02	0.87	1.19	0.83390
	rs2303067 (hom. vs. wt.)		1.18	0.98	1.41	0.07389
	FLG comb. genotype (het.) rs2303067 (het.)		0.90	0.59	1.38	0.63859
	FLG comb. genotype (het.) rs2303067 (hom.)		1.04	0.62	1.74	0.87865
IC *(D,G)*	*interaction score**	1 vs. 0	1.06	0.92	1.23	0.39907
		2 vs. 0	1.29	1.09	1.51	0.00245
		3 vs. 0	2.72	1.84	4.02	5.4×10^{-7}
pooled**						
LRM	**gene-gene interaction term**		**OR**	**95% CI**		**P-value**
(A,B,C,D ,E,F,G)	*FLG* comb. genotype (het. vs. wt.)		3.43	2.71	4.35	1.7×10^{-24}
	rs2303067 (het. vs. wt.)		1.03	0.92	1.16	0.58801
	rs2303067 (hom. vs. wt.)		1.15	1.00	1.32	0.04914
	FLG comb. genotype (het.) rs2303067 (het.)		0.81	0.60	1.09	0.16864
	FLG comb. genotype (het.) rs2303067 (hom.)		0.96	0.67	1.37	0.82037
IC *(A,B,C,D ,E,F,G)*	*interaction score**	1 vs. 0	1.10	0.99	1.23	0.08116
		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}

*score: number of copies of the "risk" allels in *FLG* comb. genotype, *SPINK5* rs2303067 and *KLK7*, for ALSPAC and pooled anaylsis the score is reduced to the number of copies of *FLG* and *SPINK5* allels

** for the pooled analyses in all models we estimated a population effect as confounder with p-values<10

Table 4: Results for random forests with 2 randomly pre-selected variables in each split. The average permutation importance +/- 2 standard errors of the mean over 100 iterations are displayed for each variable. Results for random forests with 5 randomly preselected variables, i.e. 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 1 are indicated in brackets.

	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	n.a.	n.a.	n.a.	n.a.
gender	-0.00037	-0.00032	0.01075	0.01106	-0.00004	-0.00003	-0.00022	-0.00021
FLG comb. 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	n.a.	n.a.	n.a.	n.a.
population	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	-0.00008	-0.00006

Figure 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.

Figure 2:

Residual plot of age after fitting a general linear model for SNP-SNP interaction analyses. The anscombe residuals show a functional form and do not spread randomly. Hence a general additive model is fitted for the data. The upper row reflects anscombe rediuals of the German case-control analyses; the lower row shows the anscombe residuals for the same models in the Irish case-control cohort.

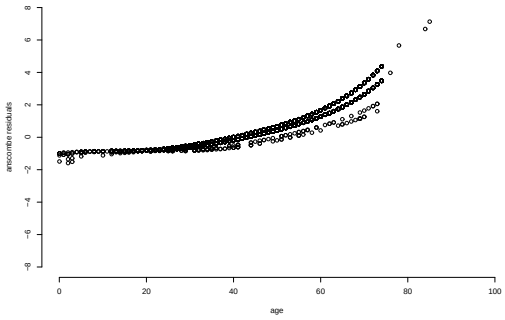
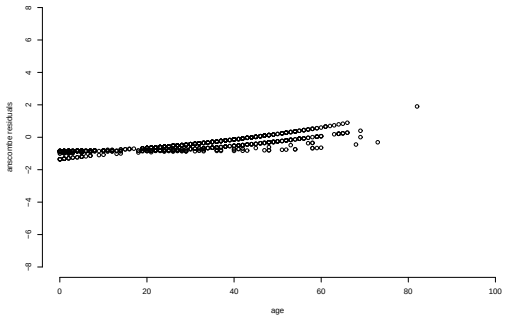
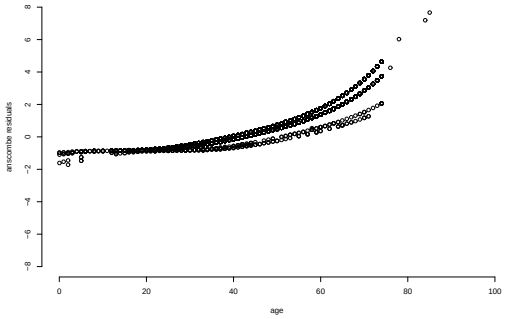
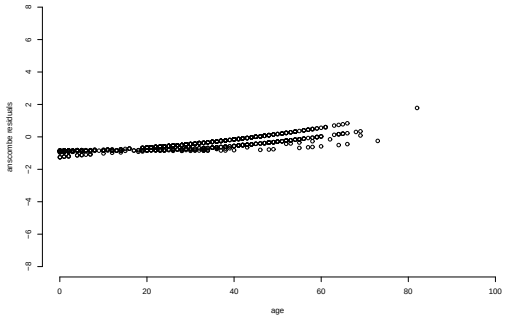
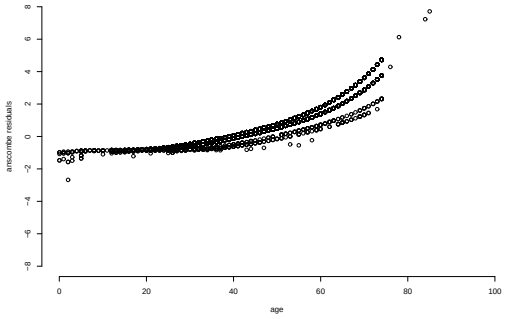
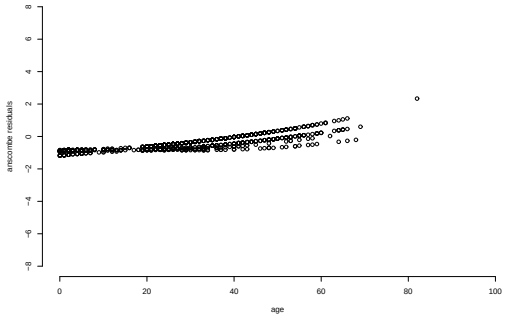
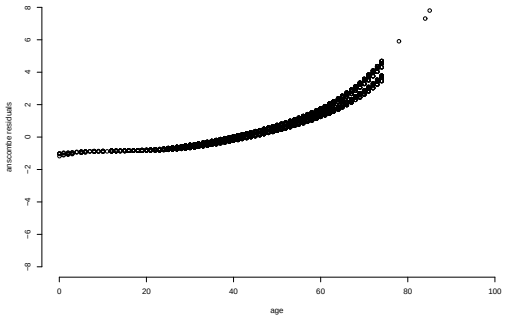
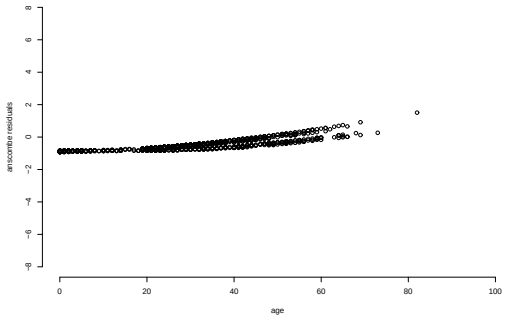
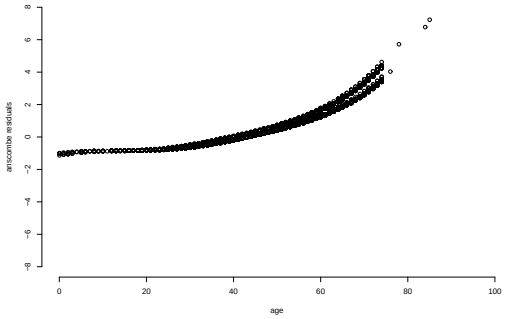
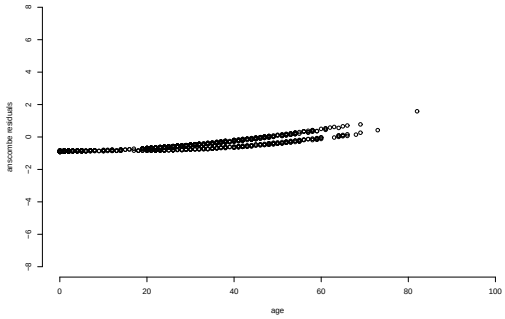
References:

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 2006.
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 2006.
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. 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.
14. 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.
15. Morar N, Cookson WO, Harper JI, Moffatt MF. Filaggrin mutations in children with severe atopic dermatitis. J Invest Dermatol 2007; 127:1667-72.
16. 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.
17. 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.

18. 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.
19. 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.
20. 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.
21. 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; 121:940-46 e3.
22. 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; in press.
23. Rodriguez E, Illig T, Weidinger S. Filaggrin loss-of-function mutations and association with allergic diseases. Pharmacogenomics 2008; 9:399-413.
24. 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.
25. 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 e9.
26. Hudson TJ. Skin barrier function and allergic risk. Nat Genet 2006; 38:399-400.
27. 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.
28. 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.
29. 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.
30. 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.
31. 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.
32. 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.
33. 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.
34. 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.
35. 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.

588 36. Walley AJ, Chavanas S, Moffatt MF, Esnouf RM, Ubhi B, Lawrence R, et al. Gene
589 polymorphism in Netherton and common atopic disease. Nat Genet 2001; 29:175-8.
590 37. Kato A, Fukai K, Oiso N, Hosomi N, Murakami T, Ishii M. Association of SPINK5
591 gene polymorphisms with atopic dermatitis in the Japanese population. Br J Dermatol
592 2003; 148:665-9.
593 38. Nishio Y, Noguchi E, Shibasaki M, Kamioka M, Ichikawa E, Ichikawa K, et al.
594 Association between polymorphisms in the SPINK5 gene and atopic dermatitis in the
595 Japanese. Genes Immun 2003; 4:515-7.
596 39. Rathmann W, Haastert B, Icks A, Lowel H, Meisinger C, Holle R, et al. High
597 prevalence of undiagnosed diabetes mellitus in Southern Germany: target populations
598 for efficient screening. The KORA survey 2000. Diabetologia 2003; 46:182-9.
599 40. Burney PG, Luczynska C, Chinn S, Jarvis D. The European Community Respiratory
600 Health Survey. Eur Respir J 1994; 7:954-60.
601 41. Hanifin JM, Rajka G. Diagnostic features of atopic eczema. Acta Derm Venereol
602 1980:44-7.
603 42. Williams HC, Burney PG, Hay RJ, Archer CB, Shipley MJ, Hunter JJ, et al. The U.K.
604 Working Party's Diagnostic Criteria for Atopic Dermatitis. I. Derivation of a minimum
605 set of discriminators for atopic dermatitis. Br J Dermatol 1994; 131:383-96.
606 43. Golding J, Pembrey M, Jones R. ALSPAC--the Avon Longitudinal Study of Parents
607 and Children. I. Study methodology. Paediatr Perinat Epidemiol 2001; 15:74-87.
608 44. Jones RW, Ring S, Tyfield L, Hamvas R, Simmons H, Pembrey M, et al. A new
609 human genetic resource: a DNA bank established as part of the Avon longitudinal
610 study of pregnancy and childhood (ALSPAC). Eur J Hum Genet 2000; 8:653-60.
611 45. Weinberg CR. Methods for detection of parent-of-origin effects in genetic studies of
612 case-parents triads. Am J Hum Genet 1999; 65:229-35.
613 46. Foster MR. Key concepts in model selection: Performance and generalizability. J
614 Math Psychol 2000; 44:205-31.
615 47. Weedon MN, McCarthy MI, Hitman G, Walker M, Groves CJ, Zeggini E, et al.
616 Combining information from common type 2 diabetes risk polymorphisms improves
617 disease prediction. PLoS Med 2006; 3:e374.
618 48. Brezger A, Kneib T, Lang S. BayesX - Software for Bayesian Inference in Structured
619 Additive Regression Models. Munich, 2007.
620 49. Strobl C, Boulesteix AL, Zeileis A, Hothorn T. Bias in random forest variable
621 importance measures: illustrations, sources and a solution. BMC Bioinformatics 2007;
622 8:25.
623 50. Fleiss JL, Tytun A, Ury HK. A Simple Approximation for Calculating Sample Sizes
624 for Comparing Independent Proportions. Biometrics 1980; 36:343-6.
625 51. Team RDC. R: A Language and Environment for Statistical Computing. Vienna: R
626 Foundation for Statistical Computing, 2007.
627 52. Weidinger S, Rodriguez E, Stahl C, Wagenpfeil S, Klopp N, Illig T, et al. Filaggrin
628 mutations strongly predispose to eatly-onset and extrinsic atopic dermatitis J Invest
629 Dermatol 2006; 126:in press.
630 53. Candi E, Schmidt R, Melino G. The cornified envelope: a model of cell death in the
631 skin. Nat Rev Mol Cell Biol 2005; 6:328-40.
632 54. Folster-Holst R, Stoll M, Koch WA, Hampe J, Christophers E, Schreiber S. Lack of
633 association of SPINK5 polymorphisms with nonsyndromic atopic dermatitis in the
634 population of Northern Germany. Br J Dermatol 2005; 152:1365-7.
635 55. Lunetta KL, Hayward LB, Segal J, Van Eerdewegh P. Screening large-scale
636 association study data: exploiting interactions using random forests. BMC Genet
637 2004; 5:32.

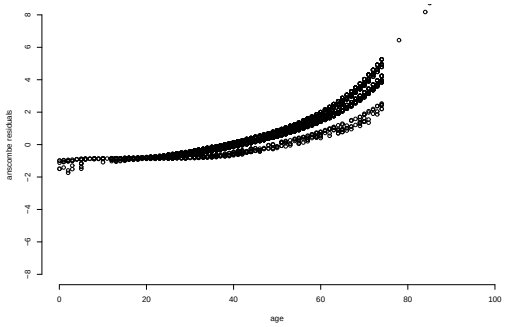
56. Moore JH. The ubiquitous nature of epistasis in determining susceptibility to common human diseases. Hum Hered 2003; 56:73-82.
57. Moore JH. A global view of epistasis. Nat Genet 2005; 37:13-4.

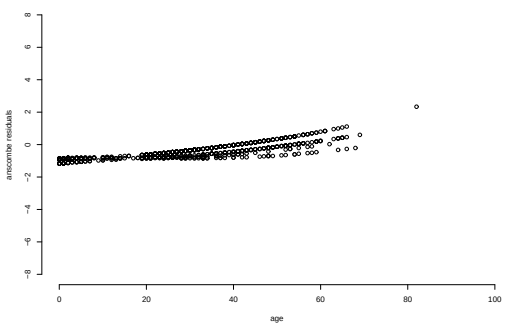
	German	Irish
R501X		
2282del4		
FLG comb.Geno		
rs2303067		
KLK7		

German

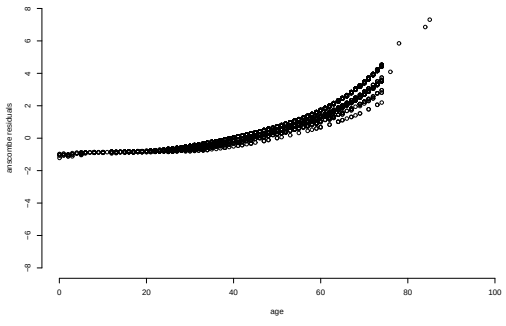
Irish

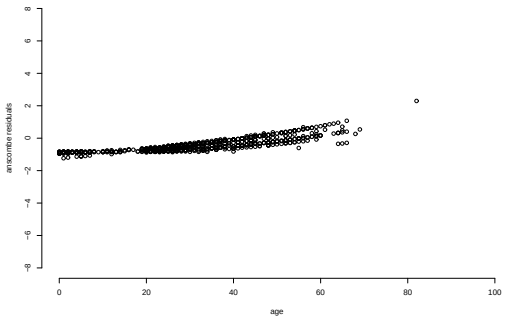
LM



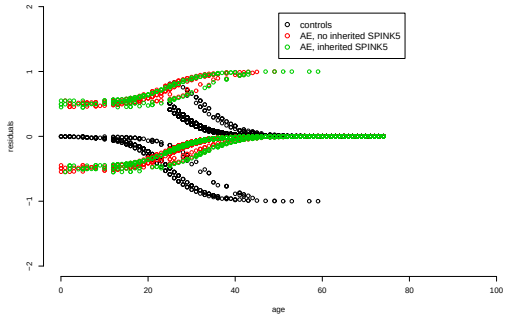


IC





MNM



Online Repository:

Analysis of the individual and aggregate genetic contributions of previously identified *SPINK5* , *KLK7* and *FLG* polymorphisms to eczema risk

Stephan Weidinger MD[1,2*], Hansjörg Baurecht MSc[2,3*], Stefan Wagenpfeil PhD[3,4,] John Henderson MD[5], Natalija Novak MD[6], Aileen Sandilands PhD[7], Huijia Chen BSc[7], Elke Rodriguez PhD[2], Grainne M. O'Regan MRCPI[8], Rosemarie Watson MD[8], Haihui Liao MD[7] , Yiwei Zhao MD[7], Jonathan NWN Barker MD[9], Michael Allen MPhil[9] , Nick Reynolds MD[10], Simon Meggit[10], Kate Northstone PhD[11], George Davy Smith Carolin Strobl MSc[12], Caroline Stahl[1,3], Thomas Kneib PhD[12], Norman Klopp PhD[13], Thomas Bieber MD[5], Heidrun Behrendt MD[2,] Colin NA Palmer PhD[14], H.-Erich Wichmann MD[13], Johannes Ring MD[1,4,] Thomas Illig PhD[13], WH. Irwin McLean PhD, DSc, FRSE[7*], Alan D. Irvine MD[8,15*]

1 Department of Dermatology and Allergy Biederstein, Technical University Munich (TUM), Munich, Germany
2 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, Munich, Germany
3 Institute for Medical Statistics and Epidemiology IMSE, Technical University Munich (TUM), Munich, Germany
4 Graduate School of Information Science in Health (GSISH), Technical University Munich (TUM), Munich, Germany
5 Department of Community-Based Medicine, University of Bristol, UK
6 Department of Dermatology and Allergy, University of Bonn
7 Epithelial Genetics Group, Human Genetics Unit, Division of Pathology and Neuroscience, University of Dundee, UK
8 Department of Paediatric Dermatology, Our Lady's Children's Hospital, Crumlin, Dublin, Ireland
9 St. John's Institute of Dermatology, Guys Hospital, London, UK
10 Department of Dermatology, University of Newcastle, UK
11 Department of Social Medicine, University of Bristol, UK
12 Institute for Statistics, Ludwig-Maximilans University (LMU), Munich, Germany
13 Department of Epidemiology, Helmholtz Zentrum Munich- German Research Center for Environmental Health
14 Population Pharmacogenetics Group, Biomedical Research Centre, University of Dundee, UK
15 Department of Clinical Medicine, Trinity College Dublin, Ireland

* contributed equally
Correspondence to SW or ADI, as below:
Stephan Weidinger
Department of Dermatology and Allergy Biederstein,
Technical University Munich (TUM)
Munich 80802, Germany
Phone: 0049-89-4140-3396, Fax: 0049-89-4140-3127
Email: weidinger@lrz.tum.de
Alan D. Irvine
Paediatric Dermatology
Our Lady's Children's Hospital, Crumlin
Dublin 12, IRELAND
Phone: +3531 428 2532, Fax: +3531 428 2651
Email: irvinea@tcd.ie

METHODS

Statistical methods:

Descriptive statistics for quantitative and qualitative values are given by mean ± standard deviation (SD) and relative frequencies or absolute numbers, respectively. Deviation from Hardy-Weinberg equilibrium was tested in parents for the family-analyses and in controls for the case-control analyses. In the family-setting we analysed association of single SNPs with eczema using the classical transmission disequilibrium test (TDT). Parent-of-origin effects were investigated with the method proposed by Weinberg[1].

Case-control analyses for single SNPs was performed using logistic regression models adjusted for age and gender. In order not to constrain the analyses to a specific genetic model we modelled the three categorical genotypes by two dummy variables. Thus we estimated separate effects for heterozygotes and homozygotes as compared to wildtype. As the English and Irish case series are ethnically related, were shown to have highly similar genotypes for *FLG*[2, 3], and here had almost identical MAFS for *KLK7* and *SPINK5* polymorphisms (supplementary table 3), we analysed these cases together.

Gene-gene interaction analyses was performed after excluding individuals with two mutant *FLG* alleles since these individuals do not express the filaggrin protein and therefore a biological interaction with post-translational modifications by SSCE and LEKTI is not plausible.

In order to estimate interaction effects between the single polymorphisms in *FLG*, *SSCE* and *LEKTI,* four different approaches were carried out. In any of these approaches we adjusted the models for the common covariates age and gender. Firstly, interaction was evaluated using the logistic regression model with product

interaction terms (LRM). We started with a sparse model of the known covariates age, gender and *FLG*. Incrementally we extended the model with additional SNPs in *SPINK5* and *KLK7* and SNP-SNP-product-interaction terms. The Akaike Information Criterion (AIC) was used to select the appropriate model[4].

Secondly, we defined an interaction score (IC) which counts the number of copies of the potentially disease-associated alleles[5], which we used as a covariate in the logistic model. Thirdly, we modelled maternal effects observed for *SPINK5* using affected offspring from families only. We constructed a three-categorical trait: unaffected controls, affected offspring with no mutant allele inherited from the mother, affected offspring with a mutant allele inherited from the mother. We then estimated ORs for both affected status compared with controls in a multinomial regression model (MRM).

For any of these regression approaches the quantitative covariate age was modelled non-parametrically in a general additive model framework (GAM) due to the functional structure of anscombe residuals after applying analyses in the general linear model framework (GLM). For the multinomial model we used a REML-approach[6] implemented in BayesX 1.507[7].

Finally, for further exploration, variable importance measures were computed by means of the random forest method. Random forests, and the related method bagging, are an ensemble method where a set of classification or regression trees is aggregated for prediction[8, 9]. Random forests provide variable importance measures that can be employed to detect variables relevant for predicting the response. The most commonly used variable importance measure is the permutation importance[10].

For variable selection purposes the advantage of the random forest permutation accuracy importance measure as compared to univariate screening methods is that it covers the impact of each predictor variable individually as well as in multivariate

interactions with other predictor variables. For example, Lunetta et al.[11] 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 like Fisher's exact test[11]. Here the random forest implementation *cforest* from the package *party*[12, 13] in the R system for statistical computing[14] is used, because it guarantees unbiased variable selection for predictor variables of different scales of measurement[10]. Predictor variables considered here were age, gender and respective SNP variables.

To assess the stability of the results 100 random forests with 500 trees each were fitted with the configuration guaranteeing unbiased variable selection suggested by Strobl et al.[10]. The random forests were built with either 2 randomly pre-selected variables in each split (argument mtry=2) or 3 randomly pre-selected 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 where the number of randomly pre-selected 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 three study cohorts to increase the power of detecting any single SNP and interaction effect. As British and German populations might be slightly different with regard to ethnicity[15] we accounted for a population effect in every analysis approach by introducing a binary variable which codes 1="British/Irish origin" and 0="German origin". Thus we corrected the estimated genetic effect for potential population differences. Power calculations[16] 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[14], unless otherwise stated.

Table E1: Descriptive characterization of cases and control populations. N.a., not available

	n	Mean age in yrs. (std)	Male gender	Mean IgE (std)	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 controls	552	35.71 (12.27)	170 (30.8%)	n.a.	Ireland	F
ALSPAC controls	6063	3.5*	3135 (51.7%)	200.4 (462.7)	England	G
Sum controls	10607					

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

Table E2: Genotyping details and minor allele frequencies in the KORA S4 population-based cohort. MAF, minor allele frequency; DIR, direction

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	

References:

1. 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.
2. 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.
3. 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 2006.
4. Foster MR. Key concepts in model selection: Performance and generalizability. J Math Psychol 2000; 44:205-31.
5. 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.
6. Kneib T, Baumgartner B, Steiner WJ. Semiparametric Multinomial Logit Models for Analysing Consumer Choice Behaviour. AStA Advances in Statistical Analysis 2007; 91:225-44.
7. Brezger A, Kneib T, Lang S. BayesX - Software for Bayesian Inference in Structured Additive Regression Models. Munich, 2007.
8. Breiman L. Arcing classifers. The Annals of Statistics 1998; 26:801-49.
9. Breiman L. Random forests. Machine Learning 2001; 45:5-32.
10. 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.
11. 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.
12. Hothorn T, Hornik K, Zeileis A. Unbiased recursive partitioning: a conditional inference framework. Journal of Computational and Graphical Statistics 2006; 15:651-74.
13. Hothorn T, Hornik K, Zeileis A. Party: A laboratory for recursive part(y)itioningR package version 0.9-0. 2006.
14. Team RDC. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing, 2007.
15. 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.
16. Fleiss JL, Tytun A, Ury HK. A Simple Approximation for Calculating Sample Sizes for Comparing Independent Proportions. Biometrics 1980; 36:343-6.