ORIGINAL ARTICLE

SKIN AND EYE DISEASES

Protein biomarkers in vernix with potential to predict the development of atopic eczema in early childhood

T. Holm[1], D. Rutishauser[2,3], Y. Kai-Larsen[4], Y. Lyutvinskiy[2], F. Stenius[5], R. A. Zubarev[2,3], B. Agerberth[4], J. Alm[5] & A. Scheynius[1]

[1]Translational Immunology Unit, Department of Medicine Solna, Karolinska Institutet and University Hospital, Stockholm; [2]Physiological Chemistry I, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm; [3]SciLifeLab, Stockholm; [4]Physiological Chemistry II, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm; [5]Department of Clinical Science and Education, Karolinska Institutet, Södersjukhuset, Sachs' Children and Youth Hospital, Stockholm, Sweden

Abstract

Background: Atopic eczema (AE) is a chronic inflammatory skin disease, which has increased in prevalence. Evidence points toward lifestyle as a major risk factor. AE is often the first symptom early in life later followed by food allergy, asthma, and allergic rhinitis. Thus, there is a great need to find early, preferentially noninvasive, biomarkers to identify individuals that are predisposed to AE with the goal to prevent disease development.

Objective: To investigate whether the protein abundances in vernix can predict later development of AE.

Methods: Vernix collected at birth from 34 newborns within the Assessment of Lifestyle and Allergic Disease During INfancy (ALADDIN) birth cohort was included in the study. At 2 years of age, 18 children had developed AE. Vernix proteins were identified and quantified with liquid chromatography coupled to tandem mass spectrometry.

Results: We identified and quantified 203 proteins in all vernix samples. An orthogonal projections to latent structures-discriminant analysis (OPLS-DA) model was found with $R^2 = 0.85$, $Q^2 = 0.39$, and discrimination power between the AE and healthy group of 73.5%. Polyubiquitin-C and calmodulin-like protein 5 showed strong negative correlation to the AE group, with a correlation coefficient of 0.73 and 0.68, respectively, and a *P*-value of 8.2 E-7 and 1.8 E-5, respectively. For these two proteins, the OPLS-DA model showed a prediction accuracy of 91.2%.

Conclusion: The protein abundances in vernix, and particularly that of polyubiquitin-C and calmodulin-like protein 5, are promising candidates as biomarkers for the identification of newborns predisposed to develop AE.

The prevalence of AE has rapidly increased and is estimated to 15–30% in children and 2–10% in adults with the highest prevalence found in Northern Europe (1). The pathogenesis of AE is likely to result from the combination of a disturbed skin barrier and inappropriate immune responses with contributions from both genetic and environmental factors. Atopic eczema is often the first symptom early in life of allergic disorders that many times progress to food allergy, asthma, and allergic rhinitis, a phenomenon known as the atopic march (2, 3). A recent report highlighted the need to identify different early biomarkers in AE with the goal to stop or even reverse the atopic march (4).

Inverse relation has been found between the number of anthroposophic lifestyle characteristics and the risk of atopy in school children of families with an anthroposophic lifestyle

(5, 6). Anthroposophic lifestyle includes biodynamic food containing live lactobacilli and restrictive use of antibiotics, antipyretics, and vaccinations (7). To investigate how lifestyle factors influence the development of allergic disease early in life, the birth cohort designated Assessment of Lifestyle and Allergic Disease During INfancy (ALADDIN) was initiated (7). In this birth cohort, the children born into families with an anthroposophic or nonanthroposophic life style have been followed prospectively from birth and many different samples have been collected, among those vernix caseosa (vernix). Vernix is a yellow-white, cheese-like, material consisting of sebum, lanugo hairs, and desquamated squamous cells, which covers the neonate's skin during the last trimester of gestation (8, 9). No other animal species produces vernix, making this material a unique human skin barrier film. It is believed to exhibit waterproofing, antioxidant, temperature-regulating, and anti-infective functions (8). The composition of vernix is mainly water (80.5%), lipids (10.3%), and proteins (9.1%) (9). Vernix is produced by the sebaceous glands, and in addition, lanugo hairs and desquamated squamous cells contribute to the formation of vernix (8). Thus, the origin of proteins found in vernix is to a large part from the fetal epidermis.

In this study, we investigated whether protein biomarkers in vernix, which reflects the composition of the infant's epidermal barrier, can predict later development of AE. As a source of biomarkers, vernix is attractive because it is usually abundant and its collection is noninvasive. We also addressed whether lifestyle can influence the protein composition of vernix by further dividing the AE and healthy control groups into three lifestyle categories: anthroposophic, partly anthroposophic, and nonanthroposophic.

Materials and methods

Study population

The study was designed to compare vernix collected at birth from children who had developed AE at 2 years of age with those who remained healthy. According to the inclusion criteria, based on doctor's diagnosis and the criteria of the UK Working Party's refinement of the Hanifin and Rajka criteria (10), 20 children had developed AE at 2 years of age. Due to limited amount of vernix from two of these children, 18 could be included. The healthy controls were selected among children with no current or history of AE or other allergic diseases and not allergen sensitized at 2 years of age. Furthermore, the healthy controls were selected from children where enough amount of vernix had been collected and to match the three lifestyle groups, gender, and parental characteristics of the AE children (see Table 1). Sixteen children who fulfilled all these requirements were included in the study as healthy controls. The anthroposophic, partly anthroposophic, and nonanthroposophic lifestyle groups were classified based on choice of Maternal-Child Health Centers and parental responses to a questionnaire, described in detail elsewhere (7). The study was approved by the Regional Ethical Review Board in Stockholm, and all parents have given their written informed consent.

Determination of sensitization

Parental allergen sensitization was defined by Phadiatop® (Phadia AB, Uppsala, Sweden) containing a mix of 11 inhalant allergens. Blood from the children at the age of 2 years were analyzed by ImmunoCAP® (Phadia AB) for IgE to seven common allergens (hen's egg, cow's milk, peanut, cat, dog, birch, and timothy). A study subject was classified as allergen sensitized if the IgE level was ≥0.35 kU_A/l for the parents in Phadiatop® and for the children in at least one of the seven allergens (7).

Collection of vernix

Vernix from newborns was collected by midwives and stored at −80°C. The estimated amount of vernix present on the whole body and the location from where the vernix was collected is presented in Table 1.

Extraction of proteins

Vernix (15 mg) from each individual was mixed with 50 mM ammonium bicarbonate (BDH Laboratory Supplies, Poole, UK)/1 M urea (Invitrogen, Carlsbad, CA, USA), incubated for 15 min on ice followed by sonication (10 cycles, each cycle 20 s with 40% amplitude) on ice using a probe (Vibra-Cell™ CV18, Sonics & Materials, Newtown, CT, USA). The lipids were pelleted by centrifugation for 30 min at 13 000 g, and the protein concentration in the supernatant was determined with the Bio-Rad detergent compatible protein assay (Bio-Rad, Hercules, CA, USA). A fraction of every sample was separated on 4–12% NuPAGE Bis-Tris gels (Invitrogen) prior to proteomic analysis, to verify the presence of proteins and confirm that the extraction of proteins was not destructive.

Liquid chromatography tandem mass spectrometry

The procedure for protein digestion is described in Supporting information, Data S1. Digested samples were acidified and cleaned with C18 StageTips according to the manufacturer's description (Thermo Fisher Scientific Inc., Waltham, MA, USA). Eluted peptides were dried and re-suspended in 3% acetonitrile (ACN) and 0.2% formic acid (Sigma-Fluka, Munich, Germany). Detailed description of the procedure of liquid chromatography tandem mass spectrometry (LC-MS/MS) analyses can be found in Data S1.

Data analysis

Tandem mass spectra were extracted using Raw2MGF v.2.1.0 [in-house developed software (11)], and the resulting Mascot generic files were searched against the human Swiss-Prot protein database (v. 2012.06, reversed protein sequences were added to the database for decoy search) using the Mascot 2.3.02 (Matrix Science Ltd., London, UK), with trypsin for generation of peptides and allowing

Table 1 Demographic data including allergic symptoms, sensitization, and environmental factors in families with an anthroposophic, partly anthroposophic, and nonanthroposophic lifestyle. [Categorical variables: n/N (%). Continuous variables: mean ± SD]

	Atopic eczema	Healthy	P*	Atopic eczema				Healthy			
				Anthro.	Partly anthro.	Nonanthro.	P**	Anthro.	Partly anthro.	Nonanthro.	P**
n	18	16		5	7	6		6	6	4	
Child											
Sex (female)	7/18 (39)	7/16 (44)	1.0	2/5 (40)	2/7 (29)	3/6 (50)	0.83	1/6 (17)	3/6 (50)	3/4 (75)	0.22
Birth weight (g)	3556 ± 500	3607 ± 481	0.77	3797 ± 578	3390 ± 584	3548 ± 278	0.75	3474 ± 510	3870 ± 465	3410 ± 374	0.25
Birth height (cm)	51.0 ± 1.4	50.9 ± 2.6	0.82	51.2 ± 1.3	50.9 ± 1.2	50.3 ± 1.2	0.59	49.7 ± 3.2	52.2 ± 2.2	51.0 ± 1.6	0.27
Gestational week	39.4 ± 1.2	39.5 ± 0.8	0.91	39.2 ± 1.1	39.4 ± 1.0	39.5 ± 1.2	0.92	39.3 ± 0.8	39.7 ± 1.0	39.5 ± 0.6	0.76
Having older siblings	10/15 (67)	9/16 (56)	0.72	3/5 (60)	5/6 (83)	2/4 (50)	0.52	4/6 (67)	1/6 (17)	4/4 (100)	0.05
24 months SCORAD median (range) (10)	19.3 (4.5–41)	–	–	24.5 (10.5–37)	16.2 (4.5–41)	14 (4.5–27.5)	0.30	–	–	–	–
Parents											
Reactions to food†	8/18 (44)	–	–	2/5 (40)	2/7 (29)	4/6 (67)	0.42	–	–	–	–
Recurrent wheeze‡	2/18 (11)	–	–	1/5 (20)	1/7 (14)	0/6 (0)	0.73	–	–	–	–
Sensitized§	7/18 (39)	–	–	2/5 (40)	2/6 (33)	3/6 (50)	1.0	–	–	–	–
Parents											
Age mother at birth of child, years	30.0 ± 5.4	31.5 ± 4.6	0.41	27 ± 5.4	31 ± 5.8	31 ± 5.1	0.41	30.8 ± 5.1	30.8 ± 5.6	33.5 ± 1.3	0.63
Age father at birth of child, years	33.8 ± 6.5	34.7 ± 5.4	0.67	32 ± 7	36 ± 7.2	33 ± 5.6	0.57	36.7 ± 6.9	31.0 ± 3.1	36.5 ± 2.6	0.16
Mother reported eczema	3/18 (17)	4/16 (25)	0.68	1/5 (20)	1/7 (14)	1/6 (17)	1.0	3/6 (50)	0/6 (0)	1/4 (25)	0.13
Father reported eczema	4/18 (22)	2/15 (13)	0.67	1/5 (20)	1/7 (14)	2/6 (33)	0.79	1/6 (17)	0/5 (0)	1/4 (25)	0.71
Mother sensitized¶	4/18 (22)	6/16 (38)	0.46	2/5 (40)	0/7 (0)	2/6 (33)	0.19	3/6 (50)	2/6 (33)	1/4 (25)	0.83
Father sensitized¶	5/17 (29)	5/14 (36)	1.0	2/5 (40)	2/6 (33)	1/6 (17)	0.82	2/6 (33)	2/5 (40)	1/3 (33)	1.0
Pregnancy											
Mother											
Antibiotics	2/18 (11)	2/16 (13)	1.0	1/5 (20)	0/7 (0)	1/6 (17)	0.50	0/6 (0)	1/6 (17)	1/4 (25)	0.70
Vegetarian diet	3/18 (17)	4/16 (25)	0.68	1/5 (20)	1/7 (14)	1/6 (17)	1.0	3/6 (50)	1/6 (17)	0/4 (0)	0.15
Organic/biodynamic diet	5/18 (28)	1/15 (7)	0.18	3/5 (60)	2/7 (29)	0/6 (0)	0.14	1/6 (17)	0/6 (0)	0/4 (0)	1.0
Smoking	1/18 (6)	0/16 (0)	1.0	0/5 (0)	0/7 (0)	1/6 (17)	0.61	0/6 (0)	0/6 (0)	0/4 (0)	–
Snuffing	0/18 (0)	0/16 (0)	–	0/5 (0)	0/7 (0)	0/6 (0)	–	0/6 (0)	0/6 (0)	0/4 (0)	–
Father smoking	4/18 (22)	1/15 (7)	0.35	1/5 (20)	2/7 (29)	1/6 (17)	1.0	1/6 (17)	0/5 (0)	0/4 (0)	1.0
Delivery											
Home delivery	4/18 (22)	2/16 (13)	0.66	2/5 (40)	2/7 (29)	0/6 (0)	0.38	2/6 (33)	0/6 (0)	0/4 (0)	0.30
Vacuum extraction	0/18 (0)	2/16 (13)	0.21	0/5 (0)	0/7 (0)	0/6 (0)	–	1/6 (17)	1/6 (17)	0/4 (0)	1.0
Acute cesarean	2/18 (11)	1/16 (6)	1.0	0/5 (0)	0/7 (0)	2/6 (33)	0.16	1/6 (17)	0/6 (0)	0/4 (0)	1.0
Elective cesarean	1/18 (6)	0/16 (0)	1.0	0/5 (0)	1/7 (14)	0/6 (0)	1.0	0/6 (0)	0/6 (0)	0/4 (0)	–
Anesthesia	7/18 (39)	11/16 (69)	0.1	2/5 (40)	0/7 (0)	5/6 (83)	**0.01**	3/6 (50)	4/6 (67)	4/4 (100)	0.33
Antibiotics	1/18 (6)	1/16 (6)	1.0	0/5 (0)	0/7 (0)	1/6 (17)	0.61	1/6 (17)	0/6 (0)	0/4 (0)	1.0
Vernix											
Location (n)††	2a, 8b, 3c, 3e, 2f (18)	6b, 2c, 1d, 4e, 3f (16)	0.62	4b, 1e (5)	1a, 2b, 2c, 2e (7)	1a, 2b, 1c, 1e, 1f (5)	0.57	2b, 1c, 1e, 2f (6)	1b, 1c, 1d, 2e, 1f (6)	3b, 1e (4)	0.86
Estimated amount on the body [Median (range)] n‡‡	3 (1–4) (18)	3 (1–4) (16)	0.92	3 (1–4) (5)	3 (1–4) (6)	3 (2–4) (5)	1.0	3 (1–4) (4)	3 (3–4) (5)	3.5 (3–4) (4)	0.22
Protein concentration after extraction (mg/ml) (n)§§	1.26 ± 0.47 (18)	1.27 ± 0.45 (16)	0.94	0.93 ± 0.17 (5)	1.33 ± 0.41(7)	1.44 ± 0.60 (6)	0.17	1.36 ± 0.61(6)	1.11 ± 0.20 (6)	1.37 ± 0.49 (4)	0.58

Categorical variables: P for trend (Fisher's exact test); continuous variables: P for comparisons of means (one-way ANOVA).
Bold indicates P < 0.05.
*P for comparison of the atopic eczema (AE) group vs the healthy group.
**P for comparison of the lifestyle groups anthro. vs partly anthro. vs nonanthro within the AE group and the healthy group, respectively.
†Reactions to food were defined as acute onset of symptoms such as skin reactions, wheezing, vomiting, or diarrhea on more than one occasion after ingestion or contact with a particular type of food (24).
‡Recurrent wheeze was defined as three or more episodes of wheeze since the last examination (24)
§Classified as sensitized if IgE level was ≥0.35 kU_A/l for at least one of the seven allergens analyzed using ImmunoCap™ (Phadia AB).
¶Classified as sensitized if IgE level was ≥0.35 kU_A/l measured using Phadiatop® (Phadia AB).
††a: arm pit, b: back, c: groin, d: leg, e: other or mixed, f: not indicated; reported by the midwives.
‡‡1: very small, 2: small, 3: medium, 4: large; reported by the midwives.
§§Measured by the DC protein assay (Bio-Rad).

for one missed cleavage site. Mass tolerance was set to 10 ppm for peptides and 0.02 Da for the fragment ions. Carbamidomethylation of cysteine was specified as a fixed modification, whereas oxidation of methionine and deamidation of asparagine and glutamine were defined as variable modifications. Proteins were grouped into protein classes with the tools in Protein ANalysis THrough Evolutionary Relationships (PANTHER) available at http://www.pantherdb.org.

Quantitation

Quantitative information was extracted using in-house developed software Quanti v 2.5.3.1 (11). For quantitative purposes only reliably identified (false discovery rate <0.01), unmodified peptides with unique sequences were considered and only proteins discovered in all samples with at least two such peptides were quantified. For each protein, one database identifier (ID) was selected, covering all the identified peptide sequences for this specific protein. If two protein IDs had partial intersection, then all the peptides belonging to this intersection were excluded from the analysis. The results from Quanti v 2.5.3.1 are reported as a set of relative protein abundances scaled to 1.0 as the geometric mean of the abundance of each protein over all samples. Relative abundance values were log-scaled to reveal down-regulated proteins with the same significance as up-regulated.

Statistical analysis

Statistical analyses were conducted using IBM SPSS Statistics 20 software (Chicago, IL, USA). A P-value < 0.05 was regarded as statistically significant. Fischer's exact test (for categorical variables) and ANOVA (for comparisons of means of continuous variables) were used for the comparison of demographic and exposure variables between the groups in Table 1. ANOVA was also used in Fig. 3C and D to compare the relative protein abundances of polyubiquitin-C and calmodulin-like protein 5 in the different groups. In Table 2 and Supplementary Table S1, P-values were calculated using Student's t-test and correlation coefficients were calculated using Pearson's Product-Moment Correlation. The orthogonal projections to latent structures-discriminant analysis (OPLS-DA) method was employed for discrimination between the AE and healthy control groups. The OPLS-DA is a supervised classification technique which convolutes systematic variation and maximizes the difference between observations according to provided classification. This will display which variables are responsible for class discrimination (12). Classification power was determined with receiver operating characteristic (ROC) analysis. Receiver operating characteristic curves show the predictive accuracy of the cross-validated OPLS-DA model depicting the equal probability of the false-positive and false-negative results (13). OPLS-DA modeling was performed using SIMCA 13 (Umetrics, Umeå, Sweden).

Results

Study population

Demographic data and the amount, location of collection, and protein concentration of vernix for the 34 children are presented in Table 1. There were no significant differences between any groups, except for the lifestyle groups within the AE group regarding anesthesia at delivery ($P = 0.01$) (Table 1).

Proteins identified

After LC-MS/MS analyses, 208 proteins were identified, with at least two peptides matching and a false discovery rate of less than 1%, and quantified in all 34 vernix samples. Three protein entries for the hemoglobin subunits (HBA_HUMAN, HBB_HUMAN, and HBD_HUMAN) were considered blood contaminants and were excluded from further analysis. Furthermore, two proteins were removed due to their high correlation with similar proteins: CALM_HUMAN highly correlates with CALL5_HUMAN ($R = 0.862$) and probably represents a different isoform of the same protein, and K2C4_HUMAN is also known to be co-expressed with K1C13_HUMAN ($R = 0.96$). Thus, a final list of 203 proteins was used for statistical analysis, classification modeling using SIMCA (Supplementary Table S1) and categorization of proteins into functional classes utilizing PANTHER (Fig. 1).

Proteins with most significant differences between the AE and healthy group

The 30 proteins with highest impact on the classification of individuals into the AE or healthy group are presented in Table 2. Among the proteins present at higher levels in the AE group, peroxiredoxin-2 and serpin A12 had the highest impact (OPLS-DA model loadings −0.127 and −0.123, respectively), and among the proteins present at lower levels in the AE group, polyubiquitin-C and calmodulin-like protein 5 contributed most (OPLS-DA model loadings 0.166 and 0.157, respectively). Possible confounders were investigated using logistic regressions models. All the variables in Table 1 were included one at a time (due to the limited amount of data) together with peroxiredoxin-2, serpin A12, polyubiquitin-C, and calmodulin-like protein 5 in individual models with AE as outcome. The association between the protein composition of vernix and AE was still statistically significant, regardless of which variable that was included; thus, no confounders were present. To investigate the presence of up- or down-regulated proteins in vernix from newborns that had developed AE at 2 years of age compared with vernix from children that remained healthy, the relative protein abundances were calculated. An OPLS-DA model was found for discrimination between AE and healthy control samples with $R^2 = 0.85$ and $Q^2 = 0.39$ (Fig. 2A). Although the OPLS-DA model itself showed clear separation, we present results of sevenfold cross-validation of this model to avoid overestimation of predictive power in the modeling (14). Receiver operating characteristic curve analysis of the OPLS-DA model cross-validation results showed 73.5% predictive

Table 2 The proteins that differ most between the atopic eczema (AE) and the healthy group

Rank*	Protein ID (SwissProt)	Protein	P-value†	OPLS-DA model loadings‡	Correlation coefficient to health status§	Protein ID (SwissProt)	Protein	P-value†	OPLS-DA model loadings‡	Correlation coefficient to health status§
	Higher relative abundance in AE					Lower relative abundance in AE				
1	PRDX2_HUMAN	Peroxiredoxin-2	6.35E-04	−0.127	−0.563	UBC_HUMAN	Polyubiquitin-C	8.21E-07	0.166	0.729
2	SPA12_HUMAN	Serpin A12	6.22E-03	−0.123	−0.469	CALL5_HUMAN	Calmodulin-like protein 5	1.80E-05	0.157	0.684
3	ASAH1_HUMAN	Acid ceramidase	6.56E-03	−0.116	−0.489	TPM4_HUMAN	Tropomyosin alpha-4 chain	5.16E-04	0.130	0.559
4	BPIA1_HUMAN	BPI fold-containing family A member 1	3.45E-03	−0.111	−0.512	CD59_HUMAN	CD59 glycoprotein	1.80E-03	0.121	0.514
5	DESP_HUMAN	Desmoplakin	7.64E-03	−0.110	−0.460	INVO_HUMAN	Involucrin	2.28E-03	0.119	0.522
6	BPIB1_HUMAN	BPI fold-containing family B member 1	3.95E-03	−0.109	−0.507	S10A7_HUMAN	Protein S100-A7	3.23E-03	0.118	0.524
7	K1C13_HUMAN	Keratin, type I cytoskeletal 13	1.83E-03	−0.109	−0.525	AATC_HUMAN	Aspartate aminotransferase. cytoplasmic	3.07E-03	0.113	0.492
8	GDPD3_HUMAN	Glycerophosphodiester phosphodiesterase domain-containing protein 3	1.83E-02	−0.106	−0.416	MYH9_HUMAN	Myosin-9	1.28E-02	0.107	0.428
9	HUTH_HUMAN	Histidine ammonia-lyase	2.45E-02	−0.105	−0.397	FLNB_HUMAN	Filamin-B	4.38E-03	0.107	0.485
10	SAP3_HUMAN	Ganglioside GM2 activator	2.53E-02	−0.104	−0.412	FILA_HUMAN	Filaggrin	2.94E-03	0.106	0.489
11	HEM2_HUMAN	Delta-aminolevulinic acid dehydratase	5.93E-03	−0.103	−0.479	PSA3_HUMAN	Proteasome subunit alpha type-3	4.69E-03	0.106	0.477
12	APOA4_HUMAN	Apolipoprotein A-IV	5.68E-03	−0.103	−0.464	CATH_HUMAN	Pro-cathepsin H	2.54E-02	0.103	0.390
13	CATD_HUMAN	Cathepsin D	1.91E-02	−0.100	−0.405	PSB5_HUMAN	Proteasome subunit beta type-5	2.76E-03	0.101	0.504
14	PLEC_HUMAN	Plectin	1.74E-02	−0.100	−0.425	S10A8_HUMAN	Protein S100-A8	2.13E-02	0.098	0.399
15	SAPL1_HUMAN	Proactivator polypeptide-like 1	1.41E-02	−0.100	−0.441	LAMP1_HUMAN	Lysosome-associated membrane glycoprotein 1	6.15E-03	0.097	0.477

*The proteins are ranked according to the orthogonal projections to latent structures-discriminant analysis (OPLS-DA) model loadings (12).
†Two-tailed Student's t-test.
‡The model loadings are a measure of how much each protein contributes to the classification into AE or healthy group (12).
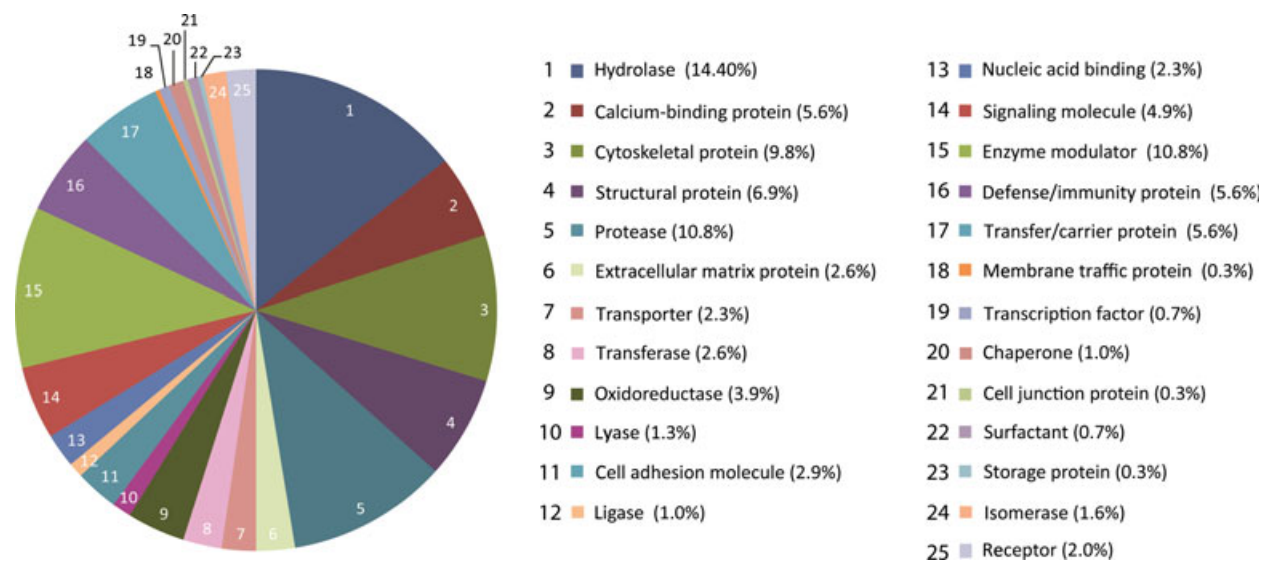§Correlation coefficient calculated using Pearson's product-moment correlation.

Figure 1 Classification of the identified and quantified 203 vernix proteins according to their function. Proteins were grouped into protein classes with the tools available in Protein ANalysis THrough Evolutionary Relationships (PANTHER).

accuracy (Fig. 2B). No significant difference was found by OPLS-DA in the protein composition of vernix between the lifestyle groups ($R^2 = 0.15$ and $Q^2 = -0.05$).

The potential AE biomarkers polyubiquitin-C and calmodulin-like protein 5

When the intergroup t-test value distributions were plotted, two proteins had much stronger *P*-values than all the rest, polyubiquitin-C and calmodulin-like protein 5 (Fig. 3A). These two proteins also showed exceptionally high correlation with the healthy state of the children, with a correlation value of 0.73 and 0.68, respectively, and a *P*-value of 8.2 E-7 and 1.8 E-5, respectively (Table 2). The OPLS-DA model providing the optimal linear combination of protein abundances for these two proteins had prediction accuracy of 91.2% (Fig. 3B). The relative abundances of polyubiquitin-C and calmodulin-like protein 5 in all 34 vernix samples were plotted and showed a significant different distribution between the AE and healthy groups (*P* = <0.0001 for both proteins) (Fig. 3C and D). These findings allow considering polyubiquitin-C and calmodulin-like protein 5 as potential biomarkers for AE prediction. No significant differences in the relative abundances of polyubiquitin-C and calmodulin-like protein 5 between the three lifestyle groups or between sensitized and nonsensitized children (Table 1) could be found (Fig. 3C and D).

Discussion

We identified and quantified 203 proteins in vernix collected from 34 newborns. For two proteins, polyubiquitin-C and calmodulin-like protein 5, the abundance variations between the AE and healthy group demonstrated strong significance (Table 2). Together with the ROC curve prediction accuracy

of 91.2%, these two proteins emerge as promising biomarkers in vernix to predict development of AE.

Most vernix proteins identified in this study are hydrolases (14.4%), proteases (10.8%), and enzyme modulators (10.8%) (Fig. 1). These functions are all inter-related (proteases are a kind of hydrolases and both groups are enzymes) and suggest that these processes have important roles in the vernix of newborns. Vernix is produced by the sebaceous glands, and in addition, lanugo hairs and desquamated squamous cells contribute to the formation of vernix (8). Thus, the origin of proteins found in vernix is to a large part from the fetal epidermis. Previously, the vernix proteome has been analyzed on pooled samples from newborns by a 2D-gel electrophoresis-based approach resulting in 41 identified proteins (15). In that study, it was found that almost half of these proteins (39%) are related to immunity, whereas we here demonstrate that only 5.6% of the identified proteins were associated with immunity (Fig. 1). This discrepancy clearly demonstrates that the method used for the proteome analysis affects not only the sensitivity but also the specificity. In the 2D-gel approach, the identification of proteins is restricted to the spots that are selected from the gel, while the method used in our study enables identification of all proteins in a solution. When analyzing the 203 proteins in our study for differences in abundance in vernix from newborns, who had developed AE at 2 years of age and those who remained healthy, an OPLS-DA model was found with $R^2 = 0.85$ and $Q^2 = 0.39$ (Fig. 2A). No significant differences in vernix protein composition could, however, be detected by OPLS-DA between the lifestyle groups.

What is the biological significance of the two potential biomarkers present at lower levels in the AE group found in this study? Polyubiquitin-C is a precursor protein that is cleaved into the active ubiquitin monomer (16). The ubiquitin conjugation system regulates a wide variety of biological processes, including protein degradation and signal transduction
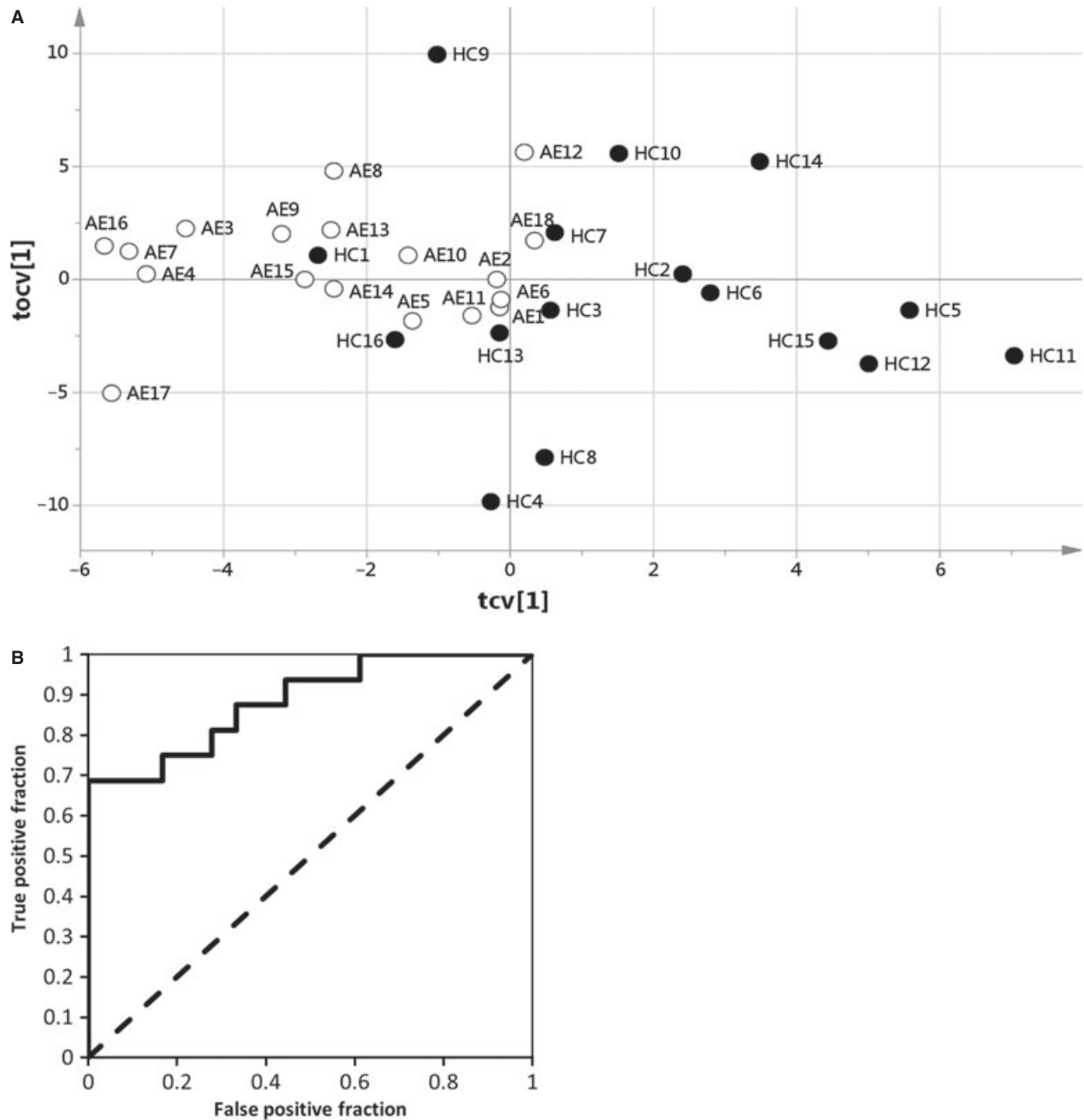
Figure 2 Analysis of the proteome dataset of 203 proteins. (A) Cross-validated score plot from the OPLS-DA model, illustrating the separation of atopic eczema (AE) (○) and healthy control (●) samples. tcv = sample projection to maximum discrimination dimension; tocv = maximum deviation orthogonal to discrimination dimension in sample set. (B) Receiver operating characteristic (ROC) curve for the predictive score of the OPLS-DA model calculated during sevenfold cross-validation (AUC = 0.885; accuracy = 73.5%).

(17). Furthermore, protein ubiquitylation is of fundamental importance in the regulation of both the innate and the adaptive immune system (18). Decreased ubiquitin level lowers the activation threshold of cells to a variety of chemical and environmental stresses (16), suggesting that the epidermis of the children in the AE group might be more sensitive to external stressors. Calmodulin-like protein 5 (also called calmodulin-like skin protein) is a calcium-binding protein and a marker for late keratinocyte differentiation (19). Its ability to associate with transglutaminase-3 (19), a key enzyme involved in the formation of the cornified envelope (20), indicates an important role of this protein in the epidermal barrier function. To our knowledge, there are no reports on mutations in the genes coding for polyubiqu-
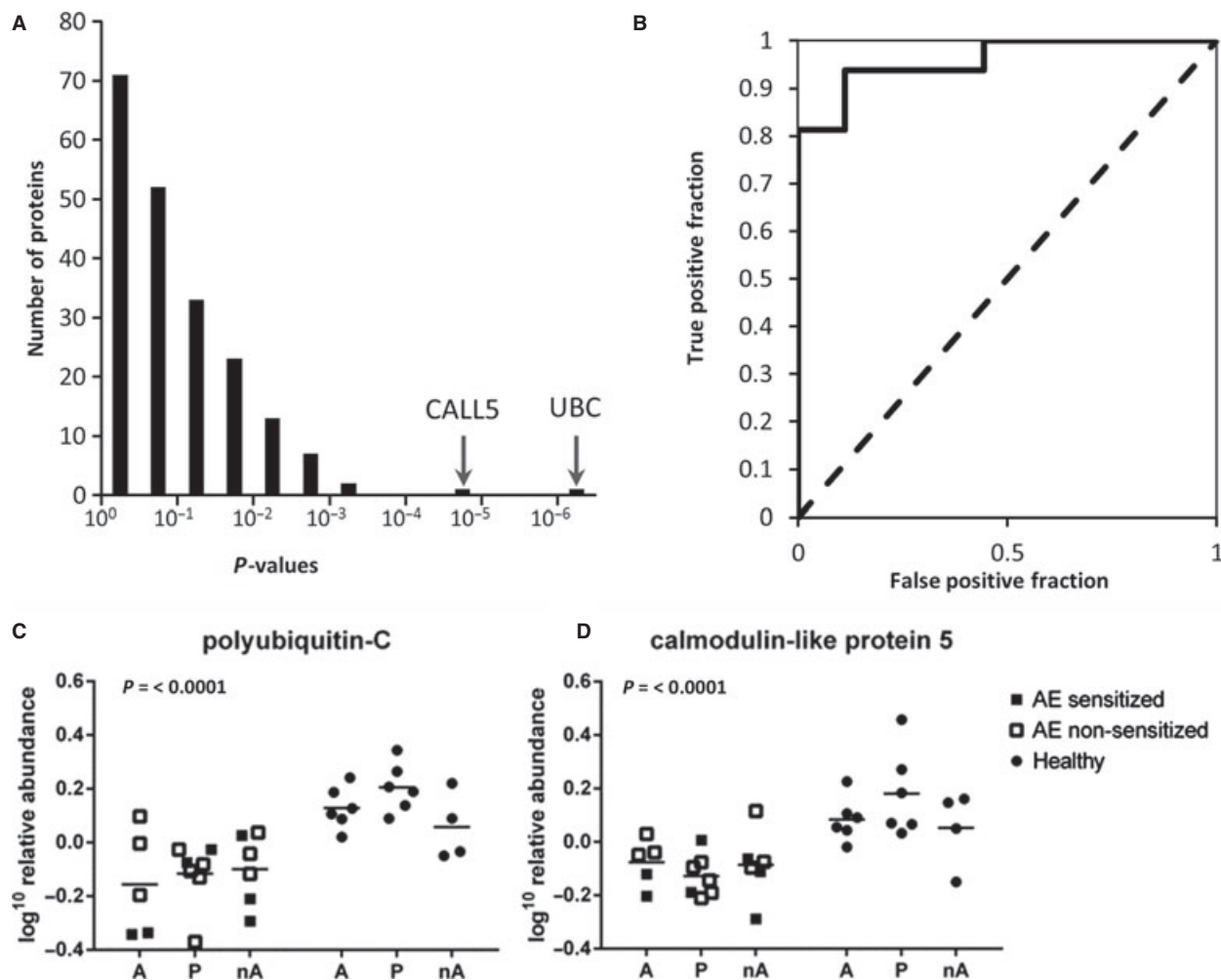
Figure 3 Analysis of polyubiquitin-C and calmodulin-like protein 5. (A) Intergroup t-test value distribution. Polyubiquitin-C (UBC), calmodulin-like protein 5 (CALL5). (B) Receiver operating characteristic (ROC) curve for the predictive score of the OPLS-DA model, based on polyubiquitin-C and calmodulin-like protein 5, calculated during sevenfold cross-validation (AUC = 0.958; accuracy = 91.2%). (C) Polyubiquitin-C and (D) calmodulin-like protein 5 relative protein abundances within the healthy group (●) and within the atopic eczema (AE) group, divided into sensitized (■) and nonsensitized (□), subdivided into lifestyle groups: anthroposophic (A), partly anthroposophic (P), and nonanthroposophic (nA). Statistical significance was assessed using two-way ANOVA.

itin-C and calmodulin-like protein 5 associated with the pathogenesis of AE, which could be a plausible explanation for their lower levels, as is the case for mutations in the human filaggrin gene (21). Another explanation to the reduced levels of polyubiquitin-C and calmodulin-like protein 5 could be that there are many proteases present in vernix (see Fig. 1).

Polyubiquitin-C and calmodulin-like protein 5 showed strong statistical significance in their negative correlation with the AE group. However, when interpreting the results, it has to be reminded that they are based on two proteins and 34 children (larger populations tend to give less accurate but more realistic models). As nontargeted proteomics analysis is increasingly being used not only in the discovery phase but also for analytical purposes (11), the complete or partial vernix proteome (protein pattern) could also be a potential biomarker. Its 73.5% prediction accuracy was lower than

that of the two proteins, polyubiquitin-C and calmodulin-like protein 5, but full-proteome measurement is more robust as it accumulates signals from hundreds of peptides.

There was also a group of proteins with significantly higher relative abundance in the AE group, although their correlation with AE was not as strong as for polyubiquitin-C or calmodulin-like protein 5 (Table 2). Among those were peroxiredoxin-2 and serpin A12. The level of peroxiredoxin-2 is increased in lymphocytes of patients with rheumatoid arthritis, with a positive correlation with the concentration of the acute-phase protein C-reactive protein (22), indicating its involvement in inflammatory responses. Serpin A12 (also called alpha-1 antitrypsin) is a member of the serpin (serine protease inhibitors) family and is an acute-phase protein, which plasma concentration increases upon inflammatory signals (23). Hence, elevated levels of peroxiredoxin-2 and serpin A12 in

vernix of the AE group indicate that the skin of these neonates might be prone to develop inflammation, which may possibly have a physiological association with AE.

In summary, we have shown that the protein composition of vernix has the potential to function as a predictor of future AE development. Identifying risk groups is of uttermost importance because restoration of the skin barrier function early in life may help prevent the outbreak of the disease.

Acknowledgment

Author contributions

TH, FS, JA, and AS involved in conception and design of the study. TH, DR, YKL, YL, RZ, BA, and AS performed acquisition, analysis, and interpretation of data. TH, DR, YKL, and AS contributed to drafting the article. FS, RZ, BA, and JA revised the article. All authors approved the final article.

Funding

Conflicts of interest

The authors claim no conflicts of interest.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Data S1. Methods.

Table S1. List of the 203 proteins identified and quantified in vernix.

References

1. Bieber T. Atopic dermatitis. *Ann Dermatol* 2010;**22**:125–137.

2. Ballardini N, Kull I, Lind T, Hallner E, Almqvist C, Ostblom E et al. Development and comorbidity of eczema, asthma and rhinitis to age 12: data from the BAMSE birth cohort. *Allergy* 2012;**67**:537–544.

3. Gustafsson D, Sjoberg O, Foucard T. Development of allergies and asthma in infants and young children with atopic dermatitis–a prospective follow-up to 7 years of age. *Allergy* 2000;**55**:240–245.

4. Bieber T, Cork M, Reitamo S. Atopic dermatitis: a candidate for disease-modifying strategy. *Allergy* 2012;**67**:969–975.

5. Alm JS, Swartz J, Lilja G, Scheynius A, Pershagen G. Atopy in children of families with an anthroposophic lifestyle. *Lancet* 1999;**353**:1485–1488.

6. Floistrup H, Swartz J, Bergstrom A, Alm JS, Scheynius A, van Hage M et al. Allergic disease and sensitization in Steiner school children. *J Allergy Clin Immunol* 2006;**117**:59–66.

7. Stenius F, Swartz J, Lilja G, Borres M, Bottai M, Pershagen G et al. Lifestyle factors and sensitization in children - the ALADDIN birth cohort. *Allergy* 2011;**66**:1330–1338.

8. Hoath SB, Pickens WL, Visscher MO. The biology of vernix caseosa. *Int J Cosmet Sci* 2006;**28**:319–333.

9. Hoeger PH, Schreiner V, Klaassen IA, Enzmann CC, Friedrichs K, Bleck O. Epidermal barrier lipids in human vernix caseosa: corresponding ceramide pattern in vernix and fetal skin. *Br J Dermatol* 2002;**146**:194–201.

10. Williams HC, Burney PG, Hay RJ, Archer CB, Shipley MJ, Hunter JJ et al. The U.K. working party's diagnostic criteria for atopic dermatitis. I. Derivation of a minimum set of discriminators for atopic dermatitis. *Br J Dermatol* 1994;**131**:383–396.

11. Lyutvinskiy Y, Yang H, Rutishauser D, Zubarev R. In silico instrumental response correction improves precision of label-free proteomics and accuracy of proteomics-based predictive models. *Mol Cell Proteomics* 2013;**12**:2324–2331.

12. Trygg J, Wold S. Orthogonal projections to latent structures (O-PLS). *J Chemometric* 2002;**16**:119–128.

13. Fawcett T. An introduction to ROC analysis. *Pattern Recogn Lett* 2006;**27**:861–874.

14. Broadhurst DI, Kella DB. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics* 2006;**2**:171–196.

15. Tollin M, Jagerbrink T, Haraldsson A, Agerberth B, Jornvall H. Proteome analysis of vernix caseosa. *Pediatr Res* 2006;**60**:430–434.

16. Hanna J, Meides A, Zhang DP, Finley D. A ubiquitin stress response induces altered proteasome composition. *Cell* 2007;**129**:747–759.

17. Iwai K. Linear polyubiquitin chains: a new modifier involved in NFkappaB activation and chronic inflammation, including dermatitis. *Cell Cycle* 2011;**10**:3095–3104.

18. Liu YC, Penninger J, Karin M. Immunity by ubiquitylation: a reversible process of modification. *Nat Rev Immunol* 2005;**5**:941–952.

19. Mehul B, Bernard D, Simonetti L, Bernard MA, Schmidt R. Identification and cloning of a new calmodulin-like protein from human epidermis. *J Biol Chem* 2000;**275**:12841–12847.

20. Candi E, Schmidt R, Melino G. The cornified envelope: a model of cell death in the skin. *Nat Rev Mol Cell Biol* 2005;**6**:328–340.

21. Irvine AD, McLean WH, Leung DY. Filaggrin mutations associated with skin and allergic diseases. *N Engl J Med* 2011;**365**:1315–1327.

22. Szabo-Taylor KE, Eggleton P, Turner CA, Faro ML, Tarr JM, Toth S et al. Lymphocytes from rheumatoid arthritis patients have elevated levels of intracellular peroxiredoxin 2, and a greater frequency of cells with exofacial peroxiredoxin 2, compared with healthy human lymphocytes. *Int J Biochem Cell Biol* 2012;**44**:1223–1231.

23. Bergin DA, Hurley K, McElvaney NG, Reeves EP. Alpha-1 antitrypsin: a potent anti-inflammatory and potential novel therapeutic agent. *Arch Immunol Ther Exp (Warsz)* 2012;**60**:81–97.

24. Stenius F, Borres M, Bottai M, Lilja G, Lindblad F, Pershagen G et al. Salivary cortisol levels and allergy in children: the ALADDIN birth cohort. *J Allergy Clin Immunol* 2011;**128**:1335–1339.