Expression of the filaggrin gene in umbilical cord blood predicts eczema risk in infancy: a birth cohort study

Ali H. Ziyab[1], **Susan Ewart**[2], **Gabrielle A. Lockett**[3], **Hongmei Zhang**[4], **Hasan Arshad**[5,6], **John W. Holloway**[3,6], and **Wilfried Karmaus**[4]

[1]Departments of Community Medicine and Behavioral Sciences, Faculty of Medicine, Kuwait University, Kuwait

[2]College of Veterinary Medicine, Michigan State University, East Lansing, Michigan, USA

[3]Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, UK

[4]Division of Epidemiology, Biostatistics and Environmental Health, School of Public Health, University of Memphis, Memphis, Tennessee, USA

[5]David Hide Asthma and Allergy Research Centre, St Mary's Hospital, Isle of Wight, UK

[6]Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, UK

Abstract

Background—Filaggrin gene (*FLG*) expression, particularly in the skin, has been linked to the development of the skin barrier and is associated with eczema risk. However, knowledge as to whether *FLG* expression in umbilical cord blood (UCB) is associated with eczema development and prediction is lacking.

Objective—This study sought to assess whether *FLG* expression in UCB associates with and predicts the development of eczema in infancy.

Methods—Infants enrolled in a birth cohort study (n = 94) were assessed for eczema at ages 3–, 6–, and 12-months. Five probes measuring *FLG* transcripts expression in UCB were available from genome-wide gene expression profiling. *FLG* genetic variants R501X, 2282del4, and S3247X were genotyped. Associations were assessed using Poisson regression with robust variance estimation. Area under the curve (AUC), describing the discriminatory/predictive performance of fitted models, was estimated from logistic regression.

Results—Increased level of *FLG* expression measured by probe A_24_P51322 was associated with reduced risk of eczema during the first year of life (RR = 0.60, 95% CI: 0.38–0.95). In contrast, increased level of *FLG* antisense transcripts measured by probe A_21_P0014075 was associated with increased risk of eczema (RR = 2.02, 95% CI: 1.10–3.72). In prediction models

Corresponding author: Ali H. Ziyab, PhD, Departments of Community Medicine and Behavioral Sciences, Faculty of Medicine, Kuwait University, Kuwait, P. O. Box 24923, Safat 13110, Kuwait, Tel: (+965) 24636545, Fax: (+965) 25338948, aziyab@hsc.edu.kw.

Conflicts of interest

All authors declare that they have no conflicts of interest.

including *FLG* expression, *FLG* genetic variants, and sex, discrimination between children who will and will not develop eczema at 3-months of age was high (AUC: 0.91, 95% CI: 0.84–0.98).

Conclusions and Clinical Relevance—This study demonstrated, for the first time, that *FLG* expression in UCB is associated with eczema development in infancy. Moreover, our analysis provided prediction models that were capable of discriminating, to a great extent, between those who will and will not develop eczema in infancy. Therefore, early identification of infants at increased risk of developing eczema is possible and such high-risk newborns may benefit from early stratification and intervention.

Introduction

Eczema (atopic dermatitis) is a common chronic inflammatory skin disorder that is often seen in early life and affects up to 30% of children worldwide [1]. In addition to social and economic burden associated with the disease, early-life eczema has been linked to the subsequent development of asthma and rhinitis, a concept termed 'the atopic march' [2, 3]. Therefore, eczema is a candidate for a disease-modifying strategy that aims at breaking, stopping, or reversing the course of atopic disease [4]. There is an urgent need for biomarkers that can identify children at increased risk of developing eczema who may substantially benefit from early treatment and preventive strategies.

Eczema is characterized by epidermal barrier defects and immunologic dysregulation [5, 6]. Filaggrin (filament-aggregating protein) is a major structural protein of the epidermal barrier and its insufficiency underlies the pathogenesis of eczema [7, 8]. Filaggrin gene (*FLG*) variants are the strongest and most replicated genetic risk factor for eczema development, in particular for eczema that develops in infancy [9–11]. Reduced expression of filaggrin in skin of eczema-cases compared to controls has been reported [12, 13].

Despite the strong association between *FLG* genetic variants and eczema, only around 30% of eczema patients carry *FLG* variants and approximately 10% of eczema cases can be attributed to *FLG* variants [1, 10]. Hence, *FLG* variants alone are not sensitive predictors of eczema development. On the other hand, family history of allergic disease and umbilical cord blood (UCB) immunoglobulin E (IgE) levels have been widely speculated as predictors of eczema; however, their predictive value remains controversial [14–16]. Hence, there is lack of reliable screening biomarker for eczema that identifies individuals at high risk.

Recent studies have shown that epidermal barrier enhancement in early infancy, through application of emollient and moisturizer, is associated with reduced incidence of eczema [17, 18]. Such simple interventions could be applied to children at increased risk of eczema. Filaggrin, due to its essential role in the formation of the epidermal barrier, is a potential candidate that can serve as a screening biomarker for eczema. Filaggrin is expressed in the skin, oral mucosa, conjunctivae, esophagus, and cervix, but assessments of its expression in blood cells have been limited [19, 20]. To test the predictive value of a novel biomarker, this study sought to (i) measure expression levels of *FLG* RNA transcripts in UCB and (ii) assess whether *FLG* RNA transcripts levels in UCB associate with and predict the development of eczema at ages 3–, 6–, and 12-months. To achieve the objectives of this study, data from a prospective birth cohort study were analyzed.

Methods

Study design and participants

A second generation (F2) birth cohort was established based on the Isle of Wight (IOW), UK, 1989 birth cohort (F1) to study the transgenerational inheritance of epigenetic markers and the developmental origins of allergic disorders. Pregnant women of the 1989 IOW birth cohort, and pregnant partners of male members of the cohort, were enrolled in the F2 IOW birth cohort during gestation. At birth, written informed consent was obtained from parents to enroll 351 newborns (F2 participants), with follow-up assessments conducted at 3–, 6–, and 12-months of age. Ethics approvals were obtained from the respective ethics committees (Southampton & South West Hampshire Research Ethics Committee) at recruitment and for subsequent follow-ups (09/H0504/129). Questionnaires capturing maternal demographic information and atopy status during pregnancy were completed during gestation. Detailed questionnaires, including questionnaires from the international study of asthma and allergies in childhood (ISAAC) [21], were completed for each child at the follow-ups.

Eczema definition

In all assessments (3–, 6–, and 12-month follow-ups), eczema was defined as chronic or chronically relapsing, itchy dermatitis lasting more than 6 weeks with characteristic morphology and distribution, following Hanifin and Rajka criteria [22].

FLG genotyping

DNA was extracted from UCB and was genotyped for three *FLG* null variants common among Europeans: R501X, S3247X, and 2282del4. Genotyping was performed using TaqMan allelic discrimination assays [23], with PerfeCTa mastermix (VWR International) and 5 ng DNA per sample. Control samples of known genotype were included to allow endpoint genotype determination. Individuals carrying the minor allele for at least one of the *FLG* variants were classified as filaggrin haploinsufficiency.

FLG gene expression

UCB was collected into PAXgene Bone Marrow RNA tubes (Qiagen, Valencia, CA, USA), from which total RNA was isolated according to the manufacturer's instructions. Total RNA yield and the absence of DNA contamination was checked using a Qubit 2.0 fluorometer (Life Technologies, Grand Island, NY, USA). RNA quality was assessed measuring the 28S/18S rRNA and the RNA Integrity Number (RIN) with a Bioanalyzer 2100, using RNA 6000 Nano Chips (Agilent Technologies, Santa Clara, CA, USA). RNA samples with RIN values > 8, 260/280 absorbance ratios >1.8 and 260/230 absorbance ratios >1.5 were considered suitable for microarray analysis.

A total of 500 ng of each RNA sample was reverse transcribed into cDNA, and subsequently amplified and labeled with Cy5 dye following Agilent's Single-Color Microarray-Based Gene Expression Analysis protocol version 6.0. The same amount of RNA from a commercially-available pool of human leucocyte total RNA (Clontech, Mountain View, CA, USA) was reverse transcribed, amplified and Cy3/5-labeled to be used as reference. Spike-in RNA (Agilent Technologies) was used as internal control. RNA references and samples

were labeled separately and then hybridized together on SurePrint G3 Human GE v2 8×60K Agilent Microarray slides (Agilent Technologies). Slides were scanned with an Agilent's G2565AA Microarray Scanner System. Dye-normalized, background-subtracted log-ratios of sample to reference expression were calculated using Agilent's Feature Extraction Software version 9.5. Hybridization quality was checked using the software's quality report. The reported expression levels are log base 2 intensities normalized at the 75% percentile shift over all samples. A subset of probes measuring expression levels of *FLG* RNA transcripts were examined in this study; Probes A_24_P51322, A_32_P387648, and A_33_P3261328 measured *FLG* mRNA (sense) transcripts and probes A_21_P0014075 and A_33_P3296200 measured antisense *FLG* transcripts.

Statistical analysis

Analyses were conducted using SAS 9.4 (SAS Institute, Cary, North Carolina, USA). The statistical significance level was set to $\alpha = 0.05$ for all association analyses. To assess whether the analytical study sample (n = 94, participants with *FLG* gene expression and *FLG* genotype information) was representative of the total cohort (n=351), proportions of categorical variables were compared using chi-square (χ^2) tests across these two samples. Associations between *FLG* genetic variants and *FLG* expression levels were assessed using the Wilcoxon rank sum test. Spearman correlation coefficients between the five probes measuring *FLG* expression were estimated.

Since eczema (the outcome variable) was repeatedly measured at ages 3–, 6–, and 12-months, generalized estimating equations (GEE) were applied [24]. First-order autoregressive covariance matrix was implemented to account for correlated repeated measurements when assessing the association between *FLG* variants and *FLG* expression with eczema over time (i.e., eczema status at ages 3–, 6–, and 12-months). The GEE approach is a useful method for the analysis of repeated measurements data (longitudinal data), especially when the outcome variable is dichotomous and the repeated outcomes are correlated. The regression coefficient estimates (e.g., risk ratio) retuned by the GEE approach are known to be population-averaged estimates [24]. Hence, since eczema was assessed at ages 3–, 6–, and 12-months, our interpretations of the GEE estimates will pertain to the first 12-months of life. Also, the GEE approach increases the statistical power due to the fact that all available measurements are analyzed. Risk ratios (RR) and their 95% confidence intervals (95% CI) were estimated by applying the modified Poisson regression with robust variance estimation using the GENMOD procedure in SAS 9.4 [25, 26]. In all GEE models, sex and age at follow-up were included as potential confounders.

Binary logistic regression models were fitted to assess the ability of *FLG* expression measured in UCB in predicting (discriminating) those who will (true cases) and will not (true non-cases) develop eczema at ages 3–, 6–, and 12-months. Receiver operating characteristics (ROC) curves, plotting sensitivity (true positives) by 1 – specificity (false positives) at different cut-points, were used to assess the discriminatory performance of models with different predictors. The area under the ROC curve (AUC) was estimated, providing a quantitative summary measure on the discriminatory performance of the fitted models. For instance, a model with AUC of 0.5 has no discrimination value, whereas a

model with AUC of 1.0 has perfect discrimination [27]. AUC values and their 95% CIs were estimated using the LOGISTIC procedure in SAS 9.4.

Sensitivity analysis

To ensure that the predictions models in which AUC was estimated were not over-fitted and that the estimated AUC were not over-estimated, internal cross-validation using the leave-one-out principle was applied. In this cross-validation method, the prediction model is fitted N (number of subjects in the dataset) times by dropping the data of one subject and re-estimating the parameter estimates using the rest of the data, producing an overall cross-validated estimate using parameter estimates from the N cross-validation subsets.

Results

Description of study population

A total of 351 newborns were enrolled and followed-up at 3–, 6–, and 12-months of age. In this report we focused on 94 participants (analytical study sample) that had information on *FLG* gene expression and genotypes. The total study sample (n = 351) and the analytical sample (n = 94) were similar with respect to all characteristics under study (Table 1). The prevalence of eczema was 10.2%, 20.5%, and 16.3% at ages 3–, 6–, and 12-months, respectively. The combined proportion of carriers of *FLG* loss-of-function variants R501X, 2282del4, or S3247X was 6.4% (Table 1).

FLG variants and eczema risk

The association between *FLG* variants with eczema at ages 3–, 6–, and 12-months and the repeated measurements of eczema were evaluated (Table 2). In the repeated measurements analysis, *FLG* variants were associated with 2.94-fold (95% CI: 1.33 – 6.51) increased risk of eczema during the first 12-months of life (Table 2). In the analytical sample we have 39 repeated measurements of infants with eczema and 212 repeated measurements without. These will be used in further repeated analyses described as k = 39 and k = 212.

Correlations between FLG expression levels

Five probes measuring expression levels of *FLG* transcripts were available from the genome-wide study of gene expression. Spearman correlation coefficients between the five probes were estimated (Table S1). There was a robust correlation between transcript levels measured by probes A_32_P387648 (sense transcript) and A_33_P3296200 (antisense transcript) (r = 0.86, *P* < 0.001; Table S1). In contrast, correlations between the others probes were weaker (r < 0.5).

FLG variants and FLG expression

Associations were evaluated to determine whether *FLG* variants influence expression of *FLG* transcripts (Table 3). The median *FLG* expression level of probe A_24_P51322 in children with *FLG* variants was significantly lower compared to children with the wild-type *FLG* genotype (−0.97 vs. 0.16; *P* = 0.007). Similarly, the expression level of antisense transcripts measured by probe A_33_P3296200 was lower among children carrying *FLG*

variants compared to children with wild-type *FLG* genotype (−0.43 vs. 0.09; *P* = 0.031; Table 3).

FLG expression and the future risk of eczema in infancy

The association between *FLG* expression levels in UCB and the risk of eczema in infancy was analyzed using the repeated measurements of eczema (3−, 6−, and 12-months). Note we have 39 repeated measurements of eczema and 212 measurements without eczema (Table 4). In a model adjusted for sex, increased levels of *FLG* mRNA expression measured by A_24_P51322 were significantly associated with a reduced risk of eczema during the first year of life (RR = 0.54, 95% CI: 0.33 – 0.89, *P* = 0.015; Table 4). This reduced risk of eczema remained after further adjustment for *FLG* variants (RR = 0.60, 95% CI: 0.38 – 0.95, *P* = 0.047). In contrast, increased levels of the *FLG* antisense transcripts measured by A_21_P0014075 were significantly associated with increased risk of eczema (sex-adjusted RR = 2.02, 95% CI: 1.20 – 3.41, *P* = 0.008; Table 4). This increased risk of eczema was also not attenuated after adjusting for *FLG* variants (RR = 2.02, 95% CI: 1.10 – 3.72, *P* = 0.024). In a regression model that included both significant biomarkers (A_24_P51322 and A_21_P0014075) plus sex and *FLG* variants as covariates, increased expression levels measured by A_24_P51322 (sense transcripts) were still associated with reduced risk of eczema (RR = 0.59, 95% CI: 0.35 – 0.99, *P* = 0.048) and increased levels of A_21_P0014075 (antisense transcripts) were associated with increased risk of eczema (RR = 2.24, 95% CI: 1.13 – 4.41, *P* = 0.020). Hence, simultaneously adjusting for the effect of both biomarkers did not attenuate their independent effects.

Prediction of eczema risk in infancy by FLG expression

The predictive value (discrimination performance) of the two biomarkers (A_24_P51322 and A_21_P0014075) that were significantly associated with eczema was evaluated by estimating the corresponding AUC using different statistical models. In general, AUC estimates were better at predicting eczema at 3-months than at 6− and 12-months of age (Table 5). The AUC for discriminating those who will and will not develop eczema at 3-months of age improved from 0.72 (95% CI: 0.53 – 0.91) in the model including only A_24_P51322 (model 1) to 0.88 (95% CI: 0.79 – 0.97) in the model adjusting for sex and *FLG* variants (model 4; Table 5). Similarly, the AUC of the crude model including only A_21_P0014075 (model 5) increased from 0.63 (95% CI: 0.45 – 0.80) to 0.91 (95% CI: 0.84 – 0.98) in the model accounting for sex and *FLG* variants (model 8; Table 5). Hence, our data indicate that models 4 and 8 predict the development of eczema at 3-months with high accuracy.

To ensure that models 4 and 8 (Table 5) are not over-fitted and the estimated AUCs are not over-estimated, internal cross-validation was performed. The cross-validated AUC estimate of model 4 was 0.82 (95% CI: 0.71 – 0.93), which is slightly lower than the naïve AUC estimate of 0.88 (95% CI: 0.79 – 0.97; Table 5). Similarly for model 8, the cross-validated AUC was estimated to be 0.88 (0.79 – 0.95), which is slightly lower than the naïve AUC estimate of 0.91 (95% CI: 0.84 – 0.98; Table 5). Hence, findings from cross-validation indicate no substantial over-estimation of the discriminatory performance of the prediction models.

Discussion

The identification of biomarkers that enable the early identification and stratification of individuals at high risk for future disease development, coupled with a cost-effective primary prevention could substantially reduce the disease burden. Thus, we investigated the association and predictive value of *FLG* gene expression measured in UCB on the development of eczema in infancy. Our study showed that the mean expression of *FLG* mRNA (sense) transcripts measured by A_24_P51322 was significantly lower in children who developed eczema compared to those who did not develop eczema during the first 12-months of life (RR = 0.60, P = 0.047). In contrast, the mean expression of *FLG* antisense transcripts measured by A_21_P0014075 was significantly higher in children who developed eczema compared to those who did not develop eczema during the follow-up period (RR = 2.02, P = 0.024). Both biomarkers (A_24_P51322 and A_21_P0014075) demonstrated similar capability and high accuracy in predicting eczema at 3-months of age. The AUC for a model including A_24_P51322 plus *FLG* genetic variants and sex was estimated to be 0.88 (95% CI: 0.79 – 0.97). In a model including A_21_P0014075 plus *FLG* genetic variants and sex, the AUC was estimated to be 0.91 (95% CI: 0.84 – 0.98). Our study is the first to demonstrate the predictive value of novel biomarkers for eczema, which can be used in the development of a new screening test that aims at early identification of newborns susceptible for the development of eczema and who may benefit from early intervention.

Prior studies have investigated various markers in predicting the risk of eczema. Family history of atopy and elevated UCB IgE are well-established risk factors for eczema; however, reports of their predictive value are contradictory [14, 28]. More recently, the role of skin barrier function indices (e.g., skin pH, transepidermal water loss, and stratum corneum hydration) in the development of eczema have been investigated [29–31]. For instance, elevated transepidermal water loss measured at 2-days and 2-months after birth were associated with 7.1-fold (P = 0.001) and 5.6-fold (P = 0.001) increased risk of eczema at 12-months of age, respectively [30]. A multivariable model including measurements of transepidermal water loss at 2-days after birth and several other factors (e.g., parental atopy, *FLG* loss-of-function mutation, use of emollient, sex) estimated the AUC to be 0.83 (95% CI: 0.7 – 0.9) for discriminating eczema cases from non-cases at age 12-months [30]. Another study evaluated the predictive value of a model including UCB IgE, genotypes of cytokine-related genes, maternal stress level during gestation, and socio-demographic factors, which estimated the AUC to be 0.73 (95% CI: 0.70 – 0.76) for predicting eczema at age 2-years [32]. Our analysis, when compared to the aforementioned studies, provided prediction models (model 4 and model 8; Table 5) that used fewer variables (biomarkers) and yielded better discrimination (AUC = 0.88, 95% CI: 0.79 – 0.97; AUC = 0.91, 95% CI: 0.84 – 0.98) between those who will and will not develop eczema as early as 3-months of age.

To determine the importance of predicting eczema at 3 months of age, we assessed the extent that eczema at 3 months of age predicts subsequent eczema development. Results of this analysis indicate that 100.0% (9/9) of infants with eczema at 3 months continued to have eczema at 6 months of age (RR = 8.25, 95% CI: 3.29 – 20.67). Similarly, 77.8% (7/9) of infants with eczema at 3 months continued to have eczema at 12 months of age (RR = 8.30,

95% CI: 2.67 – 25.75; data not shown). Hence, eczema at 3 months of age is a strong predictor of subsequent/persistent eczema. Moreover, existing literature demonstrates that early eczema (onset before 2 years) is associated with subsequent/persistent eczema manifestations [33, 34]. Furthermore, it has been widely demonstrated that eczema in infancy is a predictor of subsequent development of rhinitis and asthma [35, 36]. Therefore, developing models that can predict eczema at 3 months of age with high accuracy will guide early stratification and intervention, which could modify the course of the disease in a high-risk population.

Although *FLG* variants are the strongest and most replicated risk factors for the development of eczema, their value in predicting and explaining the burden of eczema has been limited [10, 37]. This study builds on the evidence that filaggrin expression is modulated, regardless of *FLG* variants, in eczema patients compared to healthy controls [12, 38]. Such an observation further highlights the importance of this structural protein for the development of functional skin barrier and the pathogenesis of eczema [39]. We demonstrated that the predictive value of *FLG* expression measured in UCB is slightly improved when *FLG* genetic variants were added to the prediction models (see Table 5, model 2 vs. model 4 and model 6 vs. model 8). Therefore, considering the expression levels as well as genetic variants of *FLG* gene yielded improved discriminatory performance. However, when excluding participants carrying *FLG* haploinsufficiency variants from the analysis, the prediction performance of models including *FLG* expression variable (A_24_P51322 or A_21_P0014075) and sex were not compromised (data not shown). Hence, further indicating the *FLG* expression is associated with and can predict eczema independently of *FLG* haploinsufficiency variants.

A recent study found that *FLG* variants status in mothers increased the risk of eczema in their children, even if the child did not inherit the genetic defect [40]. Therefore, to further corroborate this finding, we ran additional prediction models that included infant *FLG* expression variable (A_24_P51322 or A_21_P0014075) plus infant sex and maternal *FLG* variants status; of the 64 mothers with *FLG* genotype information, 5 (7.8%) carried *FLG* variants. We compared these results to models that used infant *FLG* variants status (models 4 and 8; Table 5). Results of these analysis did not demonstrate any substantial improvement in prediction capability (data not shown). However, since models with maternal *FLG* variants status performed as well as models that included infant *FLG* variants status, our results further support the previous observation that maternal *FLG* status has an independent effect on child's eczema risk prediction.

The results show that an increase in the mRNA (sense) transcripts measured by A_24_P51322 was associated with reduced risk of eczema (RR = 0.60, P = 0.047). In contrast, increased expression of antisense transcripts measured by A_21_P0014075 was associated with increased risk of eczema (RR = 2.02, P = 0.024). The observed opposite directions of association are in agreement with our expectation and the current state of knowledge, which suggests that increased *FLG* mRNA transcript levels should essentially correlate with higher filaggrin protein expression and subsequently a well-developed skin barrier; thus reduced risk of eczema. On the other hand, emerging scientific evidence suggest that antisense transcription is an important regulatory element of gene expression

rather than just transcriptional noise, which is capable of repressing the transcript of origin, suppressing transcription initiation, and increasing the stability of their target sense mRNAs [41, 42].

We did not find consistent correlation trends between the sense and antisense transcripts (Table S1). This may be due, in part, to the genomic physical proximity of the transcripts to each other. For instance, the genomic coordinates of the antisense transcript A_33_P3296200 place it closer to the sense transcripts than to the other antisense transcript A_21_P0014075 (Table S2). Moreover, due to the observed positive correlations between the antisense transcript A_33_P3296200 with the sense transcripts (Table S1), it could be that this antisense transcript is involved in increasing the stability of its target sense mRNA transcripts and thus demonstrating similar expression pattern as the sense transcripts [41, 42]. Whereas, the antisense transcript A_21_P0014075, which did not show correlation with the sense transcripts (Table S1), might be a suppressor of transcription initiation or act as a repressor of sense transcription. Therefore, the antisense transcript A_33_P3296200 might be associated with increased filaggrin protein levels, whereas the antisense transcript A_21_P0014075 could be associated with decreased filaggrin protein levels. These proposed explanations are plausible; however, laboratory-based investigation are needed to confirm and determine the precise role of antisense transcripts on filaggrin expression.

The prospective nature of our study and the repeated measurements during the three follow-ups (ages 3−, 6, and 12-months) allowed us to assess temporal associations. Moreover, measuring *FLG* gene expression in UCB and genotyping *FLG* genetic variants added to the strengths of the current study. To our knowledge, this is the first study that has measured and assessed the association of *FLG* expression in UCB with eczema in infancy. A limitation of our study is that 94 participants out of the total study sample (n = 351) had information on *FLG* gene expression and *FLG* genotypes; however, there was no indication of selection (Table 1). Although the group of infants carrying *FLG* variants (n = 6) is small, the present report focuses on *FLG* gene expression and only controls for *FLG* genetic variants. In addition, investigations of the *FLG* gene expression go beyond the analysis of genetic variants, since some children without *FLG* haploinsufficiency have a lack of *FLG* gene expression (Table 3). Hence, *FLG* genetic variants are considered to be confounders and not the main exposure in the current study. *FLG* gene expression was quantified using an Agilent genome-wide microarray (SurePrint G3), which included five probes to measure *FLG* transcripts. The presence of multiple potential splice sites leading to alternative splicing and thus potentially different RNA transcript isoforms of *FLG* can be explained by the large and repetitive nature of *FLG* gene, which also explains the difference between multiple transcription probes [43]. Although the observed associations were clinically and statistically relevant, lack of quantifying the measured transcripts using a polymerase chain reaction (PCR)-based method, which is needed to further validate the microarray measurements, is another limitation to our study. However, searching the Gene Expression Omnibus (GEO) database [44] revealed that prior studies were able to detect *FLG* expression signals in cord blood (GEO accession: GDS3929, ID: 73097607 [45]; GEO accession: GDS2655, ID: 37270577 [46]; GEO accession: GDS3401, ID: 55838377 [47]). Such observations further support our detection of *FLG* expression in cord blood.

In prediction models including the *FLG* expression variable (A_24_P51322 or A_21_P0014075), *FLG* genetic variants, and sex, discrimination between children who will and will not develop eczema at 3-months was high (AUC = 0.88, 95% CI: 0.79 – 0.97; AUC = 0.91, 95% CI: 0.84 – 0.98). The discrimination performance of these models were not considerably compromised by internal cross-validations (cross-validated model 4: AUC = 0.82, 95% CI: 0.71 – 0.93; cross-validated model 8: AUC = 0.88, 95% CI: 0.79 – 0.95). Thus, a screening test based on the current prediction models has strong potential to produce high prediction accuracy in future risk of eczema. These findings provide evidence that early identification of infants at increased risk of eczema is possible, which will enable early stratification and intervention to prevent and modify the course of the disease in a high-risk population. Therefore, early identification of the at-risk population with high accuracy and precision coupled with the existence of cost-effective preventive strategies (e.g., use of moisturizes and emollients) is a translational step in the research field of eczema that will reduce the associated public health burden.

In conclusion, this study demonstrated, for the first time, the association between *FLG* gene expression measured in UCB and the risk of subsequent eczema. Moreover, our analysis provided prediction models that were capable of discriminating, to a great extent, between those who will and will not develop eczema during infancy. Therefore, future studies measuring *FLG* expression in UCB and corroborating our findings are needed, and if confirmed, will contribute to early detection of infants at higher risk of eczema who may substantially benefit from early intervention.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Bieber T. Atopic dermatitis. N Engl J Med. 2008; 358:1483–94. [PubMed: 18385500]

2. Dharmage SC, Lowe AJ, Matheson MC, et al. Atopic dermatitis and the atopic march revisited. Allergy. 2014; 69:17–27. [PubMed: 24117677]

3. Suh DC, Sung J, Gause D, et al. Economic burden of atopic manifestations in patients with atopic dermatitis--analysis of administrative claims. J Manag Care Pharm. 2007; 13:778–89. [PubMed: 18062729]

4. Bieber T, Cork M, Reitamo S. Atopic dermatitis: a candidate for disease-modifying strategy. Allergy. 2012; 67:969–75. [PubMed: 22671706]

5. Brown SJ, McLean WH. Eczema genetics: current state of knowledge and future goals. J Invest Dermatol. 2009; 129:543–52. [PubMed: 19209157]

6. Boguniewicz M, Leung DY. Atopic dermatitis: a disease of altered skin barrier and immune dysregulation. Immunol Rev. 2011; 242:233–46. [PubMed: 21682749]

7. McAleer MA, Irvine AD. The multifunctional role of filaggrin in allergic skin disease. J Allergy Clin Immunol. 2013; 131:280–91. [PubMed: 23374260]

8. Cabanillas B, Novak N. Atopic dermatitis and filaggrin. Curr Opin Immunol. 2016; 42:1–8. [PubMed: 27206013]

9. van den Oord RA, Sheikh A. Filaggrin gene defects and risk of developing allergic sensitisation and allergic disorders: systematic review and meta-analysis. BMJ. 2009; 339:b2433. [PubMed: 19589816]

10. Rodriguez E, Baurecht H, Herberich E, et al. Meta-analysis of filaggrin polymorphisms in eczema and asthma: robust risk factors in atopic disease. J Allergy Clin Immunol. 2009; 123:1361–70. e7. [PubMed: 19501237]

11. Rupnik H, Rijavec M, Korosec P. Filaggrin loss-of-function mutations are not associated with atopic dermatitis that develops in late childhood or adulthood. Br J Dermatol. 2015; 172:455–61. [PubMed: 25314673]

12. Thyssen JP, Kezic S. Causes of epidermal filaggrin reduction and their role in the pathogenesis of atopic dermatitis. J Allergy Clin Immunol. 2014; 134:792–9. [PubMed: 25065719]

13. Batista DI, Perez L, Orfali RL, et al. Profile of skin barrier proteins (filaggrin, claudins 1 and 4) and Th1/Th2/Th17 cytokines in adults with atopic dermatitis. J Eur Acad Dermatol Venereol. 2015; 29:1091–5. [PubMed: 25271795]

14. Nissen SP, Kjaer HF, Host A, et al. Can family history and cord blood IgE predict sensitization and allergic diseases up to adulthood? Pediatr Allergy Immunol. 2015; 26:42–8. [PubMed: 25420698]

15. Allam JP, Zivanovic O, Berg C, et al. In search for predictive factors for atopy in human cord blood. Allergy. 2005; 60:743–50. [PubMed: 15876303]

16. Hide DW, Arshad SH, Twiselton R, et al. Cord serum IgE: an insensitive method for prediction of atopy. Clin Exp Allergy. 1991; 21:739–43. [PubMed: 1777834]

17. Simpson EL, Chalmers JR, Hanifin JM, et al. Emollient enhancement of the skin barrier from birth offers effective atopic dermatitis prevention. J Allergy Clin Immunol. 2014; 134:818–23. [PubMed: 25282563]

18. Horimukai K, Morita K, Narita M, et al. Application of moisturizer to neonates prevents development of atopic dermatitis. J Allergy Clin Immunol. 2014; 134:824–30. e6. [PubMed: 25282564]

19. De Benedetto A, Qualia CM, Baroody FM, et al. Filaggrin expression in oral, nasal, and esophageal mucosa. J Invest Dermatol. 2008; 128:1594–7. [PubMed: 18172455]

20. Cintorino M, Syrjanen S, Leoncini P, et al. Altered expression of filaggrin in human papillomavirus (HPV) lesions of the uterine cervix. Arch Gynecol Obstet. 1988; 241:235–47. [PubMed: 2835017]

21. Asher MI, Keil U, Anderson HR, et al. International Study of Asthma and Allergies in Childhood (ISAAC): rationale and methods. Eur Respir J. 1995; 8:483–91. [PubMed: 7789502]

22. Hanifin JM, Rajka G. Diagnostic features of atopic dermatitis. Acta Derm Venereol Suppl (Stockh). 1980; 92:44–47.

23. Kezic S, O'Regan GM, Yau N, et al. Levels of filaggrin degradation products are influenced by both filaggrin genotype and atopic dermatitis severity. Allergy. 2011; 66:934–40. [PubMed: 21261659]

24. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. Biometrics. 1986; 42:121–30. [PubMed: 3719049]

25. Zou G. A modified poisson regression approach to prospective studies with binary data. Am J Epidemiol. 2004; 159:702–6. [PubMed: 15033648]

26. Yelland LN, Salter AB, Ryan P. Performance of the modified Poisson regression approach for estimating relative risks from clustered prospective data. Am J Epidemiol. 2011; 174:984–92. [PubMed: 21841157]

27. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. Circulation. 2007; 115:928–35. [PubMed: 17309939]

28. Bohme M, Wickman M, Lennart Nordvall S, et al. Family history and risk of atopic dermatitis in children up to 4 years. Clin Exp Allergy. 2003; 33:1226–31. [PubMed: 12956743]

29. Flohr C, England K, Radulovic S, et al. Filaggrin loss-of-function mutations are associated with early-onset eczema, eczema severity and transepidermal water loss at 3 months of age. Br J Dermatol. 2010; 163:1333–6. [PubMed: 21137118]

30. Kelleher M, Dunn-Galvin A, Hourihane JO, et al. Skin barrier dysfunction measured by transepidermal water loss at 2 days and 2 months predates and predicts atopic dermatitis at 1 year. J Allergy Clin Immunol. 2015; 135:930–5. e1. [PubMed: 25618747]

31. Horimukai K, Morita K, Narita M, et al. Transepidermal water loss measurement during infancy can predict the subsequent development of atopic dermatitis regardless of filaggrin mutations. Allergol Int. 2016; 65:103–8. [PubMed: 26666481]

32. Wen HJ, Wang YJ, Lin YC, et al. Prediction of atopic dermatitis in 2-yr-old children by cord blood IgE, genetic polymorphisms in cytokine genes, and maternal mentality during pregnancy. Pediatr Allergy Immunol. 2011; 22:695–703. [PubMed: 21539617]

33. Illi S, von Mutius E, Lau S, et al. The natural course of atopic dermatitis from birth to age 7 years and the association with asthma. J Allergy Clin Immunol. 2004; 113:925–31. [PubMed: 15131576]

34. Burr ML, Dunstan FD, Hand S, et al. The natural history of eczema from birth to adult life: a cohort study. Br J Dermatol. 2013; 168:1339–42. [PubMed: 23301666]

35. Goksor E, Loid P, Alm B, et al. The allergic march comprises the coexistence of related patterns of allergic disease not just the progressive development of one disease. Acta Paediatr. 2016; 105:1472–79. [PubMed: 27381249]

36. Hopper JL, Bui QM, Erbas B, et al. Does eczema in infancy cause hay fever, asthma, or both in childhood? Insights from a novel regression model of sibling data. J Allergy Clin Immunol. 2012; 130:1117–22. e1. [PubMed: 23021888]

37. Weidinger S, Novak N. Atopic dermatitis. Lancet. 2016; 387:1109–22. [PubMed: 26377142]

38. Peng W, Novak N. Pathogenesis of atopic dermatitis. Clin Exp Allergy. 2015; 45:566–74. [PubMed: 25610977]

39. Armengot-Carbo M, Hernandez-Martin A, Torrelo A. The role of filaggrin in the skin barrier and disease development. Actas Dermosifiliogr. 2015; 106:86–95. [PubMed: 24674607]

40. Esparza-Gordillo J, Matanovic A, Marenholz I, et al. Maternal filaggrin mutations increase the risk of atopic dermatitis in children: an effect independent of mutation inheritance. PLoS Genet. 2015; 11:e1005076. [PubMed: 25757221]

41. Pelechano V, Steinmetz LM. Gene regulation by antisense transcription. Nat Rev Genet. 2013; 14:880–93. [PubMed: 24217315]

42. Esteller M. Non-coding RNAs in human disease. Nat Rev Genet. 2011; 12:861–74. [PubMed: 22094949]

43. Sandilands A, Sutherland C, Irvine AD, et al. Filaggrin in the frontline: role in skin barrier function and disease. J Cell Sci. 2009; 122:1285–94. [PubMed: 19386895]

44. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res. 2013; 41:D991–5. [PubMed: 23193258]

45. Votavova H, Dostalova Merkerova M, Fejglova K, et al. Transcriptome alterations in maternal and fetal cells induced by tobacco smoke. Placenta. 2011; 32:763–70. [PubMed: 21803418]

46. Goh SH, Josleyn M, Lee YT, et al. The human reticulocyte transcriptome. Physiol Genomics. 2007; 30:172–8. [PubMed: 17405831]

47. Markov V, Kusumi K, Tadesse MG, et al. Identification of cord blood-derived mesenchymal stem/ stromal cell populations with distinct growth kinetics, differentiation potentials, and gene expression profiles. Stem Cells Dev. 2007; 16:53–73. [PubMed: 17348805]

Table 1

Characteristics of total study population and analytical study sample

Variable	Enrolled study sample (n = 351)	Analytical study sample[*](n = 94)	P value
Sex, % (n)			
Male	55.8 (196)	51.1 (48)	0.409
Female	44.2 (155)	48.9 (46)	
Eczema at age, % (n)			
3-months	11.6 (35)	10.2 (9)	0.715
Missing, (n)	(50)	(6)	
6-months	15.0 (29)	20.5 (17)	0.257
Missing, (n)	(157)	(11)	
12-months	15.1 (28)	16.3 (13)	0.804
Missing, (n)	(165)	(14)	
FLG variants[†], % (n)			
Present (haploinsufficiency)	7.8 (9)	6.4 (6)	0.700
Missing, (n)	(235)	(0)	

FLG, filaggrin.

[*] Analytical sample refers to the sub-sample of participants with *FLG* gene expression and *FLG* genotype information.

[†] Individuals carrying the minor allele for at least one of the *FLG* variants R501X, 2282del4, or S3247X were classified as having filaggrin haploinsufficiency.

Table 2

Association between *FLG* variants and eczema at different ages and repeated measurements of eczema

Eczema at age	*FLG* variants		*P* value
	No (wild-type)	Yes (haploinsufficiency)	
3-months, % (n/total)	7.3 (6/82)	50.0 (3/6)	
RR (95% CI)[*]	1.00 (Ref.)	3.18 (1.05 – 9.59)	0.040
6-months, % (n/total)	16.7 (13/78)	80.0 (4/5)	
RR (95% CI)[*]	1.00 (Ref.)	3.38 (1.65 – 6.94)	< 0.001
12-months, % (n/total)	13.5 (10/74)	50.0 (3/6)	
RR (95% CI)[*]	1.00 (Ref.)	2.57 (1.03 – 7.16)	0.044
Repeated measurements[†], % (k/total)	12.4 (29/234)	58.8 (10/17)	
RR (95% CI)[‡]	1.00 (Ref.)	2.94 (1.33 – 6.51)	0.008

k, number of repeated measurements; *FLG*, filaggrin; RR, risk ratio; CI, confidence interval; Ref., reference category.

[*] Adjusted for sex

[†] Eczema was repeatedly measured at ages 3-, 6-, and 12-months. Generalized estimating equations (GEE) method was applied to account for the correlated observations.

[‡] Adjusted for sex and age at follow-up.

Table 3

Associations between *FLG* variants and *FLG* expression

	FLG variants		
FLG probe ID	**Yes (n = 6)** **(haploinsufficiency)**	**No (n = 88)** **(wild-type)**	**P value**
Sense transcripts			
A_24_P51322, Median (IQR)	−0.97 (0.81)	0.16 (1.14)	0.007
A_32_P387648, Median (IQR)	−0.39 (0.45)	0.09 (0.97)	0.052
A_33_P3261328, Median (IQR)	0.09 (0.96)	−0.01 (0.60)	0.799
Antisense transcripts			
A_21_P0014075, Median (IQR)	0.09 (0.41)	0.02 (0.78)	0.740
A_33_P3296200, Median (IQR)	−0.43 (0.47)	0.09 (0.93)	0.031

FLG, filaggrin; ID, identification; IQR, interquartile range.

FLG expression values presented as normalized log-base 2 intensities.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 4

Associations between *FLG* expression and repeated measurements of eczema [*]

FLG probe ID	Eczema, Median (IQR)		Sex-adjusted model[†]		FLG adjusted model[‡]	
	Yes (k = 39)	No (k = 212)	RR (95% CI)	P value	RR (95% CI)	P value
Sense transcripts						
A_24_P51322	−0.61 (1.19)	0.22 (1.15)	0.54 (0.33–0.89)	0.015	0.60 (0.38–0.95)	0.047
A_32_P387648	−0.22 (1.11)	0.05 (0.97)	0.83 (0.55–1.25)	0.373	0.95 (0.63–1.41)	0.782
A_33_P3261328	−0.14 (0.27)	0.01 (0.66)	0.60 (0.34–1.07)	0.077	0.67 (0.45–1.01)	0.058
Antisense transcripts						
A_21_P0014075	0.07 (0.74)	−0.01 (0.83)	2.02 (1.20–3.41)	0.008	2.02 (1.10–3.72)	0.024
A_33_P3296200	−0.26 (1.14)	0.11 (0.93)	0.79 (0.47–1.32)	0.366	0.94 (0.56–1.57)	0.811

k, number of repeated measurements; *FLG*, filaggrin; ID, identification; IQR, interquartile range; RR, risk ratio; CI, confidence interval.

[*] Eczema was repeatedly measured at ages 3–, 6–, and 12-months. Generalized estimating equations (GEE) method was applied to account for the correlated observations.

FLG expression values presented as normalized log-base 2 intensities.

[†] Adjusted for sex and age at follow-up.

[‡] Adjusted for sex, age at follow-up, and *FLG* variants status.

Table 5

Estimates of area under the receiver operating characteristic curve of models predicting eczema at 3-, 6-, and 12-months of age

	Variables	Eczema at 3 months (n = 88)	Eczema at 6 months (n = 83)	Eczema at 12 months (n = 80)
		AUC (95% CI)	AUC (95% CI)	AUC (95% CI)
Model 1	A_24_P51322 (sense transcript)	0.72 (0.53 – 0.91)	0.71 (0.56 – 0.85)	0.66 (0.48 – 0.83)
Model 2	A_24_P51322 + Sex	0.87 (0.78 – 0.96)	0.76 (0.64 – 0.88)	0.72 (0.55 – 0.89)
Model 3	A_24_P51322 + *FLG* variants	0.74 (0.54 – 0.93)	0.72 (0.57 – 0.86)	0.66 (0.49 – 0.84)
Model 4	A_24_P51322 + Sex + *FLG* variants	0.88 (0.79 – 0.97)	0.76 (0.64 – 0.89)	0.72 (0.55 – 0.89)
Model 5	A_21_P0014075 (antisense transcript)	0.63 (0.45 – 0.80)	0.61 (0.45 – 0.76)	0.60 (0.43 – 0.76)
Model 6	A_21_P0014075 + Sex	0.87 (0.78 – 0.95)	0.72 (0.59 – 0.85)	0.71 (0.56 – 0.86)
Model 7	A_21_P0014075 + *FLG* variants	0.78 (0.62 – 0.94)	0.73 (0.58 – 0.88)	0.71 (0.55 – 0.87)
Model 8	A_21_P0014075 + Sex + *FLG* variants	0.91 (0.84–0.98)	0.77 (0.63 – 0.90)	0.75 (0.61 – 0.90)

FLG, filaggrin; AUC, area under the receiver operating characteristic curve; CI, confidence interval