J. Cosmet. Sci., 72, 249–267 (May/June 2021)

Insights in Human Hair Curvature by Proteome Analysis of Two Distinct Hair Shapes

EVELYNE MAES, FRASER BELL, CHARLES HEFER, ANCY THOMAS, DUANE HARLAND, ALASDAIR NOBLE, JEFFREY PLOWMAN, STEFAN CLERENS, and ANITA GROSVENOR, *Beyond Food, AgResearch Ltd., Lincoln 7674, New Zealand (E.M., A.T., D.H., J.P., S.C., A.G.), Unilever R&D, Port Sunlight, Bebington, Wirral CH63 3JW (F.B.), United Kingdom, Knowledge and Analytics, AgResearch Ltd., Lincoln 7674, New Zealand (C.H., A.N.), Biomolecular Interaction Centre, University of Canterbury, Christchurch 8041, New Zealand (S.C.), Riddet Institute, Massey University, Palmerston North 4474, New Zealand (S.C.)*

Accepted for publication January 6, 2021.

Synopsis

Scalp hair is a universal human characteristic, and a wide range of hair shape and color variations exists. Although differences in human scalp hair shape are visually apparent, the underpinning molecular insights are yet to be fully explored. This work reports the determination of differences at the protein level between two distinct groups of hair shape: very straight samples versus very curly hair samples. An in-depth high-resolution liquid-chromatography mass spectrometry proteome analysis study was performed on hair samples from 50 individuals (pooled in 10×5 samples) with very curly hair and 50 subjects with very straight hair (pooled in 10×5 samples) to decipher differences between the two experimental groups at the protein level. Our results demonstrate that a distinction between the two experimental groups (very straight vs. very curly) can be made based on their overall protein profiles in a multivariate analysis approach. Further investigation of the protein expression levels between these two groups pinpointed 13 unique proteins which were found to be significantly different between the two groups, with an adjusted p-value < 0.05 and a fold change of more than two. Although differences between the very curly and the very straight hair sample groups could be identified, linkage between population differences and curl phenotype is currently unknown and requires further investigation.

INTRODUCTION

Human scalp hair is an important component of personal identity, and the perception of hair quality measured by numerous visual appearance and tactile aspects is of crucial importance to consumers. Individual hair-care needs and aspirations are diverse, and

Address all correspondence to Anita Grosvenor at anita.grosvenor@agresearch.co.nz.

consumers with naturally curly hair have different cosmetic needs from people with straight hair (1), thus indicating that specific product design is required for hair-care products to deliver optimal benefits to each hair type.

Hair shape can be described by consumers as straight, wavy, curly, and a number of subjective variations thereof. An objective classification system has been developed to classify these different hair shapes based on a number of morphological and phenotypical parameters, independent of biogeographic or genomic ancestry. This classification system, based on the degree of curliness, starts at type I which refers to straight hair and goes to type VIII which is tightly curly hair (2).

Many factors have been discovered over the years that have been claimed to directly or indirectly contribute to hair curvature. One of the most prominent findings was that hair curvature correlates to the morphology of the hair follicle. It has been demonstrated that curly fibers which emerge from the follicle are generated within curved follicles, whereas straight hair emerges from collinear follicles, with bulbs tilted at almost right angles with the scalp (3). Rotation of the follicle by the arrector pili muscle has been suggested as responsible for extrusion of helical fiber shape (4). Others have suggested that the inner root sheath, which supports the growing hair shaft, plays a role in molding curvature due to correlations between mutations to keratins K71 and K74 and high-curl disorders (5). Challenging these extrusion views are studies that correlate internal hair-shaft features in the follicle or mature fiber with hair shape such that follicle shape is a consequence of fiber development and not *vice versa*. In the follicle, these include reported differences across the hair shaft in cell division rates (6) and keratin structure development (7,8). In the mature hair, an elliptical profile is often cited as correlated to curliness (5,9), but there are other differences across the hair shaft at the microstructural level in keratin nanostructure organization (10–12) and cortical cell remnant length (13) that also correlate with hair curvature. Two main points are that the complexity of hair growth makes discriminating between cause and effect difficult (14,15) and that hair curliness appears to be programmed within the follicle bulb (16,17).

Data that clarify the underpinning mechanism that connects the proteins and microstructure with curvature in human hair will help us not only understand hair shape generation but more directly identify differences in damage and interventions related to hair shape. In this study, we target a comprehensive understanding of the differences in hair curvature at the molecular level by applying proteomic technologies to decipher differences in the proteome between two groups with distinct hair shapes. The first group reflects hair samples with very straight hair, whereas the second group contains hair samples with very high levels of curvature.

METHODS

OVERVIEW OF SAMPLES

Hair fiber samples were sourced from 50 subjects with very straight hair, with characteristics of type I in the scale defined by Loussouarn and de la Mettrie et al. (18), and from 50 panelists with very curly hair, with characteristics of types V–VIII of the same scale. All subjects were aged between 18 and 55 years and had not chemically treated their hair for at least 6 weeks before sampling.

Prescreening was performed to identify a large number of subjects who may meet the required study criteria by review of expert assessments of hair type from the databases of the Unilever Consumer Science Centre, Unilever R&D, Port Sunlight, UK, and of a consumer research partner. Subjects identified with the required hair types were then contacted to communicate the study objectives and criteria and the hair sampling requirements; subjects meeting the study criteria and willing to participate were then invited to a study center for a further assessment. At the study center, an expert confirmed the subject hair type met the study recruitment criteria and the absence of any hair or scalp medical conditions before fiber sampling.

To sample hair fibers with minimal exposures to cosmetic treatments, or cosmetic interventions, the hair was cut by a hairdresser very close to the scalp from four locations on the head. Before sampling, the subjects' hair style was parted in the center and then again, normal to the first, to create four quadrants. Each quadrant was then parted horizontally approximately 2 cm above eye level for sampling. Approximately 50 hair fibers were cut close to each quadrant's parting along a narrow line to minimize the cosmetic impact of the hair removal. Cut samples from each region of the head were then bundled and secured in a device with the root cut end clearly identified for further processing.

SAMPLE PREPARATION

Before proteomic analysis, all the hair samples were washed in 1.4% sodium laurel sulfate with gentle agitation and rinsed twice with water. Samples were air-dried and the root-most ~2 cm was cut off and weighed. To minimize biological variation, 20 pools (A–J: very curly hair and K–T: very straight hair) were created with similar weights of hair (±10%) from five different individuals in each pool. After pooling, the samples were snipped to approximately 1-mm lengths and stored in airtight glass containers protected from light. The snipped hair samples were further rinsed 10 times with liquid-chromatography mass spectrometry (LCMS)-grade water and air-dried before mass spectrometric analysis.

PROTEOMIC ANALYSIS

Sample preparation

Proteins from 10 mg of each fine snipped subsample were extracted with 1 mL of extraction buffer containing 7 M urea, 2 M thiourea, 50 mM dithiothreitol, 50 mM tris(hydroxymethyl) aminomethane, 5% sodium deoxycholate, and pH 7.5. After 18 hours of vigorous shaking at 35°C in an Eppendorf thermomixer (Merck, Kenilworth, NJ), the solubilized proteins were precipitated with a chloroform/methanol procedure. After removal of methanol and air-drying, the protein pellet was resuspended in 100 µL 0.01 M ammonium bicarbonate. The protein concentrations of the hair extracts were quantified using a Direct Detect instrument (Merck) according to the manufacturer's instructions, and a volume equivalent to 150 µg of protein from each sample was dried in a vacuum centrifuge. The dried protein was resuspended in 0.1 M ammonium bicarbonate and chemically reduced by agitation with 20 µL of 50 mM dithiothreitol at 56°C for 45 minutes. The proteins were then alkylated by the addition of 20 µL of 150 mM iodoacetamide, with agitation at room temperature in the dark for 30 minutes. Sequencing grade trypsin (3 µg) was

added, and the digestion mixture was incubated overnight at 37°C in the presence of 10% acetonitrile to facilitate enzyme access. The dried digests were resuspended in 50 µL 0.1% formic acid and further diluted 40 times before injection on the mass spectrometer.

LC-MS/MS

All samples were analyzed in triplicate on an Ultimate 3000 UPLC system (Thermo Scientific, Waltham, MA) coupled to a maXis Impact HD mass spectrometer *via* a CaptiveSpray source equipped with a nanoBooster device (Bruker Daltonik, Bremen, Germany). For each sample, 1 µL of the sample was loaded on a C18 PepMap100 nano-Trap column (300 µm ID × 5 mm, 5 micron, 100 Å) (Thermo Scientific) at a flow rate of 3 µL/min. The trap column was then switched in line with the analytical ProntoSIL C18AQ column (100 µm ID × 150 mm, 3 micron, 200 Å). The reverse phase elution gradient was from 2% to 20% to 45% B over 60 minutes, totaling 84 minutes at a flow rate of 600 nL/min. Solvent A was LCMS-grade water with 0.1% formic acid; solvent B was LCMS-grade acetonitrile with 0.1% formic acid.

To profile quantitative protein expression patterns, the LC-MS analysis of the samples was performed in positive ion mode, with a mass range between 130 and 2,200 m/z and a sampling rate of 2 Hz. To link the expression levels with identifications, a pool of all curly and a pool of all straight samples was created, and triplicate LC-MS/MS runs of these pools were performed using data-dependent acquisition with the following settings: the same LC parameters as described before, a full scan MS spectrum, with a mass range of 50–2,500 m/z, was followed by a maximum of 10 collision-induced dissociation tandem mass spectra (350–1,500 m/z) at a sampling rate of 2 Hz for MS scans and 1–20 Hz for MS/MS (depending on precursor intensity). Precursors with charges 2+ and 3+ were preferred for further fragmentation, and a dynamic exclusion of 60 seconds was set.

Following the LC-MS run, the Q-TOF data were analyzed with Compass DataAnalysis 4.4 software (Bruker Daltonics) to evaluate the LC chromatogram and the overall quality of both MS and MS/MS spectra. From each triplicate sample run, the two LC-MS runs of highest quality were chosen to be included in the quantitative analysis.

Protein Identification

The Peaks X Studio software package (Bio informatic Solutions Inc., Waterloo, ON, Canada) was used to analyze the LC-MS/MS data. The raw data were refined by a built-in algorithm, including only ions with a charge 2+ to 5+ and allowing chimeric spectra. The proteins and peptides were identified with the following parameters: a precursor mass error tolerance of 10 ppm and fragment mass error tolerance of 0.05 Da were allowed, the UniProt human database (v2018.07; 173,324 sequences) was used, semi-trypsin was specified as digestive enzyme, and up to two missed cleavages were allowed. Carbamidomethylation of cysteine was set as a fixed modification. Oxidation (M) and deamidation (NQ) were chosen as variable modifications for the Peaks DB search, and in the optimized Peaks posttranslational modification (PTM) search, pyro-Glu from Q and carbamylation (C) were added to the variable modification list. A maximum of three PTMs per sample were permitted. False discovery rate (FDR) estimation was made based on decoy fusion. An FDR of $< 1\%$ at the peptide-spectrum match level (associated with a peptide score threshold of $-10 \log p > 22$) and a peptide PTM A-score of 50 and a protein -10lgP score > 20 were considered adequate for confident peptide and protein identification. At least one unique peptide per protein is required for both identification and quantification purposes.

Label-Free Quantification

To quantify the protein expression levels, label-free quantification (LFQ) was performed using Peaks Studio X software. Here, expression levels between the 20 curly (duplicate LC-MS runs of 10 samples) and 20 straight samples were compared. The following parameters were set: a mass tolerance error of 10 ppm and a retention time (RT) shift tolerance of 2 minutes were allowed. To determine the relative protein abundances in the LC RT-aligned samples, peptide feature–based quantification was performed. Comparison of the relative levels of each protein identified in the samples was based on the LC peak areas of up to three unique peptides per protein that were automatically selected by the software. A further manual inspection of each of the selected unique peptides was performed, and the top three best unique peptide candidates were selected for LFQ abundance estimations.

Statistical analysis

To assess whether proteins were expressed in different amounts between the curly and the straight hair groups, protein area values were compared between the curly and straight groups using a Welch t-test, followed by Benjamini–Hochberg correction for the p values. A p value ≤ 0.05 was considered as statistically significant. Protein area values were screened before the means comparison so that only those observed in at least 75% of samples in either the curly or the straight hair group were compared between the groups. This screening excluded protein area values not consistently observed in both groups from analysis, to ensure that comparison of the groups was only made on reliable protein area values. All statistical analyses were carried out with statistical software R (version 4.0.0).

Data visualization

Partial least square-discriminant analysis (PLS-DA) plots were generated with the mixOmics package (19), volcano plots with the EnhancedVolcano package (version 1.7.2, https://github.com/kevinblighe/EnhancedVolcano), and expression abundance plots using ggplot. All graphs were generated using *R* (version 4.0.0).

RESULTS

In this study, proteomic technologies were applied to determine whether differences in protein abundance could be observed between the very straight hair samples and the very curly hair samples. To achieve this goal and to minimize the effect of biological variance between individuals, pools of five curly or five straight hair samples of our batch were created. Ten very curly pooled samples (named sample A to sample J) and 10 very straight pooled hair samples (sample K to sample T) were used for further analysis. Proteomic analysis was performed on a high-resolution mass spectrometer instrument, and state-of-the-art data analysis tools were applied, to maximize the number of identified peptide sequences and thus protein identifications.

Overall, 362 proteins were be identified across all samples, with 217 confident protein identifications that were detected and quantified in at least 75% of all samples. To ensure rigorous statistical analysis, further analysis was only based on the 217 proteins that were detected and quantified in most samples. An overview of these proteins is provided in Appendix 1.

DIFFERENCES IN PROTEIN PROFILES

To investigate differences in protein profiles between samples, a multivariate data analysis approach was performed using a supervised partial least squares-discriminant analysis (PLS-DA). When comparing the protein profiles of the 217 quantified proteins in each pool, the pooled samples belonging to the very curly hair group or the very straight hair group were able to be distinguished based on the protein abundance profiles of the proteins (Figure 1).

To better understand which proteins were the key contributors in the discrimination between these two groups in the PLS-DA, the top 20 proteins that contributed the most to this discrimination in the first component were determined. An overview of the protein groups they belong to is listed in Table I. Keratin 85 was the strongest contributor to the distinction of the very straight hair shaft pools, whereas KAP 13-2 had the most discrimination power in the very curly hair pooled samples.

DIFFERENCES IN PROTEIN ABUNDANCE

Next, a univariate analysis was performed to detect proteins whose abundance differed according to the level of curliness. Two different parameters in this statistical analysis are of major importance: (i) the *p* value: how significantly different was the protein abundance

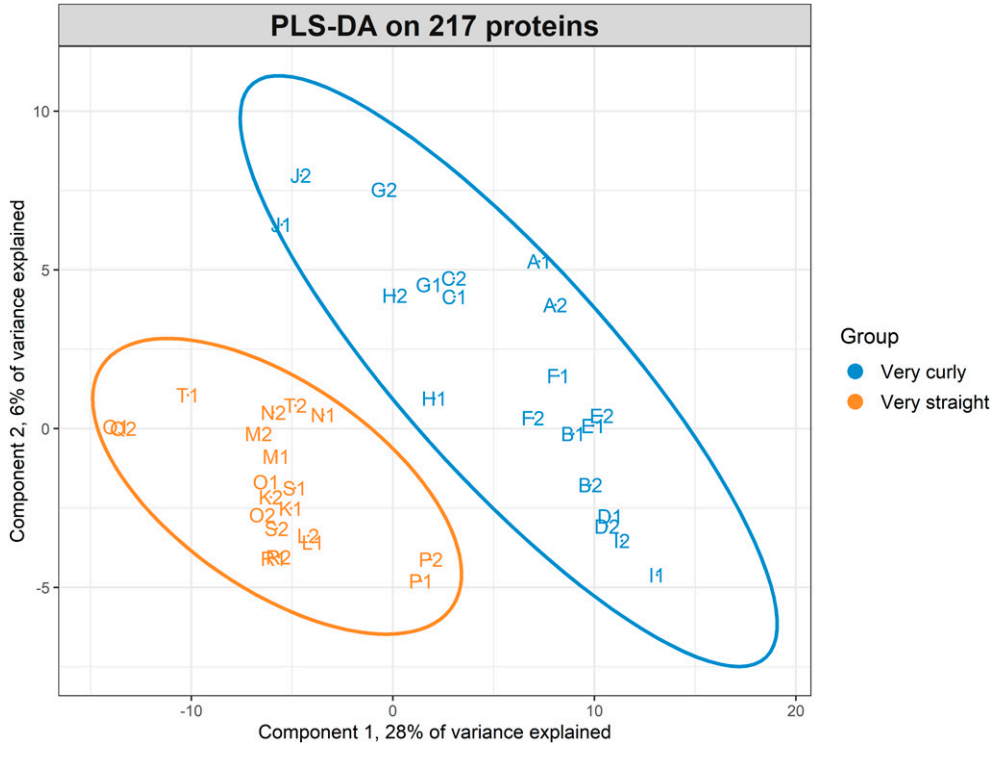

Figure 1. Multivariate analysis *via* PLS-DA enables clustering of the 20 pooled samples measured in duplicate (A1 to T2) based on the protein abundance levels of 217 proteins present in the samples. A clear discrimination between the very curly hair group and the very straight hair group is seen.

Table I
Overview of the proteins most important in discriminating the two experimental groups based on PLS-DA

Protein group	Group contributor	Importance
K85	Very straight	−0.1429
Methanethiol oxidase	Very straight	−0.1308
K34	Very straight	−0.1299
K81	Very straight	−0.1127
Protein S100-A3	Very straight	−0.1122
Histone H2AX	Very straight	−0.1098
KAP 9-8	Very straight	−0.1083
KAP 2-1	Very straight	−0.1082
KAP 4-2	Very straight	−0.1062
K38	Very curly	0.1069
K10	Very curly	0.1083
KAP 3-3	Very curly	0.1099
Protein band 4.1	Very curly	0.1169
KAP 4-4	Very curly	0.1170
KAP 13-2	Very curly	0.1192

difference between the two experimental groups, and (ii) the fold change: how large was the difference in protein abundance between the two groups. An overview of these data is represented in the volcano plot in Figure 2. In this volcano plot, every dot represents a specific protein species. All gray dots are proteins that were not statistically significantly different between the two experimental groups. The purple and yellow dots represent the proteins that passed the significance threshold and are therefore of high interest. Each of these proteins had a significant p value (so the proteins were significantly different between the two experimental groups), and the proteins were also at least twofold higher or lower in abundance in a specific experimental group. The yellow dots represent proteins with higher abundances in the very straight samples, whereas the purple dots represent proteins that were higher in abundance in the very curly hair shape samples.

Interestingly, 14 unique proteins in our dataset displayed a significant difference in the abundance level (adjusted p-value ≤ 0.05) between the very curly versus the very straight hair group with at least a twofold difference.

Table II lists all the significant proteins (most significant protein listed first), with their respective significance level (adjusted p value *via* Benjamini–Hochberg correction) and fold change. Some proteins (e.g., K1) detected in these ethnically diverse samples are known to exist in several protein isoforms, and as these different isoforms are detected with the same peptides, only one of them is displayed in the table. Of these 14 proteins of interest, eight proteins have a higher abundance in the very straight hair samples, whereas six proteins have a higher abundance level in the very curly hair samples.

Because the fold-change calculations were based on mean values across the different pools within the same experimental group, additional analysis of the variability of the protein abundances within and between these experimental groups can generate additional insights. In Figure 3, the abundance levels of the 14 proteins found to have significantly different expression levels between the curly hair groups and the straight hair groups are represented. For each protein, the relative abundance levels for every pooled sample (thus 20 datapoints for very straight and 20 datapoints for very curly hair) are represented in a boxplot. The upper, middle, and lower lines represent the first quantiles, medians, and third quantiles

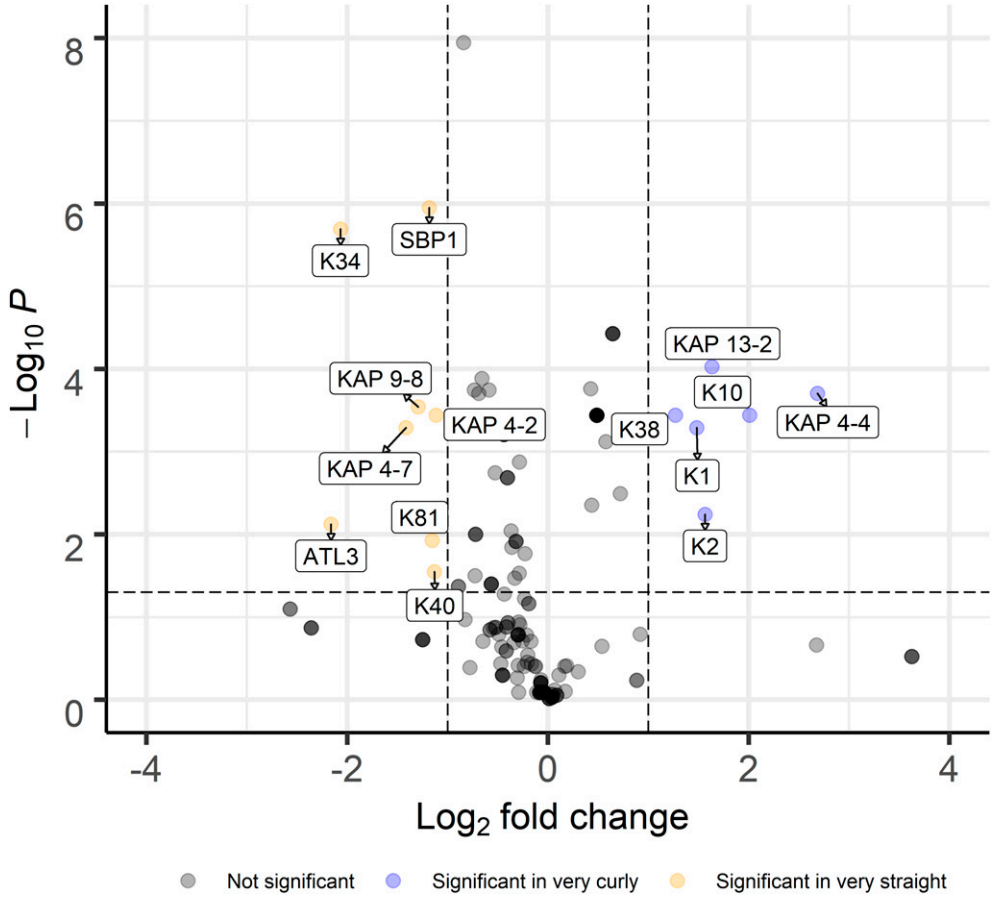

Figure 2. Volcano plot representing the differential abundancies of proteins in very curly versus very straight hair samples is observed. On the *x*-axis, the log2 fold change of each protein is represented, whereas on the *y*-axis, the -$\log_{10}$(*p*-value) is shown. Proteins that have a high fold change and a significant difference in the abundance level with an adjusted *p* value are of high interest; these proteins are labeled in the graph. The yellow dots on the left represent proteins higher in abundance in the very straight samples, whereas the purple dots on the right represent proteins that were higher in abundance in very curly hair shape samples.

of the abundance levels, respectively. The larger these rectangular boxes in each boxplot are, the more variation between the protein abundance levels within an experimental group is seen. For KAP4-7, e.g., a large variation of protein abundances between individuals was seen within the very curly group, whereas a very tight distribution of protein abundances was seen in the straight hair sample group.

DISCUSSION

In this study, we sought to explore whether differences could be found at the proteome level when we compared two groups of human hair shafts with a very different grade of curliness: very straight hair versus very curly hair. Because hair fibers are considered to have large intra- and inter-sample phenotypical variabilities, a pooled

Table II
List of proteins showing differences in abundance between very curly versus very straight hair samples. The proteins in the list have a significant p value of $p < 0.05$ (which is corrected for multiple testing *via* Benjamini–Hochberg) and a fold change of at least 2 ($\log_2$ fold change = 1)

Protein	Fold change (very curly/very straight)	Adjusted p value	Full protein description
SBP1	−1.181081415	1.13E-06	Methanethiol oxidase
K34	−2.066207146	2.04E-06	Keratin 34
KAP 13-2	1.634917291	9.42E-05	Keratin-associated protein 13-2
KAP 4-4	2.68765369	0.000198	Keratin-associated protein 4-4
KAP 9-8	−1.2927477	0.000288	Keratin-associated protein 9-8
K10	2.010710525	0.000362	Keratin 10
KAP 4-2	−1.112588238	0.000362	Keratin-associated protein 4-2
K38	1.271281473	0.000362	Keratin 38
K1	1.484219137	0.000514	Keratin 1
KAP 4-7	−1.414189772	0.000514	Keratin-associated protein 4-7
K2	1.567179628	0.005741	Keratin 2
ATL3	−2.161080371	0.007585	ADAMTS-like protein 3
K40	−1.129873144	0.028445	Keratin 40
K81	−1.153647498	0.011886	Keratin 81

strategy was adopted, whereby each sample consisted of a pool of five individuals. Thus, from the 50 individual samples available from each experimental group, 10 pools were created. Each pool was measured in duplicate to evaluate consistency of the protein profile within and between pools.

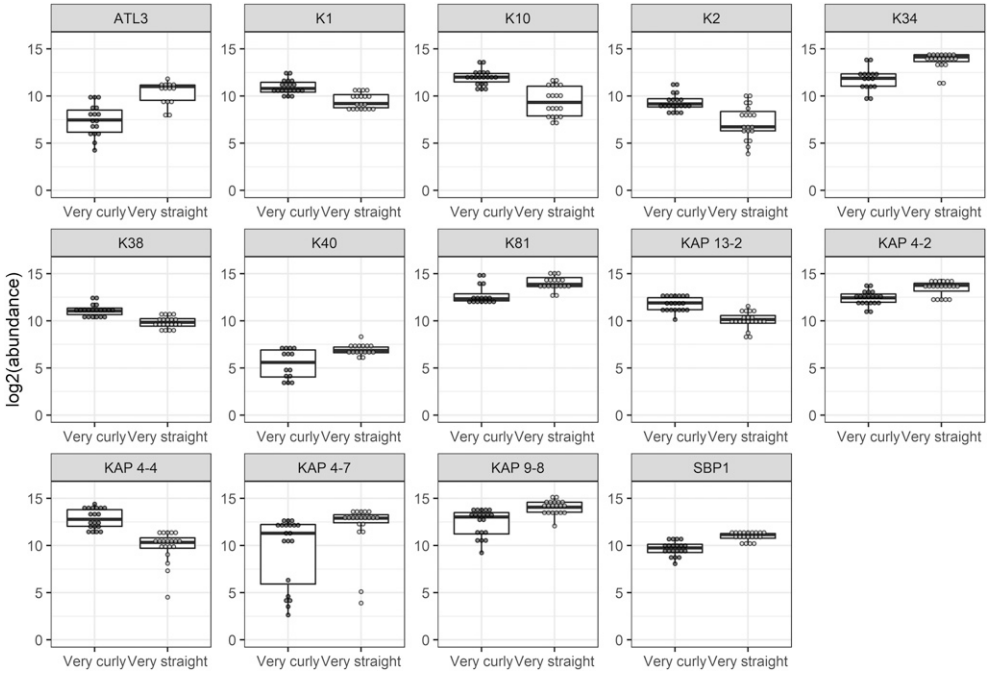

Figure 3. Abundance levels of the 14 proteins found to be significantly different between the two hair types. The upper, middle, and lower lines represent the first quantiles, medians, and third quantiles of the abundance levels, respectively.

Overall, we identified 362 different proteins across all samples but restricted the statistical comparisons to 217 protein species for which we had confident protein identifications and that were quantified in at least 75% of the samples. A twofold approach was taken for the analysis of the proteomics data. First, a supervised multivariate clustering method was used to compare overall protein profiles between the different samples. Second, univariate analyses were performed to find proteins that had significantly different abundance levels between the two experimental groups.

In the supervised multivariate analysis, a clear discrimination of the two experimental groups was observed based on two components (Figure 1). Key observations were as follows: first, many of the duplicates in the pooled samples were in close proximity to each other. This means that the overall protein profiles in these duplicate samples were very similar, which indicates good proteomic profiling performance. Second, a clear difference between the very straight and the very curly samples was seen when taking into account the overall profiles of the 217 confidently identified proteins per sample, and this discrimination was mainly based on the first component (x-axis) of the PLS-DA plot, which explains 28% of the variance.

Differences in structural proteins are likely to inform us about potential differences in hair performance and those associated with insults and interventions. When looking at the top 20 proteins contributing to the discrimination of the curly versus straight hair shape in the first component (Figure 1, Table I), most can be linked to structural protein components of hair, that is, keratin or KAP families, with keratin K85, the strongest contributor to the distinction of the very straight hair shaft pools, whereas KAP 13-2 has the most discrimination power in the very curly hair pooled samples. Interestingly, K85, one of the first keratins to be expressed during the hair growth process, has been shown to be expressed across the entire developing shaft in straight human hair (20) and in curly Wiltshire wool, but in the latter, its distribution is asymmetric (21). However, no evidence for the link between curly fibers and KAP 13-2 has previously been reported in the literature.

Next, univariate analyses were performed for pairwise comparisons between the two experimental groups, which highlighted proteins with different levels of expression between the very straight hair group and the very curly hair group. Interestingly, 14 proteins were identified as significantly differentially expressed (Table II). From these 14 proteins, the majority can be linked to the keratin or KAP families, indicating microstructural composition differences are important for fiber shape. The remaining nonstructural proteins are interesting because they may echo differences in processes of fiber growth and maturation (e.g., cornification) between fibers of different shapes.

As hair shafts are known to display marked intra- and between-person variation, we opted to visualize the protein abundance levels of each of the 20 samples measured per experimental group to see whether large differences in these abundances could be found across the samples within an experimental group. Interestingly, for many proteins, the detected abundance levels were quite similar within each experimental group, with only minimal differences between the first quantiles, median, and third quantiles in the boxplot, indicating that the sample pooling approach was an efficient way to minimize variability across samples (Figure 3). The proteins found to have statistically significant fold changes of greater than two are a good starting point for defining protein differences between these hair shape sample populations. However, little is known about the functions of most of these proteins in the hair fiber.

PROTEINS WITH A PREVIOUS LINK TO CURL IN HUMAN OR SHEEP HAIR

Interpretation of the functions of structural proteins (especially keratins and KAPs) is hindered by the less differentiated cortical organization of human hair than that of sheep wool. All mammalian hair is made of modified cell remnants, with the "cells" composed of structural components called macro-fibrils (keratin filament bundles with a KAP matrix), digested waste left over from the cytoplasm and nucleus, and sometimes melanin granules. The cells are glued to one another by a cell membrane complex that is composed of highly transformed plasma membranes, cell junction complexes, and extracellular matrix (22). In wool fibers, cortical cell remnants form clusters of similar structural organization and protein chemistry. The orthocortex contains helically twisted macro-fibrils and is dominated by high–glycine–tyrosine KAP species, and the paracortex contains macro-fibrils with less twist and a matrix dominated by high-sulfur KAP species (23,24). The cortex characteristics of human hair cell types are less clear cut (25), macrofibril helical twist varies across the cortex (26), and most cells have high sulfur content.

Notable among the structural proteins found to be more abundant in the curly human hair samples is K38 (Figure 2, Table II). K38 has a history of association with hair curl in humans and sheep. In follicles from straight human hair, its distribution across the hair shaft cortex is sporadic but uniform (27), but in Western blot studies of curly human hair, K38 was found on the concave side of the curvature and evenly distributed in straight hair (28). In sheep follicles, K38 expression is restricted to orthocortical cell remnants (21) and has been noted as potentially playing a key role in keratin polymerization and structure self-assembly (29). In sheep wool, the orthocortex is dominated by twisted fibril architectures within which keratin filaments are embedded in a mixture of high–glycine–tyrosine KAPs (21,24) and, when associated with one side of the cortex, is associated with high curvature fibers (12,30,31). Because cortical cell types are less clear-cut in human hair than in wool (25), the association between K38 and different KAP families is unlikely to be the same. However, our finding does raise the intriguing prospect that K38 may be associated with a particular level of macrofibril internal twist or with a particular set of KAP species.

KAP4-2, KAP4-4, and KAP9-8 are all ultrahigh-sulfur proteins that, like other KAPs, are major constituents of the matrix between keratin filaments in macro-fibrils. In wool, members of the KAP4 family have been found to be expressed primarily in the paracortex (21,30,32). Thus, the prominence of KAP4-2 and KAP9-8 in the straight hair fibers would be consistent with them having some aspect in common with wool paracortex, but whether this is structural (e.g., less intense macrofibril internal twist) or in terms of protein chemistry has yet to be established.

KAP13-2, a high-sulfur family KAP, was found in higher amounts in curly hair (Figure 2, Table II). Although no direct link has been reported in the literature, another protein from the KAP13 family, KAP13-1, is present in significantly higher amounts in curly sheep wool fibers than in the same-diameter straight fibers (Plowman et al. 2020 unpublished results).

PROTEINS NOT PREVIOUSLY ASSOCIATED WITH HAIR CURL

K34 and K81 are cortical proteins that have previously been observed to be uniformly distributed across the follicle in both human straight hair and sheep wool. K40 is notably one of the last keratins to be expressed during the hair formation process and is found

only in the cuticle of beard and scalp hair (33). Other than K85, which is expressed in both cuticle and cortex, K40 is the only cuticle protein of interest. Interestingly, a recent microstructure study found statistically fewer cuticle layers in curly hair of individuals of African genetic ancestry than those with straight hair from European and East Asian genetic ancestries (9).

K1, K2, and K10 are epithelial keratins, normally playing a role in the follicle root sheathes rather than the hair shaft (34). They are, however, found in the medulla of hair fibers (35). It is likely that these are associated with the peculiar assembly process and structure of the medulla by which, possibly random, combinations of various keratin species (trichocyte and epithelial) are brought together to form the medulla structures (36).

The two differential non-keratin proteins, SBP1 and ADAMTS-like protein 3 (ALT3), had higher abundance in the straight hair samples. Punctin (also known as ALT3), when it occurs in living tissue, is an extracellular matrix protein associated with epithelial cells (37). During hair development, the extracellular material is transformed into cell-type–specific continuous structured cement called the cell membrane complex (22,38). Although identified as significantly different between the two groups, it should be noted that this is based on just one peptide identification, and hence, should be considered with extreme care. Nevertheless, one intriguing possibility is that the higher abundance observed in straight hair is an indication of differences in cortical cell lengths across the cortex between straight and curly hair similar to that correlated to the extent of fiber curvature in wool (39).

Finally, methanethiol oxidase is a cellular selenium-binding protein (SBP1) expressed in multiple tissues, also previously found in hair samples (40), which is likely a leftover from fiber development. However, its function such as interacting with thiol groups to potentially generate formaldehyde molecules (41) may indicate an unknown role in the final stages of hair maturation, which is an environment in which sulfur chemistry is critical to the function.

In conclusion, this proteome analysis between two extreme groups of hair shape indicates that differences at the protein abundance level can provide useful insights into the physiochemistry underpinning differences in hair shape and the relevant hair growth processes which should inform technological interventions for hair care. However, although differences between the very curly and the very straight hair sample groups could be identified, the linkage between biogeographic genetic differences and curl phenotype is currently unknown and would require further controlled investigation.

REFERENCES

(1) R. D. Sinclair, Healthy hair: what is it? *J. Invest. Dermatol. Symp. Proc.*, 12(2), 2–5 (2007).
(2) R. De la Mettrie, D. Saint-Leger, G. Loussouarn, A. Garcel, C. Porter, and A. Langaney, Shape variability and classification of human hair: a worldwide approach, *Hum. Biol.*, 79(3), 265–281 (2007).
(3) B. Lindelof, B. Forslind, M. A. Hedblad, and U. Kaveus, Human hair form. Morphology revealed by light and scanning electron microscopy and computer aided three-dimensional reconstruction. *Arch. Dermatol.*, 124(9), 1359–1363 (1988).
(4) R. E. Chapman, "The ovine arrector pili musculature and crimp formation in wool," in Biology of the Skin and Hair Growth, A. G. Lyne, B. F. Short. Eds. (Angus and Robertson, Sydney, Australia, 1965), pp. 201–232.
(5) G. E. Westgate, R. S. Ginger, and M. R. Green, The biology and genetics of curly hair. *Exp. Dermatol.*, 26(6), 483–490 (2017).

(6) P. I. Hynd, N. M. Edwards, M. Hebart, M. McDowall, and S. Clark, Wool fibre crimp is determined by mitotic asymmetry and position of final keratinisation and not ortho- and para-cortical cell segmentation, *Animal*, **3**(6), 838–843 (2009).
(7) D. P. Harland and A. J. McKinnon, "Macrofibril formation," in *The Hair Fibre: Proteins, Structure and Development*. Advances in Experimental Medicine and Biology, Vol. 1054, J. E. Plowman, D. P. Harland, and S. Deb Choudhury, Eds. (Springer, New York, NY, 2018), pp. 155–169.
(8) J. McKinnon and D. P. Harland, A concerted polymerization-mesophase separation model for formation of trichocyte intermediate filaments and macrofibril templates 1: relating phase separation to structural development. *J. Struct. Biol.*, **173**(2), 229–240 (2011).
(9) S. L. Koch, M. D. Shriver, and N. G. Jablonski, Variation in human hair ultrastructure among three biogeographic populations, *J. Struct. Biol.*, **205**(1), 60–66 (2018).
(10) W. G. Bryson, D. P. Harland, J. P. Caldwell, J. A. Vernon, R. J. Walls, J. L. Woods, S. Nagase, T. Itou, and K. Koike, Cortical cell types and intermediate filament arrangements correlate with fiber curvature in Japanese human hair, *J. Struct. Biol.*, **166**(1), 46–58 (2009).
(11) D. P. Harland, R. J. Walls, J. A. Vernon, J. M. Dyer, J. L. Woods, and F. Bell, Three-dimensional architecture of macrofibrils in the human scalp hair cortex, *J. Struct. Biol.*, 185(3), 397–404 (2014).
(12) D. F. G. Orwin, J. L. Woods, and S. L. Ranford, Cortical cell types and their distribution in wool fibres, *Aust. J. Biol. Sci.*, **37**, 237–255 (1984).
(13) D. P. Harland, J. A. Vernon, J. L. Woods, S. Nagase, T. Itou, K. Koike, D. A. Scobie, A. J. Grosvenor, J. M. Dyer, and S. Clerens, Intrinsic curvature in wool fibres is determined by the relative length of orthocortical and paracortical cells, *J. Exp. Biol.*, **221**(Pt 6), 1–9 (2018).
(14) Y. S. Lim, D. P. Harland, and T. L. Dawson, Jr., Wanted, dead and alive; why a multidisciplinary approach is needed to unlock hair treatment potential, *Exp. Dermatol.*, **28**, 517–527 (2019).
(15) E. Cloete, N. P. Khumalo, and M. N. Ngoepe, The what, why and how of curly hair: a review, *Proc. R. Soc. A*, **475**(2231), 20190516 (2019).
(16) S. Thibaut, O. Gaillard, P. Bouhanna, D. W. Cannell, and B. A. Bernard, Human hair shape is programmed from the bulb, *Br. J. Dermatol.*, **152**(4), 632–638 (2005).
(17) S. Thibaut and B. A. Bernard, The biology of hair shape, *Int. J. Dermatol.*, **44**(Suppl. 1), 2–3 (2005).
(18) G. Loussouarn, A. L. Garcel, I. Lozano, C. Collaudin, C. Porter, S. Panhard, D. Saint-Leger, and R. de La Mettrie, Worldwide diversity of hair curliness: a new method of assessment, *Int. J. Dermatol.*, **46**(Suppl. 1), 2–6 (2007).
(19) F. Rohart, B. Gautier, A. Singh, and K. A. Le Cao, mixOmics: an R package for 'omics feature selection and multiple data integration, *PLos Comput. Biol.*, **13**(11), e1005752 (2017).
(20) L. Langbein, M. A. Rogers, M. A. Winter, S. Praetzel, and J. Schweizer, The catalog of human hair keratins. II. Expression of the six type II members in the hair follicle and the combined catalog of human type I and II keratins, *J. Biol. Chem.*, **276**(37), 35123–35132 (2001).
(21) Z. Yu, S. W. Gordon, A. J. Nixon, C. S. Bawden, M. A. Rogers, J. E. Wildermoth, N. J. Maqbool, and A. J. Pearson, Expression patterns of keratin intermediate filament and keratin associated protein genes in wool follicles, *Differentiation*, **77**(3), 307–316 (2009).
(22) J. E. Plowman and D. P. Harland, "Fibre ultrastructure," in *The Hair Fibre: Proteins, Structure and Development. Advances in Experimental Medicine and Biology*, 1st Ed., J. E. Plowman, D. P. Harland, and S. Deb-Choudhury. Eds. (Springer Nature, Singapore, 2018), pp. 3–13.
(23) J. P. Caldwell, D. N. Mastronarde, J. L. Woods, and W. G. Bryson, The three-dimensional arrangement of intermediate filaments in Romney wool cortical cells, *J. Struct. Biol.*, **151**(3), 298–305 (2005).
(24) D. P. Harland, J. P. Caldwell, J. L. Woods, R. J. Walls, and W. G. Bryson, Arrangement of trichokeratin intermediate filaments and matrix in the cortex of merino wool, *J. Struct. Biol.*, **173**(1), 29–37 (2011).
(25) W. G. Bryson, D. P. Harland, J. P. Caldwell, J. A. Vernon, R. J. Walls, J. L. Woods, S. Nagase, T. Itou, and K. Koike, Cortical cell types and intermediate filament arrangements correlate with fiber curvature in Japanese human hair, *J. Struct. Biol.*, **166**(1), 46–58 (2009).
(26) D. P. Harland, R. J. Walls, J. A. Vernon, J. M. Dyer, J. L. Woods, and F. Bell, Three-dimensional architecture of macrofibrils in the human scalp hair cortex, *J. Struct. Biol.*, **185**(3), 397–404 (2014).
(27) L. Langbein, M. A. Rogers, H. Winter, S. Praetzel, U. Beckhaus, H. R. Rackwitz, and J. Schweizer,. The catalog of human hair keratins. I. Expression of the nine type I members in the hair follicle, *J. Biol. Chem.*, **274**(28), 19874–19884 (1999).
(28) S. Thibaut, P. Barbarat, F. Leroy, and B. A. Bernard, Human hair keratin network and curvature, *Int. J. Dermatol.*, **46**(Suppl. 1), 7–10 (2007).

(29) A. J. McKinnon, D. P. Harland, and J. L. Woods, Relating self-assembly to spatio-temporal keratin expression in the wool follicle, *J. Textile Eng.*, **62**(6), 123–128 (2016).
(30) S. W. Li, H. S. Ouyang, G. E. Rogers, and C. S. Bawden, Characterization of the structural and molecular defects in fibres and follicles of the Merino felting lustre mutant, *Exp. Dermatol.*, **18**(2), 134–142 (2009).
(31) J. E. Plowman, D. P. Harland, D. R. Scobie, D. O'Connell, A. Thomas, P. H. Brorens, M. Richena, E. Meenken, A. J. Phillips, J. A. Vernon, and S. Clerens, Differences between ultrastructure and protein composition in straight hair fibres, *Zoology (Jena)*, **133**, 40–53 (2019).
(32) B. C. Powell, A. Nesci, and G. E. Rogers, Regulation of keratin gene expression in hair follicle differentiation, *Ann. N. Y. Acad. Sci.*, **642**, 1–20 (1991).
(33) L. Langbein, M. A. Rogers, S. Praetzel-Wunder, D. Bockler, P. Schirmacher, and J. Schweizer, Novel type I hair keratins K39 and K40 are the last to be expressed in differentiation of the hair: completion of the human hair keratin catalog, *J. Invest. Dermatol.*, **127**(6), 1532–1535 (2007).
(34) R. M. Lavker, T.-T. Sun, H. Oshima, Y. Barrandon, M. Akiyama, C. Ferraris, G. Chevalier, B. Favier, C. A. B. Jahoda, D. Dhouailly, A. A. Panteleyev, and A. M. Christiano, Hair follicle stem cells, *J. Invest. Dermatol. Symp. Proc.*, **8**(1), 28–38 (2003).
(35) H. J. Stark, D. Breitkreutz, A. Limat, C. M. Ryle, D. Roop, I. Leigh, and N. Fusenig, Keratins 1 and 10 or homologues as regular constituents of inner root sheath and cuticle cells in the human hair follicle, *Eur. J. Cell Biol.*, **52**(2), 359–372 (1990).
(36) L. Langbein, H. Yoshida, S. Praetzel-Wunder, D. A. Parry, and J. Schweizer, The keratins of the human beard hair medulla: the riddle in the middle, *J. Invest. Dermatol.*, **130**(1), 55–73 (2010).
(37) B. H. Koo, T. Hurskainen, K. Mielke, P. P. Aung, G. Casey, H. Autio-Harmainen, and S. S. Apte, ADAMTSL3/punctin-2, a gene frequently mutated in colorectal tumors, is widely expressed in normal and malignant epithelial cells, vascular endothelial cells and other cell types, and its mRNA is reduced in colon cancer, *Int. J. Canc.*, **121**(8), 1710–1716 (2007).
(38) D. P. Harland and J. E. Plowman, "Development of hair fibres," In. *The Hair Fibre: Proteins, Structure and Development. Advances in Experimental Medicine and Biology*. Vol. 1054. J. E. Plowman, D. P. Harland, and S. Deb Choudhury, Eds. (Springer, New York, NY, 2018), pp. 109–154.
(39) D. P. Harland, J. A. Vernon, J. L. Woods, S. Nagase, T. Itou, K. Koike, D. A. Scobie, A. J. Grosvenor, J. M. Dyer, and S. Clerens, Intrinsic curvature in wool fibres is determined by the relative length of orthocortical and paracortical cells, *J. Exp. Biol.*, **221**, jeb172312 (2018).
(40) Y. J. Lee, R. H. Rice, and Y. M. Lee, Proteome analysis of human hair shaft: from protein identification to posttranslational modification, *Mol. Cell. Proteomics*, **5**(5), 789–800 (2006).
(41) A. Pol, G. H. Renkema, A. Tangerman, E. G. Winkel, U. F. Engelke, A. P. M. de Brouwer, K. C. Lloyd, R. S. Araiza, L. van den Heuvel, H. Omran, H. Olbrich, M. Oude Elberink, C. Gilissen, R. J. Rodenburg, J. O. Sass, K. O. Schwab, H. Schäfer, H. Venselaar, J. S. Sequeira, H. J. M. Op den Camp, and R. A. Wevers, Mutations in SELENBP1, encoding a novel human methanethiol oxidase, cause extraoral halitosis, *Nat. Genet.*, **50**(1), 120–129 (2018).

Appendix 1

Overview of the 217 proteins identified and quantified in at least 75% of the samples. Proteins (but not protein groups) are listed with the obtained sequence coverage, number of peptides identified for that protein, number of unique peptides identified per protein. Additionally, some proteins can be linekd to the proteomic procedure used and are indicated with an 'x' as artefact.

Protein Description	Sequence Coverage (%)	#Peptides	#Unique peptides	Artefact
Keratin-associated protein 2-3	99	41	28	
Keratin 36	97	44	4	
Keratin type I cytoskeletal 39	96	27	3	
Keratin-associated protein 3-1	96	26	2	
Keratin-associated protein 4-2	96	26	2	
Keratin 81	93	105	1	
Keratin 83	93	100	5	
Keratin 85	92	63	15	
Keratin 85	91	96	8	
Keratin 33B	90	82	16	
Keratin 32	90	57	12	
Keratin-associated protein 2-4	90	43	5	
Keratin 33A	88	68	17	
Keratin-associated protein 9-9	87	41	2	
Keratin 31	85	88	1	
Keratin 38	84	37	5	
Histone H4 L	84	22	20	
Keratin-associated protein 9-8	83	55	1	
Keratin-associated protein 4-6	83	25	1	
Keratin-associated protein 11-1	82	23	13	
Keratin-associated protein 1-1	81	35	4	
Keratin-associated protein 9-6	80	31	1	
Keratin-associated protein 13-2	78	15	3	
Keratin-associated protein 9-3	75	39	1	
Trypsin I (Fragment)	75	15	7	x
Keratin associated protein 4-6	73	29	12	
cDNA FLJ25625 fis clone STM02974	72	19	5	
PRSS1 protein	71	15	6	x
Keratin-associated protein 4-3	70	19	2	
Trypsin-1 (Fragment)	69	22	13	x
Keratin-associated protein 9-4	68	33	1	
Protease serine 1 (Fragment)	68	7	6	x
Trypsin-1	67	12	3	x
Trypsin-6	66	16	1	x
Histone H4 H	62	18	6	
Protease serine 1 (Fragment)	62	6	1	x
Trypsin-2	62	6	1	x
Histone H4 A	57	16	3	
V-set and immunoglobulin domain-containing protein 8	57	16	3	
Protease serine 1 (Fragment)	56	12	2	x
Anionic trypsinogen	56	16	1	x
Protein S100-A3	55	19	1	
PRSS2 protein (Fragment)	47	6	6	x
Protease serine 2 preproprotein	47	6	6	x
Protease serine 2	43	9	7	x
Keratin-associated protein 3-3	37	22	2	
Transcription factor 4	36	4	4	

Appendix 1
Continued

Protein Description	Sequence Coverage (%)	#Peptides	#Unique peptides	Artefact
Histone H2AX	36	4	4	
TJP3 protein	36	4	4	
Histone H3	34	4	4	
tetratricopeptide repeat domain 21B	30	2	2	
Potassium voltage-gated channel subfamily H member 8	29	2	2	
Protease serine 1 (Fragment)	26	11	10	x
Keratin-associated protein 2-1	25	9	2	
TRANSCRIPTION FACTOR 4	25	2	2	
LIM domain only protein 7	24	1	1	
LIM domain only 7 protein	24	1	1	
LIM domain only protein 7	24	1	1	
LIM domain only protein 7	24	1	1	
Sensor protein	24	1	1	
Keratin-associated protein 4-4	22	8	4	
Trypsin-2	22	10	9	x
Histone H2B type 1-C/E/F/G/I	20	2	2	
Protease serine 2	19	11	1	x
Trypsin-1	19	11	1	x
Protease serine 2 preproprotein	19	11	1	x
Prolow-density lipoprotein receptor-related protein 1	19	11	1	
Epididymis secretory sperm binding protein Li 134P	19	11	1	
Methanethiol oxidase	19	11	1	
Histone H3	19	2	1	
Histone H3 (Fragment)	19	2	1	
Histone H3 B	19	2	1	
Histone H3.1	19	2	1	
Histone H3.2	19	2	1	
Histone H3.3	19	2	1	
Histone H3.1t	19	2	1	
Histone H3.3	19	2	1	
NEDD4-binding protein 1	19	2	1	
Epididymis secretory protein Li 50	19	2	1	
Ubiquitin C	19	2	1	
Ubiquitin-40S ribosomal protein S27a	19	2	1	
Helicase/primase complex	19	2	1	
Nucleolar protein 8	19	2	1	
Nucleolar protein 8	19	2	1	
DNA polymerase epsilon catalytic subunit	19	2	1	
DNA polymerase epsilon catalytic subunit	19	2	1	
DNA polymerase epsilon catalytic subunit A	19	2	1	
DNA polymerase epsilon catalytic subunit	19	2	1	
Mitochondrial import inner membrane translocase subunit TIM50	19	2	1	
Plexin-B3	19	2	1	
LIM domain only protein 7	18	1	1	
NOL8 protein	17	2	1	
Keratin-associated protein 1-5	16	9	1	
Histone H2B	15	2	2	
Proliferation marker protein Ki-67	14	2	1	
DNA polymerase epsilon catalytic subunit	14	2	1	

Appendix 1
Continued

Protein Description	Sequence Coverage (%)	#Peptides	#Unique peptides	Artefact
Splicing factor proline/glutamine-rich protein	14	1	1	
Keratin 10	13	5	1	
Histone H2B	12	2	2	
Histone H3 B	11	2	2	
SFPQ protein	11	1	1	
Histone H2B	11	2	2	
Histone H2B type 2-E	11	2	2	
Keratin-associated protein 1-3	10	11	2	
Keratin-associated protein 4-7	10	11	2	
Histone H2B (Fragment)	10	2	2	
Histone H2B type 1-B	10	2	2	
Histone H2B type 1-J	10	2	2	
Histone H2B type 3-B	10	2	2	
Histone H2B type 1-O	10	2	2	
Keratin-associated protein 13-1	10	2	2	
NPC1-like intracellular cholesterol transporter 1	10	1	1	
TIGHT JUNCTION PROTEIN ZO-1	9	5	2	
Keratin 1	9	4	4	
Keratin 2	9	4	4	
Keratin-associated protein 7-1	9	2	2	
LIM domain 7	8	1	1	
Splicing factor proline- and glutamine-rich protein	8	1	1	
RASGRF1 protein OS	8	1	1	
Ubiquitin carboxyl-terminal hydrolase 6	8	1	1	
Zinc finger protein 40	8	1	1	
DNA polymerase theta variant	8	1	1	
Polymerase (DNA directed) theta isoform CRA	8	1	1	
DNA polymerase theta	8	1	1	
DNA polymerase theta	8	1	1	
DNA polymerase theta	8	1	1	
Uncharacterized protein	8	1	1	
Keratin 1 OS=Homo sapiens OX=9606 GN=KRT1 PE=3 SV=1	8	1	1	
Histone H2B type F-S	7	2	2	
Keratin 34	7	1	1	
Fibronectin type-III domain-containing protein 3A	7	1	1	
Kinectin 1	6	1	1	
Trinucleotide repeat-containing gene 6B protein	6	1	1	
Microtubule-associated serine/threonine-protein kinase 4	6	1	1	
Keratin 1	5	1	1	
Uncharacterized protein (Fragment)	4	1	1	
Histone H2B type 1-N	4	2	2	
FAST kinase domain-containing protein 1	4	1	1	
NPC1-like intracellular cholesterol transporter 1	4	1	1	
Spectrin-like protein	3	1	1	
Nesprin-1	3	1	1	
Zinc finger homeobox protein 4	3	1	1	
Zinc finger homeobox protein 4	3	1	1	
Zinc finger homeodomain 4	3	1	1	
DNA-directed RNA polymerase III subunit RPC2	3	1	1	

Appendix 1
Continued

Protein Description	Sequence Coverage (%)	#Peptides	#Unique peptides	Artefact
DNA-directed RNA polymerase subunit beta	3	1	1	
BTB/POZ domain-containing protein 8	3	2	2	
Fibronectin type III domain containing 3A	3	2	2	
Conserved oligomeric Golgi complex component 1	3	1	1	
Histone H3	2	1	1	
Histone H2B	2	1	1	
Histone H2B type 1-K	2	1	1	
Histone H2B	2	1	1	
KRT83 protein	2	1	1	
Nesprin-1	2	1	1	
Nesprin-1	2	1	1	
Kinectin	2	1	1	
FLJ20980 fis clone	2	1	1	
Receptor protein-tyrosine kinase	2	1	1	
NPC1-like intracellular cholesterol transporter 1	2	1	1	
Keratin 1 (Fragment)	2	1	1	
CTD small phosphatase-like protein 2	2	1	1	
CTD small phosphatase-like protein 2	2	1	1	
Coiled-coil domain-containing protein 129	2	1	1	
Keratin 1	1	1	1	
Keratin 1	1	1	1	
Histone H2B	1	2	2	
Histone H2B type 1-D	1	2	2	
EPB41 protein	1	1	1	
Protein 4.1	1	1	1	
Probable ribonuclease ZC3H12B	1	1	1	
Zinc finger CW-type PWWP domain protein 2	1	1	1	
HCTP4	1	1	1	
Keratin-associated protein 4-9	1	1	1	
Ryanodine receptor 3	1	1	1	
Ryanodine receptor 3	1	1	1	
Ryanodine receptor 3	1	1	1	
Ryanodine receptor 3	1	1	1	
Cell division cycle-associated protein 2	1	1	1	
cDNA FLJ78763	1	1	1	
RNA binding motif protein 10 isoform 2	1	1	1	
FAST kinase domain-containing protein 1	1	1	1	
Keratin-associated protein 9-2	1	1	1	
Neurexin-3-beta	1	1	1	
Neurexin-3	1	1	1	
Neurexin-3-beta	1	1	1	
1-phosphatidylinositol 4 5-bisphosphate phosphodiesterase gamma-1	1	1	1	
Keratin 40	1	1	1	
T-cell antigen CD7	1	1	1	
ATP-binding cassette sub-family A member 5	1	1	1	
ATP-binding cassette protein	1	1	1	
Nipped-B-like protein	1	1	1	
Nipped-B protein	1	1	1	
MX2	1	1	1	
Interferon-induced GTP-binding protein	1	1	1	

Appendix 1
Continued

Protein Description	Sequence Coverage (%)	#Peptides	#Unique peptides	Artefact
Keratin 1 (Fragment)	1	1	1	
Histone H2B type 1-M	1	1	1	
Histone H2B	1	1	1	
Histone H2B type 1-H	1	1	1	
Histone H2B	1	1	1	
Histone H2B type 1-L	1	1	1	
HIST1H2BC protein	1	1	1	
Histone H2B type 2-F	1	1	1	
Spectrin beta chain	1	2	2	
Mucin-2	1	2	2	
Protein 4.1	1	2	2	
EPB41 protein	1	2	2	
Collagen type XI alpha 1	1	1	1	
Collagen type XI alpha 1	1	1	1	
Keratin-associated protein 4-1	1	1	1	
Neurexin-3-beta	1	1	1	
Conserved oligomeric Golgi complex subunit 1	1	1	1	
Conserved oligomeric Golgi complex subunit 1	1	1	1	
Keratin-associated protein 4-1	1	1	1	
Carboxy-terminal domain RNA polymerase II polypeptide A	1	1	1	
ADAMTS-like protein 3	1	1	1	
Uncharacterized protein	1	1	1	