Published in final edited form as:

J Dermatol Sci. 2015 June ; 78(3): 173–180. doi:10.1016/j.jdermsci.2015.02.013.

Quantitative proteogenomic profiling of epidermal barrier formation *in vitro*

Jason M. Winget, Ph.D.^a, Julian D. Watts, Ph.D.^a, Michael R. Hoopmann, Ph.D.^a, Teresa DiColandrea, Ph.D.^b, Michael K. Robinson, Ph.D.^b, Tom Huggins, Ph.D.^b, Charles C. Bascom, Ph.D.^b, Robert J. Isfort, Ph.D.^b, and Robert L. Moritz, Ph.D.^{a,*}

^aInstitute for Systems Biology, 401 Terry Ave N., Seattle, WA 98109 USA

^bThe Procter & Gamble Company, Mason Business Center, Cincinnati, OH 45040 USA

Abstract

Background—The barrier function of the epidermis is integral to personal well-being, and defects in the skin barrier are associated with several widespread diseases. Currently there is a limited understanding of system-level proteomic changes during epidermal stratification and barrier establishment.

Objective—Here we report the quantitative proteogenomic profile of an *in vitro* reconstituted epidermis at three time points of development in order to characterize protein changes during stratification.

Methods—The proteome was measured using data-dependent “shotgun” mass spectrometry and quantified with statistically validated label-free proteomic methods for 20 replicates at each of three time points during the course of epidermal development.

Results—Over 3600 proteins were identified in the reconstituted epidermis, with more than 1200 of these changing in abundance over the time course. We also collected and discuss matched transcriptomic data for the three time points, allowing alignment of this new dataset with previously published characterization of the reconstituted epidermis system.

Conclusion—These results represent the most comprehensive epidermal-specific proteome to date, and therefore reveal several aspects of barrier formation and skin composition. The limited correlation between transcript and protein abundance underscores the importance of proteomic analysis in developing a full understanding of epidermal maturation.

Keywords

Skin equivalent; epidermal differentiation; barrier; proteomics

Introduction

The skin carries out a variety of protective functions that must be maintained despite the constant turnover of skin tissue and are collectively termed the epidermal barrier. These functions include water retention, antibacterial action, protection from toxic substances, and initial immune responses [1]. Barrier dysfunction is tied to many acute and chronic conditions, several of which are prevalent in occurrence such as ichthyosis vulgaris, atopic dermatitis, and psoriasis [2].

The barrier is initially formed *in utero* at approximately 34 weeks gestation [2] and is comprised of several functional components. Keratinocyte-derived squames of the outer epidermis are sheathed in a layer of lipids and proteins called the cornified envelope [3]. Disruption of the lipid “mortar” (e.g. with detergents) causes barrier disruption and skin irritation. In lower epidermal layers, protein-based cell-cell junctions are another important component of the barrier. Loss of tight junctions in the central region of the epidermis (the stratum granulosum) leads to death in neonatal mice [4, 5]. Additional proteins including Loricrin, Involucrin, Keratins, and Desmosome components also contribute to the barrier function [6]. A comprehensive, quantitative proteomic profile of the temporal differences in protein abundance would aid in understanding barrier health and functionality.

Comprehensive proteomic studies of the skin have been hampered by a number of factors. The dynamic range of proteins in skin, where Keratins can comprise 70% of the cells by dry weight [7], complicates detection of lower abundance proteins in the sample. To overcome this issue, past efforts have frequently employed separation of proteins via gel electrophoresis [8–10], a procedure with low throughput as well as poor sensitivity for lower abundance proteins [11]. Several published studies have been carried out using relatively undifferentiated cultured cells which do not suffer from such extreme dynamic range [12–15]. Such studies can produce a large number of protein identifications but do not accurately represent the stratified structure and resulting protein profile of natural epidermis. Here we report a quantitative proteomic time course analysis of a previously described reconstituted epidermis (RE) [16]. Extensive characterization of this model demonstrated many strong biological parallels with natural skin, including stratification, similar lipid and natural moisturizing factor composition, a functional water barrier & gradient, appropriate pH, and proper localization of epidermal marker proteins. Transcript analysis of this model revealed several time points where major changes in RNA patterns were observed for marker proteins of skin functions such as keratinization, desquamation, cell-cell junctions, and lipid metabolism. Many of these marker proteins exhibit most transcriptional changes during the first 10 days of culture, followed by stabilization for the remainder of the time course (to 31 days). Based on this data, we chose to focus our proteomic analyses at culture days 3, 10, and 18 to examine early, mid, and late time points in epidermal maturation.

Materials and Methods

These experiments on human-derived samples were approved by the Western Institutional Review Board. Reconstituted epidermis cultures were prepared as described previously [16]. Briefly, human skin from surgical waste is treated to remove the endogenous epidermis and

render the dermal tissue nonviable. This prepared substrate is then seeded with primary keratinocytes isolated from individual donors (Lonza). Cultures are initially submerged in media then raised to the air-water interface at Day 3. To separate the epidermis, samples were first removed from the transwell and placed in a new 6-well plate. The sample was then covered in ammonium thiocyanate (3.8%) and incubated for 15 minutes at room temperature. Epidermis was peeled off using a dissecting scalpel, and flash-frozen in liquid nitrogen.

Sample Preparation

Isolated epidermis was incubated in 50% Trifluoroethanol (TFE), 1% SDS, 100 mM Ammonium Bicarbonate (AMBIC) at 60°C for 30 minutes. Samples were vortexed and then sonicated for 10 minutes total process time using a Misonix 3000 cup-horn sonicator on a 30% duty cycle at a power output of 75 W at 4°C. Samples were vortexed again and cleared via centrifugation. Protein content of cleared extracts was measured in triplicate with the μ BCA assay (Thermo Fisher, USA).

Experimental blocks generated consisting of one sample from each of the three time points selected by a pseudo-random number generator (*random* function in Python 2.7). Pools were then randomly generated in a similar fashion consisting of 5 blocks. The 60 initial samples were therefore combined into 20 blocks and 12 pools. Protein pools were generated by combining 50 μ g aliquots of the five component samples, yielding a 250 μ g pool. In addition, 5 μ g aliquots of each individual sample were processed separately.

Yeast alcohol dehydrogenase (Sigma Aldrich, USA) was added to each sample at 10 fmol/ μ g protein. Samples were then reduced with 5 mM DTT at 60°C for 30 minutes and alkylated with 10 mM Iodoacetamide at room temperature for 30 minutes in the dark. 100 mM AMBIC was added to dilute TFE to 5%, and trypsin was added at 1:100 enzyme:protein, to a final concentration of 2.5 μ g/ml. Digestions were performed at 37°C for 16 hours, and halted by addition of Trifluoroacetic Acid (TFA) to pH < 2. Peptides were purified/desalted on tC18 columns (Waters, USA) and dried to completion.

Individual samples were resuspended to 2.5 μ g/ μ l in 2% Acetonitrile (ACN), 0.1% TFA (loading buffer) and run on LC-MS/MS. Pooled samples were resuspended in H₂O and fractionated on 13 cm immobilized pH 3–11 strips (GE Healthcare, USA) using a 3100 OFFGEL Fractionator (Agilent, USA) according to the manufacturer specifications. The 12 fractions were again purified on tC18, dried to completion, and resuspended in loading buffer prior to injection.

LC-MS/MS

Chromatography consisted of a 2 cm trap column with 100 μ m I.D. followed by a 20 cm analytical column with 75 μ m I.D. packed with 3 μ m ReproSil-Pur C18-AQ (Dr. Maisch, Germany). The LC gradient was carried out on a Nano 2D Plus nanoLC (AB Sciex, Canada) from 0 – 20% B (0.1% formic acid in Acetonitrile) over 65 minutes, then from 20 – 40% B over 25 minutes, for a total gradient length of 90 minutes. Buffer A was 0.1% formic acid in

water, and the flow rate was set to 200 nl/min. Samples were injected onto the instrument in a random order, again selected via random number generator.

Eluted peptides from the capillary RP-HPLC column were analyzed by shotgun MS using an LTQ Velos Orbitrap (Thermo Fisher, USA). The instrument was run in data-dependent mode, with up to 20 MS² scans with CID fragmentation per MS¹ event. Dynamic exclusion was activated for 30 seconds after two observations of a given precursor ion, with a maximum exclusion list length of 500 precursors.

Mass Spectrometry Data Analysis

All data processing was performed using the Trans-Proteomic Pipeline, version 4.7 POLAR VORTEX rev. 1 [17]. Raw files were converted to mzML using ProteoWizard msConvert [18]. Resulting mzML files were searched with four separate proteomics search engines, namely Comet [19], OMSSA [20], MS-GF+ [21], and X!Tandem [22]. The search database consisted of UniRef90 human proteins [23] plus yeast alcohol dehydrogenase (spike-in standard), Glu-1-Fibrinopeptide (QC standard), trypsin, and bovine serum albumin (contaminants). Decoys were generated via pseudo-randomization and interleaved with target sequences. Data were also searched by MS2 spectral matching using SpectraST [24] against a consensus spectral library built from 6 reconstituted epidermis samples from a set of test cultures. Search results were processed with PeptideProphet [25] to return peptide identifications as a pepXML file. Resulting PepXML files from all search engines were combined with iProphet [26], and proteins were inferred using ProteinProphet. Identifications were filtered at a 1% false positive error rate according to iProphet (peptide) or ProteinProphet (protein) error models. All raw data and search results have been deposited in the PeptideAtlas [27] and are accessible at <http://www.peptideatlas.org> with the database identifier PASS00363.

The Normalized Spectral Index algorithm [28] was implemented in Python 2.7 and extended to support TPP files as input. Protein identifications were filtered at a 1% FDR based on ProteinProphet error models. Proteotypic peptides were parsed based on the ProteinProphet nondegenerate evidence flag. Fragment ion intensities for +1 charged b- and y- ions were matched and summed, then compiled to protein-level intensities. Values were then normalized based on global matched intensity and protein length. All values reported here have been log₂ transformed. Power analysis on the pilot RE quantification was performed for a variety of SI_N values using the “pwr” package in R with the following parameters: two-sample t-test, $p = 0.05$, $power = 0.8$, $\sigma = 1.41$ (calculated from pilot quantification results). K-means clustering was performed using the “kmeans” function in R. Missing values were imputed using k -nearest neighbors with $k = 100$. The within-cluster sum of squares was plotted to determine the optimal number of clusters (data not shown).

Microarray analysis

Gene expression profiling was performed on 5 matched cultures at each of the three time points described above. Epidermal samples were homogenized in Trizol (Invitrogen) and RNA was extracted according to manufacturer specifications. Extracted RNA was further purified, and analyzed using the Affymetrix Human Genome U219-96 GeneTitan array as

described previously [16]. To compare microarray data to protein abundances, intensity values for probesets mapping to a single entry in the protein database were averaged.

Results

Proteomic pilot study

In order to validate sample processing and data analysis methods, perform statistical modeling of quantification power, and generate skin-specific spectral libraries for use in the full experiment, we first analyzed two epidermal samples from each time point (culture days 3, 10, and 18) as described in the experimental section. We employed a large fraction of the organic solvent trifluoroethanol in the lysis buffer as well as vigorous sonication of the samples to maximize protein extraction from the lipid-rich cornified envelope. Given the extreme dynamic range of protein abundances in skin, we expect that fractionation of the samples would be extremely beneficial in producing more protein identifications. Due to the limited quantity of protein recovered from a single epidermal sample, it was necessary to pool several samples prior to fractionation. Therefore, protein extracts were pooled by time point prior to digestion. A portion of the pool was analyzed directly via LC-MS/MS, while another was fractionated using OFFGEL electrophoresis (OGE) [29]. All fractions were analyzed in technical singlet. Protein identifications from this experiment are summarized in Figure 1. We found that fractionation gave a strong increase in the number of protein identifications at each time point. Across all samples in this pilot experiment, 2412 proteins were identified at a 1% FDR, with nearly complete coverage of unfractionated results in the fractionated samples (Figure 1C). Despite analyzing only six samples in this test set, this number of protein identifications in a skin-related system is rivaled only by studies on undifferentiated primary keratinocytes or cell lines [12].

Quantification and power analysis

We quantified protein abundances using the normalized spectral index (SI_N) [28]. This algorithm uses a normalized sum of MS^2 fragment ion intensities of proteotypic peptides as the basis of protein quantification. We developed an implementation of the SI_N algorithm which can take protein results from the Trans-proteomic pipeline as input, schematically diagrammed in Figure 2. To inform sample size for the full experiment, we then performed a power analysis. Power analysis involves four parameters: sample size, effect size, significance (or false positive probability), and power (1 - false negative probability). Definition of any three parameters allows for calculation of the fourth. To perform the analysis, we set $power = 0.8$, and $p = 0.05$. The effect size (Cohen's d) was generated using the standard deviation of quantified results described above over a range of SI_N differences. The results are shown in Figure 1D. Based on this analysis we chose to analyze 20 samples at each time point for the full study to provide the greatest power within resource limitations.

Proteomic analysis of full sample set

Next, epidermis from an expanded set of 20 biological replicates at each of 3, 10, and 18 days in culture was processed and analyzed by the same methods utilized for the pilot experiment. In order to minimize technical bias, samples were processed in randomly

assigned blocks with 1 sample per time point per block, and injected onto the mass spectrometer in a randomized order. In addition to analyzing each individual sample, we also generated a total of 12 pools, each consisting of 50 µg protein from each of 5 randomly selected samples from a given time point. Following digestion, each pool was then separated into 12 fractions via OFFGEL electrophoresis and these were analyzed by LC-MS/MS as for the individual samples, again with randomized injection order to minimize technical bias. In total across these samples we identified 3661 proteins at a 1% false positive error rate (excluding decoys). To our knowledge this is the most comprehensive proteomic profile of a stratified epidermal system reported to date, and includes the extra dimension of development over time.

The results from the full study mirror those from the test samples. Despite increasing the sample number 10-fold, only 34% more proteins were identified. The general profile of identification number across time points is similar as well. Complexity of the observed proteome decreases over time, with 2962 proteins at day 3, 2860 at day 10, and 1906 at day 18. The distribution of abundances of the novel proteins identified in the full experiment is shifted downward compared to the total pool of IDs (Figure 3A). In addition, these new proteins are only weakly enriched for specific gene ontology biological process (GO BP) terms. These facts support the conclusion that the increased identifications of the full experiment are due to a deeper general sampling of the proteome. This is expected due to biological and sample preparation variance, stochastic precursor ion selection for fragmentation, and increasing confidence of identifications due to statistical models built into the data analysis software (TPP [26]). 46% of the identified proteins were observed at all three time points (Figure 3B). Days 3 and 10 also shared a pool of 708 proteins which were not observed at day 18. GO BP enrichment analysis [30] of identified proteins showed significant association with keratinization & keratinocyte development for those proteins observed at days 10 or 18 but not observed at day 3. Conversely, proteins observed only at day 3 were enriched for protein transport.

To ascertain the effect of fractionation on quantification, we examined the linear correlation of $\log_2(SI_N)$ values for proteins identified in both unfractionated and fractionated samples. Data from all three time points showed a good correlation, with $m = 0.98$, and $p < 2.2e^{-16}$ (Figure 4A). As fractionation affects all time points in a similar manner and the fact that relative abundance changes are the relevant metric for this study, we report quantification on the fractionated samples here. A heatmap of SI_N values is shown in Figure 4B. As expected given the decrease in identified proteins over time, many proteins which were quantified at days 3 & 10 were not quantified at day 18.

Differential abundance and clustering

Differentially abundant proteins were defined as those with a $\log_2(SI_N) > 1.43$ between two time points; the sensitivity indicated by our preliminary power analysis. Based on spike-in experiments (data not shown), we estimate this value to represent roughly a five-fold change in protein abundance. In total 1230 unique proteins show a differential abundance between at least one pair of time points. Counts of (non-unique) differential proteins for each time point comparison are shown in Table 1.

To group proteins by abundance profile, we employed k-means clustering using all SI_N data. Missing values were imputed using the k-nearest neighbor approach [31] with $k = 100$. Protein abundances were then grouped using k-means [32] into five clusters (based on within-groups sum-of-squares analysis). Resulting clusters are shown in Figure 5. Cluster 2 contains those proteins for which imputation of missing data failed at day 3 due to sparse information.

Of the five clusters, clusters 2 & 3 are enriched for proteins annotated with GO BP terms relating to keratinocyte differentiation and keratinization. These clusters also share the common profile of increased abundance over the time course. Several proteins in cluster 2 are known to be strongly up-regulated in fully differentiated skin, such as Beta-defensin 4A, Bleomycin hydrolase, and Late cornified envelope proteins 1B, 2B, & 3C.

Cluster 3 contains many high-abundance proteins related to keratinization, including Keratins 1 & 10 (among a number of other keratins), Kallikreins 5, 6, 7, 8 & 10, and Filaggrin. This cluster also contains a large number of proteins involved in cornified envelope formation including S100 family proteins, Loricrin, Involucrin, and Cystatins. Many proteins of the so-called epidermal differentiation complex (EDC) [33] are represented in clusters 2 & 3 combined.

Clusters 1 & 4, which represent decreasing proteins and those that “dip” at Day 10 respectively, are enriched for GO BP terms such as intracellular transport, translation, and protein localization. The protein families found in these clusters are heterogeneous, but include some interrelated groups such as Collagens, Serine/threonine-protein kinases, Integrins, and Myosins. Cluster 5, the largest cluster at 1914 proteins, consists of those which show little change in abundance over the time course. These include a large number of ribosomal proteins, DNA polymerases, translation initiation factors; a membership which is characteristic of ubiquitously expressed “housekeeping proteins” [34].

Quantitative analysis of specific barrier-related proteins

Next we were interested in examining specific abundance profiles for proteins previously implicated in barrier function. To avoid any distortion which could be caused by the imputation and scaling required for k-means clustering, we relied on the non-adjusted SI_N values as represented in Figure 4B for this more detailed analysis.

Desmosomes are intracellular junctions which link keratin intermediate filaments of adjacent cells [35]. Desmosome-related genes show differential expression patterns throughout the layers of the epidermis, and transgenic experiments have shown that alterations of these patterns can lead to a dysfunctional barrier [36]. We obtained good coverage of desmosome proteins in our dataset, quantifying all three Desmocollins, 3 out of 4 Desmogleins, three Plakophilins, and several additional proteins such as Plakoglobin and Desmoplakin. Observed protein abundances correlate well with previous studies [16], with Desmocollin 1, Desmoglein 1, and Plakoglobin increasing over the time course (Figure 6). Desmocollin 2 decreases, while other proteins do not change given our limits of significance.

Tight junctions are another form of cell-cell junction crucial to barrier function, specifically with regard to water and small molecule diffusion [37, 38]. Tight junction (TJ) composition is more heterogeneous than desmosomes; however we do detect many TJ-related proteins such as Claudin-1, Occludin, ZO-1 & 2, JAMA, CAR, Cingulin, and PAR3. In contrast to desmosome components, many tight junction proteins including ZO-1 & 2, Cingulin, Occludin, and E-cadherin decrease over the timecourse (Figure 6). Tight junctions form in the periderm, prior to establishment of a fully competent barrier [39]. Our observed down-regulation of tight junction proteins could be correlated with terminal differentiation and establishment of the lipid-based barrier of the outer stratum corneum, which complements several barrier functions of tight junctions [40], however additional experiments are required to confirm this hypothesis.

Integration of proteomic and transcriptomic data sets

We previously characterized the transcriptome of reconstituted epidermis over an extended time course [16]. In order to align that dataset with this novel proteomic information, we repeated the transcript analysis on matched samples at each of the three time points studied here. Comparison between the novel and previous transcript results gave very strong agreement between the two studies (data not shown), indicating that the expression time course follows a similar profile at the RNA level.

The direct correlation between protein and transcript abundance is low, with $r^2 \approx 0.25$. This is expected, as the reported r^2 for such a correlation in cultured cells is modest [41–43]. Given the nature of skin, with outer layers composed of denucleated cellular remnants which are rich in protein but not undergoing active DNA transcription, we might expect the correlation to be lower than average as we have seen here.

To overcome this issue and enable comparison between the transcriptome and proteome data sets, we employed rank-order normalization across the time course. For a given protein, we assigned integer values based on its measured abundance at each time point with 0 being the least abundant and 2 being the most abundant. To map transcript data to proteomic namespace, we averaged the values for all RNA probesets unambiguously mapping to each protein, then applied the same rank-order reassignment used for proteomic data. 574 proteins returned an exact rank-order match between proteomic & transcriptomic datasets across the three time points studied here, and are represented in Figure 7. The largest subgroup of these proteins decreases over the time course, and GO BP analysis indicates enrichment for RNA processing. Proteins which increase over the time course are enriched for epidermal development and differentiation. Selected proteins associated with these terms are shown in Table 2, and the full results are provided in Supplementary Table 1.

Discussion

We have generated a quantitative proteome profile of changes during epidermal development, finding that 1230 proteins (36%) change in abundance over the time course given the sensitivity limits of this dataset. This profile of over 3400 proteins in the reconstituted epidermis, backed up by quantitative transcriptome microarray analysis, is the most comprehensive on stratified human epithelium reported to date. The complexity of the

observed proteome as well as average protein abundance decreases over the time course. To increase power and confidence in the quantitative proteomic profile, as well as to relate this dataset to previous studies, we have collected matched transcriptomic data. By using rank-order normalization, we found 574 matched quantitative profiles between protein and transcript data corresponding to 47% of all differentially abundant proteins. These profiles constitute the highest confidence subset of proteins which change in abundance during epidermal development.

K-means clustering of protein abundances reveals two groups of proteins which increase in abundance over the time course, and which are enriched for GO terms related to epidermal development. Known protein complexes associated with epidermal barrier function were examined and their individual protein components yield similar abundance trajectories. We find that desmosomes increase in abundance while tight junctions decrease in abundance during epidermal development. The increase in desmosomes during barrier maturation and wound healing has been previously established via live-cell imaging [44]. The tight junctions, which are restricted to the granular layer, may decrease over the course of development as the lipid barrier is established at later stages and performs similar functions such as reduction of trans-epidermal water loss. At a more focused level, one of the proteins with the greatest differential abundance between 3 and 18-day reconstituted epidermis is Arginase-1, an enzyme linked to nitric oxide production in keratinocytes and hyperproliferation in psoriasis [45]. The depth of proteome coverage in this study provides detection of even low-abundance markers of skin health, such as Proactivator polypeptide-like 1, a protein which has only recently been identified as a risk marker for pediatric atopic eczema [46].

In addition to these known barrier-related complexes, the list of differentially abundant proteins presented here provides an expanded list of potential pharmaceutical targets to address conditions related to barrier dysfunction or to be used as markers in measuring efficacy of therapeutic interventions. For instance, Twinfilin-1, a protein tied to actin dynamics and cytoskeletal organization [47] shows similar regulation to Arginase-1 and Proactivator polypeptidelike 1. One example of a protein which increases over the time course in both proteomic and transcriptomic datasets is interleukin-36 gamma, which is a vital component of innate immunity and barrier function but when overexpressed has been linked to psoriasis and other inflammatory conditions. The 1230 differentially abundant proteins identified here, and especially the 574 proteins with matched transcript and protein profiles, provide many potentially novel targets of interest.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to Patrick Flores (ISB Proteomics Core) for instrumentation support and Dionne Swift (P&G statistical department) for input on the study design. This work was supported by a collaborative agreement between P&G and ISB, and in part with federal funds from the National Science Foundation MRI grant No. 0923536 and from the National Institutes of Health National Institute of General Medical Sciences under grant Nos. 2P50 GM076547/Center for Systems Biology, GM087221, and S10RR027584.

References

1. Cartlidge P. The epidermal barrier. *Semin Neonatol*. 2000; 5(4):273–280. [PubMed: 11032710]
2. Segre JA. Epidermal barrier formation and recovery in skin disorders. *J Clin Invest*. 2006; 116(5):1150–1158. [PubMed: 16670755]
3. Candi E, Schmidt R, Melino G. The cornified envelope: a model of cell death in the skin. *Nat Rev Mol Cell Biol*. 2005; 6(4):328–340. [PubMed: 15803139]
4. Furuse M, Hata M, Furuse K, Yoshida Y, Haratake A, Sugitani Y, et al. Claudin-based tight junctions are crucial for the mammalian epidermal barrier: a lesson from claudin-1-deficient mice. *J Cell Biol*. 2002; 156(6):1099–1111. [PubMed: 11889141]
5. Tunggal JA, Helfrich I, Schmitz A, Schwarz H, Gunzel D, Fromm M, et al. E-cadherin is essential for in vivo epidermal barrier function by regulating tight junctions. *EMBO J*. 2005; 24(6):1146–1156. [PubMed: 15775979]
6. Nemes Z, Steinert PM. Bricks and mortar of the epidermal barrier. *Exp Mol Med*. 1999; 31(1):5–19. [PubMed: 10231017]
7. Steinert PM, Parry DA, Idler WW, Johnson LD, Steven AC, Roop DR. Amino acid sequences of mouse and human epidermal type II keratins of Mr 67,000 provide a systematic basis for the structural and functional diversity of the end domains of keratin intermediate filament subunits. *J Biol Chem*. 1985; 260(11):7142–7149. [PubMed: 2581964]
8. Shen J, Fischer SM. Molecular profiling of the epidermis: a proteomics approach. *Methods Mol Biol*. 2010; 585:225–252. [PubMed: 19908007]
9. Hannigan A, Burchmore R, Wilson JB. The optimization of protocols for proteome difference gel electrophoresis (DiGE) analysis of preneoplastic skin. *J Proteome Res*. 2007; 6(9):3422–3432. [PubMed: 17696380]
10. Shen J, Pavone A, Mikulec C, Hensley SC, Traner A, Chang TK, et al. Protein expression profiles in the epidermis of cyclooxygenase-2 transgenic mice by 2-dimensional gel electrophoresis and mass spectrometry. *J Proteome Res*. 2007; 6(1):273–286. [PubMed: 17203971]
11. Gygi SP, Corthals GL, Zhang Y, Rochon Y, Aebersold R. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc Natl Acad Sci U S A*. 2000; 97(17):9390–9395. [PubMed: 10920198]
12. Sprenger A, Weber S, Zarai M, Engelke R, Nascimento JM, Gretzmeier C, et al. Consistency of the proteome in primary human keratinocytes with respect to gender, age, and skin localization. *Mol Cell Proteomics*. 2013; 12(9):2509–2521. [PubMed: 23722187]
13. Lee KA, Kang JW, Shim JH, Kho CW, Park SG, Lee HG, et al. Protein profiling and identification of modulators regulated by human papillomavirus 16 E7 oncogene in HaCaT keratinocytes by proteomics. *Gynecol Oncol*. 2005; 99(1):142–152. [PubMed: 16038965]
14. Okazaki M, Yoshimura K, Uchida G, Harii K. Correlation between age and the secretions of melanocyte-stimulating cytokines in cultured keratinocytes and fibroblasts. *Br J Dermatol*. 2005; 153(Suppl 2):23–29. [PubMed: 16280018]
15. Chen YQ, Mauviel A, Ryynanen J, Sollberg S, Uitto J. Type VII collagen gene expression by human skin fibroblasts and keratinocytes in culture: influence of donor age and cytokine responses. *J Invest Dermatol*. 1994; 102(2):205–209. [PubMed: 8106749]
16. Bachelor M, Binder RL, Cambron RT, Kaczvinsky JR, Spruell R, Wehmeyer KR, et al. Transcriptional profiling of epidermal barrier formation in vitro. *J Dermatol Sci*. 2014; 73(3):187–197. [PubMed: 24314759]
17. Keller A, Eng J, Zhang N, Li XJ, Aebersold R. A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Mol Syst Biol*. 2005; 1:2005 0017. [PubMed: 16729052]
18. Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol*. 2012; 30(10):918–920. [PubMed: 23051804]
19. Eng JK, Jahan TA, Hoopmann MR. Comet: an open-source MS/MS sequence database search tool. *Proteomics*. 2013; 13(1):22–24. [PubMed: 23148064]
20. Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM, et al. Open mass spectrometry search algorithm. *J Proteome Res*. 2004; 3(5):958–964. [PubMed: 15473683]

21. Kim S, Mischerikow N, Bandeira N, Navarro JD, Wich L, Mohammed S, et al. The generating function of CID, ETD, and CID/ETD pairs of tandem mass spectra: applications to database search. *Mol Cell Proteomics*. 2010; 9(12):2840–2852. [PubMed: 20829449]
22. Craig R, Beavis RC. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics*. 2004; 20(9):1466–1467. [PubMed: 14976030]
23. Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH. UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics*. 2007; 23(10):1282–1288. [PubMed: 17379688]
24. Lam H, Deutsch EW, Eddes JS, Eng JK, King N, Stein SE, et al. Development and validation of a spectral library searching method for peptide identification from MS/MS. *Proteomics*. 2007; 7(5): 655–667. [PubMed: 17295354]
25. Keller A, Nesvizhskii A, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Analytical Chemistry*. 2002; 74:5383–5392. [PubMed: 12403597]
26. Shteynberg D, Deutsch EW, Lam H, Eng JK, Sun Z, Tasman N, et al. iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. *Mol Cell Proteomics*. 2011; 10(12):M111 007690. [PubMed: 21876204]
27. Desiere F, Deutsch EW, King NL, Nesvizhskii AI, Mallick P, Eng J, et al. The PeptideAtlas project. *Nucleic Acids Res*. 2006; 34(Database issue):D655–D658. [PubMed: 16381952]
28. Griffin NM, Yu J, Long F, Oh P, Shore S, Li Y, et al. Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis. *Nat Biotechnol*. 2010; 28(1):83–89. [PubMed: 20010810]
29. Heller M, Michel PE, Morier P, Crettaz D, Wenz C, Tissot JD, et al. Two-stage Off-Gel isoelectric focusing: protein followed by peptide fractionation and application to proteome analysis of human plasma. *Electrophoresis*. 2005; 26(6):1174–1188. [PubMed: 15704250]
30. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009; 4(1):44–57. [PubMed: 19131956]
31. Ripley, BD. Pattern recognition and neural networks. Cambridge ; New York: Cambridge University Press; 1996. p. 403xi,
32. Forgy EW. Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *Biometrics*. 1965; 21:768–769.
33. Mischke D, Korge BP, Marenholz I, Volz A, Ziegler A. Genes encoding structural proteins of epidermal cornification and S100 calcium-binding proteins form a gene 21 complex ("epidermal differentiation complex") on human chromosome 1q21. *J Invest Dermatol*. 1996; 106(5):989–992. [PubMed: 8618063]
34. Zhu J, He F, Song S, Wang J, Yu J. How many human genes can be defined as housekeeping with current expression data? *BMC Genomics*. 2008; 9:172. [PubMed: 18416810]
35. Delva E, Tucker DK, Kowalczyk AP. The desmosome. *Cold Spring Harb Perspect Biol*. 2009; 1(2):a002543. [PubMed: 20066089]
36. Elias PM, Matsuyoshi N, Wu H, Lin C, Wang ZH, Brown BE, et al. Desmoglein isoform distribution affects stratum corneum structure and function. *J Cell Biol*. 2001; 153(2):243–249. [PubMed: 11309406]
37. Anderson JM, Van Itallie CM. Physiology and function of the tight junction. *Cold Spring Harb Perspect Biol*. 2009; 1(2):a002584. [PubMed: 20066090]
38. Furuse M. Molecular basis of the core structure of tight junctions. *Cold Spring Harb Perspect Biol*. 2010; 2(1):a002907. [PubMed: 20182608]
39. Kalia YN, Nonato LB, Lund CH, Guy RH. Development of skin barrier function in premature infants. *J Invest Dermatol*. 1998; 111(2):320–326. [PubMed: 9699737]
40. Downing DT. Lipid and protein structures in the permeability barrier of mammalian epidermis. *J Lipid Res*. 1992; 33(3):301–313. [PubMed: 1569381]
41. Gygi SP, Rochon Y, Franza BR, Aebersold R. Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol*. 1999; 19(3):1720–1730. [PubMed: 10022859]

42. Baliga NS, Pan M, Goo YA, Yi EC, Goodlett DR, Dimitrov K, et al. Coordinate regulation of energy transduction modules in *Halobacterium* sp. analyzed by a global systems approach. *Proc Natl Acad Sci U S A*. 2002; 99(23):14913–14918. [PubMed: 12403819]
43. Tian Q, Stepaniants SB, Mao M, Weng L, Feetham MC, Doyle MJ, et al. Integrated genomic and proteomic analyses of gene expression in Mammalian cells. *Mol Cell Proteomics*. 2004; 3(10): 960–969. [PubMed: 15238602]
44. Green KJ, Getsios S, Troyanovsky S, Godsel LM. Intercellular junction assembly, dynamics, and homeostasis. *Cold Spring Harb Perspect Biol*. 2010; 2(2):a000125. [PubMed: 20182611]
45. Bruch-Gerharz D, Schnorr O, Suschek C, Beck KF, Pfeilschifter J, Ruzicka T, et al. Arginase 1 overexpression in psoriasis: limitation of inducible nitric oxide synthase activity as a molecular mechanism for keratinocyte hyperproliferation. *Am J Pathol*. 2003; 162(1):203–211. [PubMed: 12507903]
46. Holm T, Rutishauser D, Kai-Larsen Y, Lyutvinskiy Y, Stenius F, Zubarev RA, et al. Protein biomarkers in vernix with potential to predict the development of atopic eczema in early childhood. *Allergy*. 2014; 69(1):104–112. [PubMed: 24205894]
47. Goode BL, Drubin DG, Lappalainen P. Regulation of the cortical actin cytoskeleton in budding yeast by twinfilin, a ubiquitous actin monomer-sequestering protein. *J Cell Biol*. 1998; 142(3): 723–733. [PubMed: 9700161]

Highlights

- Proteomics and transcriptomics were performed for a model of epidermal development
- 3661 proteins were identified across three developmentally crucial time points
- Quantitation revealed over 1200 differentially abundant proteins
- 574 proteins matched transcript abundance on a rank-order basis

Figure 1.

Summary of results from pilot study. A) Bar chart of protein identifications at each time point for unfractionated (orange) and fractionated (blue) samples. B) Venn diagram of protein identifications for the pilot study at each time point. C) Venn diagram of combined protein identifications for fractionated (OGE, grey area) and unfractionated samples (red area). D) Power analysis of quantification based on the standard deviation measured in the pilot study. Shown are the number of biological replicates needed vs. detectable differences in normalized spectral index (given $power = 0.8$ and $p = 0.05$).

Figure 2. Schematic diagram of quantification software. The user provides a protein result file (ProtXML) and desired FDR cutoff. The program then automatically extracts and compiles fragment ion intensities, returning $\log_2(SI_N)$ values.

Figure 3.

A) Distributions of protein abundances (averaged across time points) for all proteins identified (blue area) and those not identified in the pilot experiment (red area). B) Venn diagram of all protein identifications at a 1% FDR for all three time points studied.

Figure 4.

A) Correlation plots of unfractionated vs. fractionated SI_N values for days 3, 10, and 18 respectively. B) Heatmap of SI_N protein quantification for each time point. Results were filtered for proteins which were quantified in at least 2/3 time points. Missing data were replaced with a minimum value and appear as blocks of solid green in the image.

Figure 5.
K-means clusters of protein abundances. Missing SI_N values were imputed using 100 knearest neighbors, and scaled prior to clustering. Cluster 2 contains proteins for which imputation failed at Day 3, indicating very sparse abundance data at this time point.

Desmosomes

Tight Junctions

Figure 6.

Abundance trajectories for selected proteins related to Desmosomes and Tight Junctions. SI_N values at Day 3 were normalized to baseline to demonstrate increasing or decreasing abundance over the time course.

Figure 7. Heatmap of proteins with matching rank-order normalization profiles in proteomic and transcriptomic datasets. General trajectory trend and related biological process enrichments are shown to the right.

Table 1

Counts of differentially expressed proteins for each time point comparison

Comparison	Up	Down
D10 vs. D3	221	445
D18 vs. D10	110	396
D18 vs. D3	144	610

Table 2

Selected proteins which returned exact rank-order match between proteomic and transcriptomic data, along with their associated GO BP terms.

UniProt Accession	Name	Ordering Rank	GO BP term
P20930	Filaggrin	0-1-2	Epidermis development
P04264	Keratin 1	0-1-2	
P13645	Keratin 10	0-1-2	
P23490	Loricrin	0-1-2	
Q08188	Transglutaminase 3	0-1-2	
Q5T7P3	Late cornified envelope protein 1B	0-1-2	
P15924	Desmoplakin	0-1-2	
P52597	Heterogeneous nuclear ribonucleoprotein F	2-1-0	RNA splicing
O43390	Heterogeneous nuclear ribonucleoprotein R	2-1-0	
O75937	DnaJ (Hsp40) homolog, subfamily C, member 8	2-1-0	
Q07955	Splicing factor, arginine/serine-rich 1	2-1-0	
Q08170	Splicing factor, arginine/serine-rich 4	2-1-0	