The author(s) shown below used Federal funds provided by the U.S. Department of Justice and prepared the following final report:

Document Title:	**Human Hair Proteomics - Improved Evidence Discrimination**
Author(s):	**Robert H. Rice, Pei-Wen Wu, Selena M. Mann**
Document No.:	**249102**
Date Received:	**August 2015**
Award Number:	**2011-DN-BX-K543**

This report has not been published by the U.S. Department of Justice. To provide better customer service, NCJRS has made this federally funded grant report available electronically.

Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Final Technical Report

Title: Human Hair Proteomics - Improved Evidence Discrimination
Award Number: 2011-DN-BX-K543
Authors: Robert H. Rice, Pei-Wen Wu, Selena M. Mann
Collaborators: Chelsea N. Laatsch, Blythe Durbin-Johnson, David M. Rocke, Richard A. Eigenheer, Brett S. Phinney, Sophie Mukwana

Abstract

This project has developed a method to obtain potentially probative information from human hair shaft samples by analyzing their profiles of constituent proteins. We hypothesize such information will increase the forensic value of hair as evidence. Although it is well known that hair from different individuals can be distinguished by physical properties, more quantitative methods of comparison are sought. Individual protein differences within the human hair shaft, not thoroughly investigated until recently, could offer a novel approach to identification of suspects. To this end, the project was devised in two parts, first to determine whether hair from different individuals is distinguishable by shotgun protein profiling and second to develop a set of proteotypic peptides (obtained reproducibly in high yield) that could be used in targeted analyses (multiple reaction monitoring) to distinguish hair samples in a streamlined fashion. Pursuant to previous work with inbred mouse strains, the first phase of shotgun proteomic analysis revealed considerable variation in profile among hair samples from Caucasian, African-American, Kenyan and Korean subjects. Within these ethnic groups, prominent keratin proteins served to distinguish individual profiles. Differences between ethnic groups, less marked, relied to a large extent on levels of keratin associated proteins. In samples from Caucasian subjects, hair shafts from axillary, beard, pubic and scalp regions exhibited distinguishable profiles, with the last being most different from the others. Identification of proteotypic peptides for streamlined sample comparisons appeared feasible by inspection of the results from shotgun analysis. However, multiple reaction monitoring in the second phase of targeted proteomic analysis was not. The high content of hydrophobic keratin peptides rapidly degraded the instrumental capability so that quantitation and reproducibility were deficient. Instead of pursuing this direction by expensive and time consuming avenues, two other important questions were addressed. First, we asked whether identical twins can be distinguished by their hair protein profiles. We collected triplicate samples of scalp hair from 8 pairs of twins, processed them and submitted them for shotgun analysis. The results indicated that the twin pairs were readily distinguishable from each other, but distinguishing the two twins of a given pair was more difficult (but not impossible) since the protein profiles were more similar. Second, we have modified the original standard protocol for hair processing to permit analysis of small sample amounts. The original protocol involves recovery of hair protein by ethanol precipitation from sodium dodecyl sulfate solution. Although fractional losses with starting amounts of ≈4 milligrams of hair are small (several percent), samples of 0.1 mg or less would be expected to suffer high fractional losses. Replacing the sodium dodecyl sulfate with sodium dodecanoate or ammonium perfluoro-octanoate permits detergent removal by organic solvent extraction at low pH and thus direct application to the mass spectrometer after digestion. Use of sodium dodecanoate instead of sodium dodecyl sulfate was demonstrated to be successful with the hair samples from identical twins. Future research areas of investigation to which our results point include effects on individuals of aging, further characterization of hair samples from different anatomic sites and devising statistical methods to compare and search proteomic profiles. Present findings so far support the hypothesis that proteomic analysis could increase the value of hair in crime scene evidence for distinguishing among individuals.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table of Contents

	Page
Abstract	1
Table of Contents	2
Executive Summary	2-6
Introduction	6-8
Methods	8-11
Results	11-17
Conclusions	17-19
References	19-21
Dissemination of Research Findings	22

Executive Summary

Problem. Hair shafts are often found at crime scenes but are underutilized as evidence. Lacking genomic DNA from the bulb, the shaft contribution to identification of the donor generally is mitochondrial. While useful in providing exclusionary evidence, the latter provides only limited evidence for personal identification due to its small size and direct transmission through the maternal lineage. In view of the widely acknowledged variation in gene expression among individuals in the human population, the possibility exists that polymorphic levels of proteins in the hair shaft could provide a basis for distinguishing individuals. Barriers to exploring that possibility and ultimately to exploiting it have included the difficulties of identifying the proteins and ascertaining their relative amounts thanks to the complexity of the hair shaft and its content of hydrophobic proteins cross-linked covalently by disulfide bonds and transglutaminase-mediated isopeptide bonds. Moreover, the hair shaft is difficult to process to a state amenable to proteolytic digestion with high efficiency by the enzyme trypsin, the most effective protease for this purpose due to its reproducibility of cleavage to yield analyzable peptides. Thanks to the advent of mass spectrometric analysis of protein digests with database searching, these barriers have been overcome. We have been able to catalog hair shaft proteins and to judge relative amounts in samples processed in parallel. Applying this approach to analysis of hair from inbred mouse strains, a survey of 11 strains and 5 mutant stocks across a wide spectrum of estimated evolutionary divergence, over 100 proteins exhibited significant differences in amount according to statistical analysis. Effects on the protein profile of single gene mutations causing hair shaft defects were profound. Inasmuch as a collection of inbred mouse strains with respect to the gene pool is analogous to an outbred species such as humans, these results strongly suggested that human hair would display individual differences.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Purpose. The hypothesis of the present project is that individual humans can be distinguished on the basis of their hair shaft protein profiles. The ultimate purpose of this investigation is to establish the feasibility of using hair proteomic analysis to assist in connecting hair evidence to hair donors. The degree of difference among individual donors is an important consideration as well as the variance in protein levels from a single donor. The most important outcome of the project is to establish whether this proteomic approach merits further exploration. To this end, the first specific aim of the project was to analyze hair samples from various donors by shotgun mass spectrometry. Comparison of the profiles was envisioned to reveal whether individuals could be distinguished in this way. The second specific aim was to develop a targeted method of comparing proteins by their proteotypic peptides obtained reproducibly in high yield. This was envisioned to simplify the analysis and increase the sensitivity. Several additional factors will remain to be addressed if the results are sufficiently promising.

Research Design. With a major focus on hair from the scalp, the first aspect of the work involved collecting hair samples from numerous individuals. On the basis of the earlier mouse experiments, we anticipated that the variance from triplicate samples would permit sufficient discrimination among subjects. However, in most cases, the present samples were collected from three different regions of the scalp and analyzed independently in parallel. This strategy guarded against underestimating the variance in protein levels if the profile differs substantially in different locations. To find whether ethnic differences in protein levels could be discerned, we concentrated on subjects of four ethnic origins. We collected hair samples from non-related individuals of Caucasian, African-American, Korean and Kenyan origin. For the last group, we were fortunate to have the help of a collaborator in Nairobi, Kenya, a graduate of our Forensic Science Program now in private practice there. We analyzed samples from 5 subjects of each group except the Caucasian one (6 subjects). In addition to collecting scalp hair, we analyzed a sample of axillary, beard and pubic hair from 4 male Caucasian subjects. Previous work with mouse strains did not find significant differences in hair from males or females of the inbred strains and, although we did not study this aspect specifically, sex-specific differences were not seen in the present work.

Except as indicated, processing the hair samples followed the procedure developed for human and mouse hair in earlier studies. After being rinsed twice in 2% sodium dodecyl sulfate (SDS) to remove loosely adherent invisible extraneous material (e.g., small dust particles), the hair was reduced with dithioerythritol in the presence of 2% SDS and 0.1 M phosphate buffer (pH 7.8), and then sulfhydryl groups were alkylated with iodoacetamide. Under this condition, the hair becomes swollen and flexible, and up to half of the protein is solubilized. The soluble and insoluble proteins were recovered as a flocculent precipitate upon addition of ethanol (final concentration 70%). SDS and other small molecules were removed by centrifugation. The recovered protein was rinsed with 67% ethanol and then with 0.1 M ammonium bicarbonate, resuspended in the bicarbonate buffer adjusted to 10% in acetonitrile and digested for one day at room temperature after adding reductively methylated trypsin (1% by mass). The digestion was continued for two more days after adding an equal aliquot of trypsin each day. During this period, the remaining hair shaft fragments were pulverized by magnetic stirring. Reductively methylated trypsin is not readily susceptible to autolysis and, although stabilized in this way, retains activity better when the digestion is conducted at room temperature. After the digestion period, the released (solubilized) peptides were recovered by centrifugation and submitted for mass spectrometric analysis. The insoluble material was rinsed with water several times to remove ammonium bicarbonate and then quantitated by reaction with ninhydrin after digestion to amino acids using sulfuric acid.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Digested samples were analyzed for released peptides by the Proteomics Core facility at the University of California, Davis mostly using a Thermo-Finnigan LTQ iontrap mass spectrometer. The latest samples (from the twin study) were analyzed using a Thermo Q-Exactive mass spectrometer. Tandem mass spectra were extracted with Xcalibur, analyzed by searching a Uniprot human database with X!Tandem and validated with Scaffold proteome software. Because certain keratins are well known to contain identical peptides, we used a distributed spectral count (called weighted spectral count in Scaffold) to model spectral counts more accurately across the proteins identified in this study. For statistical analysis, total hair samples from different subjects were compared using mixed-effects overdispersed Poisson regression models, including a fixed effect for ethnicity (to adjust for imbalances between total hair donors) and a random effect for subject. Hair from different sites (scalp, axillary, facial, and pubic) were likewise compared using mixed-effects overdispersed Poisson regression models, including a fixed effect for site and a random effect for subject. The comparisons were conducted for each protein that was present in sufficient amount. When the comparison showed a significant difference ($p<0.05$) globally for a given protein among the samples being compared, pairwise comparisons between subjects or sites were conducted using the Tukey HSD method.

Findings. Within each ethnic group, individual hair profiles were readily distinguishable. Among the 76 proteins that were abundant enough to be amenable to statistical testing, pairwise comparisons of the individual profiles gave as few as 2 and as many as 40 differences, indicating the individuals in each group were distinguishable. The averages for each group ranged from 15 (Kenyan samples) to 25 differences (Korean samples). Similar to the earlier analyses of mouse hair, the keratins (K) were major contributors to discriminating ability, especially K32, K33B, K34, K83 and to a lesser extent K31, K35, K39 and K86. A consequence of the high variability of protein levels within ethnic groups, distinguishing among samples between ethnic groups was less striking. In two way comparisons, the range of differences extended from 3 to 8 proteins. In this regard, keratin associated proteins were most useful. Within the Caucasian group, the samples of scalp hair analyzed above were compared to single samples of hair from other bodily locations. When analyzed as a group, the scalp hair profiles were quite different from the others. Axillary and public were the most similar to each other, and beard hair was intermediate between these and scalp hair.

Once the proteins (up to perhaps a dozen) most useful in distinguishing among individuals were determined, our original strategy was to identify proteotypic peptides (those detected reproducibly in high yields) in these proteins to use in targeted analyses. Such a focused approach would be anticipated to be more sensitive and simpler to analyze. Unfortunately, the high content of keratin-derived peptides presented technical problems, primarily degrading peptide separation and detection, resulting in loss of reproducibility and quantitative precision. The possibility of simplifying the digests by removal of unwanted peptides, an ideal goal, appeared impractical by biochemical means. However, this problem might be amenable to an immunochemical solution. Successful targeting of proteotypic peptides has been demonstrated using enrichment of digests with specific anti-peptide antibodies, a commercially available technology for limited purposes from SISCAPA Assay Technologies (http://www.siscapa.com/index.html). While attractive, the development of such antibodies in the present instance was judged beyond the budget and time frame of this project. We elected to concentrate on resolving more immediate uncertainties in determining the utility of protein profiling for forensic samples. The targeted approach is seen as an aid to simpler analyses but not critical for ultimate success in discriminating among individuals even using low sample amounts.

An important issue to be addressed is the dependence of protein profile on individual genetics. If a person's hair protein profile is genetically determined, then establishing a database of

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

profiles would be anticipated to be feasible and likely useful. On the other hand, if the genetic endowment is not the predominant factor in the protein profile, then the effect of environment must be explored before deciding on the usefulness of profiling. For that reason, we solicited the cooperation of the Twins Research Registry maintained by SRI International Biosciences Division in California, which focuses on twins in the San Francisco Bay Area. Hair samples from 9 sets of identical twins were received, processed and submitted for mass spectrometry in parallel to a tenth pair recruited locally. The results showed that the profiles of each twin from a pair were much more similar to each other than to unrelated individuals, consistent with a dominant contribution of genetics to the hair protein profile. Nevertheless, differences were observed between twins of a pair, and some twin pairs differed this way more than others. No effect of age on the differences within twin pairs was evident, consistent with the lack of effect of age on the profiles or with age-related changes occurring in the same direction within a pair.

A second important issue to be addressed concerns the processing of hair to maximize the sensitivity of the analysis. The current protocol uses 4-5 milligrams of hair. This is convenient and minimizes effects of contamination by adventitious protein in the environment (especially desquamated epidermis in dust). However, mass spectrometric analysis is exquisitely sensitive, requiring 1 µg of protein or less. The present protocol is not suited to small sample sizes due to the ethanol precipitation step. In this step, although satisfactory for milligram amounts of protein, a considerable fraction (perhaps the majority) of a small sample is likely to be lost. Thus, use of detergents other than SDS was explored. Commercially available detergents compatible with mass spectrometry did not permit sufficient disaggregation of hair structure and thus were unsatisfactory for hair. By contrast, two detergents with long alkyl side chains, but containing carboxyl instead of sulfate charged moieties, were as effective as SDS for human hair. Thus, sodium dodecanoate and ammonium perfluoro-octanoate were removable after hair reduction and alkylation by organic solvent extraction after acidification. Subsequent adjustment of the pH to 8 permitted trypsinization and subsequent mass spectrometry. Sodium dodecanoate was employed in processing the twin samples, demonstrating its effectiveness.

Conclusions

● Present experiments verify the potential utility of hair profiling. Individual differences in hair shaft proteomic profile are clearly evident, confirming predictions based on analysis of hair from inbred mouse strains. Based on pairwise comparisons of individual proteins, the differences are substantial. The results so far permit us to speculate that this method may have discrimination capability similar (but complementary) to that of mitochondrial DNA.

● Differences between hair samples from different ethnic origins are detectable, but the degrees of difference are smaller than among individuals. Among various protein classes, the keratin associated proteins appear most useful in such discrimination and merit further examination.

● Protein profiles of hair from different body sites are distinguishable. A worthwhile goal for the future, entailing finding the variance of profiles from various sites in an individual, would be to determine whether hair from the scalp and from other body sites of a single individual can be predicted from the profile of one of them.

● Processing of small hair samples is now possible by modification of our standard treatment procedure. This observation raises the likelihood that small samples will have sufficient material to be divided for both protein profiling and mtDNA analysis.

● As mass spectrometry instrumentation continues to increase in sensitivity, more proteins will be amenable to inclusion in the analyses, increasing the power of comparisons among samples

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

from individuals. For example, the newly introduced Thermo Q-Exactive instrument has several fold the sensitivity of the LTQ instrument used in most of the present work. It appears to improve the resolution among samples, but this needs further testing.

● The results of our twin study suggest a dominant contribution of genetics to hair protein profiles, but the role of epigenetic differences is unclear. While some differences in gene expression patterns could occur during development, our data are consistent with the influence of environmental factors after childhood being small by contrast. An influence of ageing cannot be ruled out for hair profiles but, if it exists, it appears to be either small or the same for each twin. The influence of ageing could be investigated in humans or, more easily, in inbred strains of mice.

Main Body

I. Introduction

1. Statement of the problem

This project has developed a method to obtain potentially probative information from human hair samples by analyzing their profiles of constituent proteins. Such information could increase the forensic value of hair as evidence.

2. Literature citations and review

Hair is commonly recovered at crime scenes. It is readily lost by individuals, who shed ≈100 scalp hairs/day (Robertson, 1999), it is easily transferred among individuals or to inanimate objects, and it is quite durable, leading to its frequent recovery. Microscopic examination of hair evidence can take into account features such as pigmentation (granule distribution and density, spectral analysis), cosmetic treatment (dye, bleach), diameter, appearance in cross-section and abnormalities of the shaft (cuticle, cortex, medulla). Careful examination may determine the anatomic site of origin of human hair and can distinguish it from animal hair. Using all the features available permits useful discrimination whether a given hair originates from a specific individual (Gaudette, 1999). Thus, the Scientific Working Group on Materials Analysis recommends that hair analysis begin with a thorough microscopic examination (Scientific Working Group on Materials Analysis, 2009). Nevertheless, since this method has a substantial error rate (National Research Council, 2009), the search for more objective criteria by which to judge hair matches continues (Taupin, 2004).

Hair evidence is ordinarily supplemented by DNA evidence whenever possible (Rowe, 2001). When follicle tissue is present, nuclear DNA extracted from a hair sample may identify the donor. In exceptional situations, nuclear DNA may be found in the hair shaft, a polymorphic trait (Szabo, et al, 2012), or attached to the shaft of unclean hair (Robertson, et al, 2007). In the majority of cases, only mitochondrial DNA (mtDNA) from the shaft is available. mtDNA can provide valuable exclusionary evidence, but is not sufficient alone for individual identification. The present approach explores whether proteomic information can be obtained that can assist suspect identification or to offer an investigative lead. The information from proteomic characterization is complementary to that from microscopic examination and DNA analysis.

Hair structure. Hair shaft, nail plate and epidermal callus consist of terminally differentiated keratinocytes called corneocytes. Mature corneocytes are designed by nature to resist external

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

physical stress and chemical exposures. Virtually none of the protein in the hair shaft can be extracted, even under strongly denaturing conditions, without damaging the protein unless reducing agent is added to break disulfide bonds. A convenient protein extraction method employs the detergent sodium dodecyl sulfate (SDS) in the presence of a strong reducing agent such as dithioerythritol (DTE). Most of the constituent corneocyte protein, primarily keratins and keratin associated proteins, can be extracted under denaturing conditions, but 15-20% is not solubilized due to transglutaminase-mediated isopeptide cross-linking.

Electron microscopy of hair cross-linked structures after extraction reveals intricate architectural features (Rice, et al, 1994). In cross-section, the boundaries of cuticle, cortex and medulla cells are easily distinguished. Typically, cells of the cortex exhibit melanin pigment granules but little else inside the cell boundaries, while those in the medulla contain amorphous material and remnant nuclei. The latter are obvious in longitudinal sections. Medulla cells are less obvious in human hair. Such pictures indicate that corneocytes are composed largely of keratin, but they contain proteins from throughout the cell including membranes and nuclei. In contrast to other cells in the shaft, those of the cuticle appear full of non-extractable protein and incompletely degraded organelles.

As evidence of individual variation that can occur, our examination of extracted mouse (Rice, et al, 1999a) and human hair (Rice, et al, 1996) has permitted visualization of characteristic defects. Normally, exhibiting little effect of vigorous detergent extraction, a dramatic increase in the degree of protein loss is visible in the cuticle cells from individuals afflicted by lamellar ichthyosis (Rice, et al, 1999b). Loss of the marginal band (A layer) can be seen in some cases of congenital ichthyosiform erythroderma (Rice, et al, 2005). These changes are attributable to defective cross-linking by keratinocyte transglutaminase (TGM1). Analogous changes are also observable in corneocytes of human nail plate and epidermal callus (Rice, et al, 2003; Rice, et al, 2005).

While the proteins solubilized from hair shaft can be isolated and studied by routine biochemical procedures, the identities of the remaining cross-linked proteins was largely a mystery until they were characterized in our recent study using “shotgun” mass spectrometry (Lee, et al, 2006). In this procedure, a mixture of proteins is examined without purification of individual components. The proteins are all digested with trypsin to produce peptides that are then separated and matched by mass/charge ratio to a database containing peptide masses generated by tryptic proteolysis of all the proteins in that species. The results typically identify >100 proteins in a given sample, including novel components, a benefit of this discovery approach. Improvements in sensitivity of mass spectrometers are increasing the numbers of proteins identified. Note that a small fraction of peptides would not be matched in the database searching because they participate in isopeptide cross-links that are not hydrolyzed by proteolysis. We previously determined that only ≈ 15% of the protein lysines participate in the cross-linking (Rice and Green, 1977; Rice, et al, 1994), so this would not prevent protein identification.

By necessity, our work has focused on the corneocyte cross-linked proteome. Very little protein is solubilized from hair shaft, nail plate or epidermal callus by aqueous solutions without denaturants and reducing agent. This is not unexpected, since these mature structures are subject to frequent exposure to water (and soap) during bathing. Thus, samples can be rinsed in ionic detergent such as SDS to remove adventitious contamination with little or no loss of constituent protein. Typical sample preparation involves subsequent extensive extraction in SDS plus DTE. Samples can be extracted 4-5 times in this way to separate nearly all the protein held in the complex by disulfide bonding from the isopeptide cross-linked insoluble protein

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

(Rice, 2011). However, analysis is streamlined and more discriminating without separation into solubilized and insoluble components (Rice, et al, 2012).

Hair proteomics – mouse model. Mouse strains provide monomorphic populations to test the efficiency of analytical methods before progressing to the generally highly pleomorphic human population. An inbred mouse strain is analogous to a single human with both sexes. The availability of numerous mouse strains with anomalous hair structures raised the possibility that proteomic analysis could be useful in characterizing molecular bases for the defects. AKR mice, for example, display hair with an "interior defect" consisting of haphazard patterns of medulla cell arrangement due to mutations in steroyl O-acetyltransferase 1 (Wu, et al, 2010). In strains without this defect, where medulla cells are typically in an orderly alignment with regular spacing, electron microscopy of mature hair shafts reveals projections from the cortex into the middle of each medulla cell. The AKR strain lacks these indentations, evidently permitting irregular or disorderly spacing of the medulla cells. Proteomic analysis revealed only a low level of trichohyalin, a major component of the projections, in the cortical cells of AKR mice compared to two other strains not showing this phenotype. In the course of this work, it was noted that all three mouse strains were distinguishable by their proteomic profiles (Rice, et al, 2009). Further work has corroborated the finding that inbred mouse strains have distinguishable hair protein profiles (Rice, et al, 2012). In a survey of 11 strains and 5 mutant stocks of mice across an estimated wide spectrum of evolutionary divergence (Petkov, et al, 2004), over 100 proteins exhibited significant differences in amount according to statistical analysis. Effects on the protein profile of single gene mutations causing hair shaft defects were profound.

3. Statement of hypothesis and rationale

Supported by our previous results with mouse strains, our hypothesis was that the human population harbors substantial differences in hair protein profile, consistent with the known wide variation in gene expression that occurs among individual humans (Storey, et al, 2007). Such variations may have served as selective features during human evolution as groups migrated from Africa to populate the rest of the world (Lappalainen and Dermitzakis, 2010). This variation likely occurs as a consequence of DNA sequence polymorphisms at a variety of, apparently abundant, regulatory loci (Skelly, et al, 2009). These include gene promoter and enhancer regions affecting transcription levels or sequence elements affecting mRNA stability. DNA sequence variants could also affect local chromatin structure or mRNA splicing. Single nucleotide polymorphisms could affect protein stability and function. Most regulatory loci have local effects, but some (e.g., transcription factors) act at a distance and could affect multiple genes. Whatever their basis, the observed differences can help distinguish individuals and thus increase the value of hair as forensic evidence. As the most direct way to reveal whether the protein content permits distinguishing among individuals, the strategy in the first specific aim was to compare the profiles from different individuals by shotgun mass spectrometry. We anticipated that likely proteotypic peptides from proteins most useful in distinguishing among individuals could be selected from the data in this discovery phase of the work. A subsequent targeted phase of the work was planned to focus on the prototypic peptides for a simplified screening of samples. Thus prototypic peptides synthesized with heavy isotopes (to distinguish them from endogenous hair peptides) would be added to samples for quantitation of relative yields in different hair samples.

II. Methods

1. Hair samples

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Samples of scalp and body hair were collected from unrelated volunteer adult subjects with written consent approved by the University of California Davis Institutional Review Board (protocol 217868). Scalp hair from Caucasian (six male), African-American (five male), Korean (three male, two female) and Kenyan (four male, one female) individuals were analyzed. Three samples from each subject were analyzed; in most cases, scalp hair was collected at the same time and processed from three sites on the head (left, right, center). Minimizing contributions from other ethnicities, the parents of Korean and Kenyan subjects were also Korean or Kenyan, respectively. Samples of axillary, facial (beard) and pubic hair were collected from male subjects.

Using samples from such divergent backgrounds was expected to maximize the opportunity to observe differences in protein profile and to offer the chance to detect ancestry informative markers. Since Africa is the continent with the highest genetic diversity, individuals of African heritage are advantageous for this purpose. Surveys of African Americans exhibit an average West African genetic heritage of nearly 80% and average European genetic heritage of nearly 20%, with much variation around these means (Bryc, et al, 2010). Thus African Americans are likely to differ from East Africans in Kenya. The Korean population, primarily a Northeast Asian group, has inhabited the Korean peninsula since the Neolithic and Bronze ages (Kim, et al, 2011). Americans of European ancestry are distinct from all these divergent groups.

2. Sample processing

Samples (2-4 mg) were rinsed twice in 2% SDS – 0.1 M sodium phosphate (pH 7.8) and incubated in 0.4 ml of this buffer containing 25 mM DTE for disulfide reduction and then alkylation with iodoacetamide. Proteins were recovered as a flocculent precipitate by centrifugation after addition of 1 ml of ethanol, rinsed twice with 67% ethanol, once with freshly prepared 0.1 M ammonium bicarbonate, resuspended in 0.1 M ammonium bicarbonate - 10% acetonitrile and digested for 3 days with daily additions of 40 µg of stabilized trypsin (Rice, et al, 1977). Clarified digests, containing almost 90% of the digested protein, as determined by reaction of insoluble material with ninhydrin after sulfuric acid digestion (Lee, et al, 2006), were submitted for mass spectrometric analysis.

Subsequent to processing the hair samples from subjects of different ethnicity, the method was modified for use with hair samples available only in small amount to avoid the ethanol precipitation step. To this end, hair samples were reduced with DTE and alkylated with iodoacetamide in 2% sodium dodecanoate or ammonium perfluoro-octanoate in 0.05 M ammonium bicarbonate (instead of 2% SDS in 0.1 M sodium phosphate). This innovation is based on reports using sodium dodecanoate (Lin Y, 2013) and ammonium perfluoro-octanoate (Kadiyala, et al, 2010) in conjunction with mass spectrometric peptide analysis. The sample was then adjusted to 0.1% trifluoroacetic acid and extracted twice with ethyl acetate before readjusting the pH to 8 and digesting with trypsin. This procedure removes the detergent, preventing denaturation of the trypsin. Residual ethyl acetate after extraction (≈8%) functions in place of the previously added 10% acetonitrile as well to help solubilize hydrophobic peptides. As a demonstration that this innovation performs well in practice, hair samples from identical twins were processed using sodium dodecanoate instead of sodium dodecyl sulfate.

3. Mass spectrometry and protein identification

Samples adjusted to approximately equal peptide amounts by A^{280} and 0.5% trifluoracetic acid were directly loaded onto an Agilent ZORBAX 300SB C_{18} reverse-phase trap cartridge which, after loading, was switched in-line with a Michrom Magic C18 AQ 200 µm x 150 mm nano-LC

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

column connected to a Thermo-Finnigan LTQ iontrap mass spectrometer through a Michrom Advance Plug and Play nanospray source with CTC Pal autosampler. The nano-LC column was used with a binary solvent gradient; buffer A was composed of 0.1% formic acid and buffer B composed of 100% acetonitrile. The 120 min gradient consisted of the steps 2-35% buffer B for 85 min, 35-80% buffer B for 23 min, hold for one min, 80-2% buffer B for one min, then hold for 10 min, at a flow rate of 2 µl/min for maximal separation of tryptic peptides. An MS survey scan was obtained for the m/z range 375-1400, and MS/MS spectra were acquired from the 10 most intense ions in the MS scan by subjecting them to automated low energy CID. An isolation mass window of 2 Da was used for the precursor ion selection, and normalized collision energy of 35% was used for the fragmentation. A 2 min duration was used for the dynamic exclusion. Monitoring of column washes indicated negligible intersample contamination.

Tandem mass spectra were extracted with Xcalibur version 2.0.7. All MS/MS samples were analyzed using X!Tandem (The GPM, thegpm.org; version CYCLONE (2012.10.01.2)). X!Tandem was set up to search a 2012 Uniprot human database appended to a database of common non-human contaminants (cRAP, http://www.thegpm.org/crap/), both of which were appended to an identical but reversed database for calculating false discovery rates (136,252 proteins total), assuming the digestion enzyme was trypsin. X!Tandem was searched with a fragment ion mass tolerance of 0.40 Da and a parent ion tolerance of 1.8 Da. Iodoacetamide derivative of cysteine was specified in X!Tandem as a fixed modification. Deamidation of asparagine and glutamine, oxidation of methionine and tryptophan, sulfone of methionine and tryptophan oxidation to formylkynurenin of tryptophan were specified in X!Tandem as variable modifications. Scaffold version_4.2.1 (Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 90% probability as specified by the Peptide Prophet algorithm (peptide decoy false discovery rate 0.2%) (Keller, et al, 2002). Protein identifications were accepted if they could be established at greater than 99% probability and contained at least two identified peptides (protein decoy false discovery rate 4.1%). Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii, et al, 2003). Numbers of distributed spectral counts (called weighed spectral counts in Scaffold) were tabulated using experiment-wide grouping. The Scaffold file containing all the peptide data used in the analysis is now available in the public database on the MassIVE repository (ID: MSV000078650; http://massive.ucsd.edu).

Because certain keratins are well known to contain identical peptides, we used a distributed spectral count (called weighted spectral count in Scaffold) to model spectral counts more accurately across the proteins identified in this study. Scaffold's weighted spectral counts appropriate a percentage of each count divided among the protein groups that share that peptide. The formulation for that percentage for peptide(i) assigned to protein(j) is PPS(j)/sum(PPS(1...n)), where PPS(j) is the sum of the peptides(1...m) for protein(j) and (1...n) is the set of proteins that contain peptide(i). The weighted spectral count is the sum of those percentages for each protein group. This is similar to method 3-a in Zhang et al (2010), but Scaffold uses the summed probabilities as normalizers instead of summed exclusive counts. In addition, every protein identified was required to have at least one unique peptide that was not shared with any other protein.

4. Statistical analysis

Hair samples from different subjects were compared using mixed-effects overdispersed Poisson regression models, including a fixed effect for ethnicity (to adjust for imbalances between total hair donors) and a random effect for subject. This analysis included scalp total hair samples

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

from six Caucasian subjects, five Kenyan subjects, five African American subjects, and five Korean subjects. Hair from different sites (scalp, axillary, facial, and pubic) was likewise compared within subjects using mixed-effects overdispersed Poisson regression models, including a fixed effect for site and a random effect for subject. Pairwise comparisons between sites were conducted using the Tukey HSD method. This analysis included scalp, axillary, facial, and pubic hair samples from four Caucasian subjects. Scalp samples were compared among subjects of the same ethnicity using an overdispersed Poisson model, treating subject as a fixed effect. The comparisons were conducted for each protein that was present in sufficient amount. When the comparison showed a significant difference ($p<0.05$) globally for a given protein among the samples being compared, pairwise comparisons between subjects or sites were then conducted using the Tukey HSD method. Error bars in figures represent 95% confidence intervals for parameter estimates from the models described above. Hierarchical clustering was performed with the hclust function in the R statistical software environment using the complete linkage method as described in http://nlp.stanford.edu/IR-book/html/htmledition/single-link-and-complete-link-clustering-1.html (Manning, et al, 2008).

III. Results

A. Discovery Phase - Analyze hair samples from adults of different ancestries by shotgun proteomics. The advantage of the shotgun approach is that >100 proteins can be examined from a single sample. The analysis was anticipated to identify proteins that were the most discriminating among individuals. This part of the Results section and the corresponding methods are excerpted directly from our open access publication “Human hair shaft proteomic profiling: individual differences, site specificity and cuticle analysis” by Laatsch CN, Durbin-Johnson BP, Rocke DM, Mukwana S, Newland AB, Flagler MJ, Davis MG, Eigenheer RA, Phinney BS, Rice RH (2014) PeerJ 2:e506 (see Dissemination of Research Findings).

1. Distinguishing among hair shaft samples within ethnic groups

Scalp hair samples from Caucasian, African-American, Korean and Kenyan individuals were analyzed. Analysis of triplicate samples from each subject (one each from different regions of the scalp) provided assurance that the differences among samples reflected individual subject differences. Considerable variation was evident in the profiles of prevalent proteins within each ethnic group. Using pairwise comparisons of 76 proteins present in sufficient amounts for statistical analysis, each group was analyzed separately. As shown in Table 1 below, samples from individuals were distinguishable in each group, although the number of significant protein differences varied greatly. Figure 1 below displays for the Caucasian group nine proteins that were present at significantly different levels in comparisons among at least some of the 6 individuals analyzed. In such analyses, K32, K33B, K34 and K83 exhibited significant differences in at least 40% of the pairwise comparisons within all four ethnic groups, while K31, K35, K39 and K86 exhibited similarly high inter-subject variability in at least three of the ethnic groups.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Caucasian	CA2	CA3	CA4	CA5	CA6
CA1	7	5	11	15	26
CA2		4	16	13	27
CA3			14	23	23
CA4				23	34
CA5					30

African-American	AA2	AA3	AA4	AA5
AA1	26	22	24	18
AA2		18	20	20
AA3			5	6
AA4				9

Kenyan	KE2	KE3	KE4	KE5
KE1	6	14	13	16
KE2		10	5	21
KE3			22	22
KE4				18

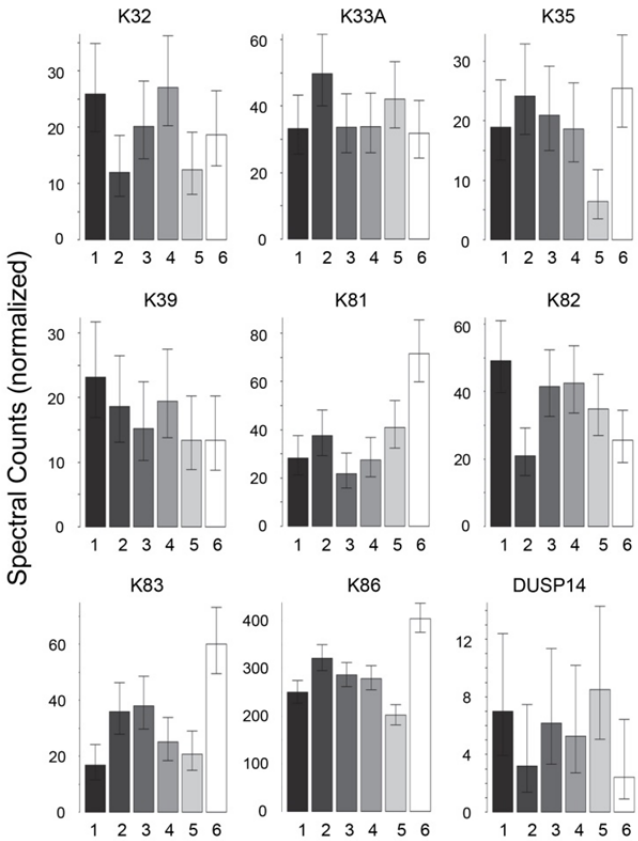
Korean	KO2	KO3	KO4	KO5
KO1	2	3	37	39
KO2		4	40	40
KO3			28	32
KO4				21

Table 1. Pairwise comparisons of 76 proteins in profiles from samples in each of 4 ethnic groups. Values are numbers of proteins that were significantly different in weighted spectral counts. Subjects in each group - 6 Caucasian and 5 each African-American, Kenyan and Korean - are numbered and labeled CA, AA, KE and KO, respectively.

2. Distinguishing samples from different ethnic groups

In contrast to the high variation in protein expression within ethnic groups, consistent differences among ethnic groups were less marked. As shown in Table 2 below, the groups were distinguishable, with African-American samples being the most distinctive. Most useful for pairwise comparisons were KAPs, accounting for 21 of the 32 significant differences (66%) observed overall. Illustrated in Figure 2, KAPs 2-4, 4-3, 13-1 and 13-2 accounted for 16 (50%), whereas K40, selenium binding protein-1 (SBP1) and epidermal transglutaminase (TGM3) accounted for seven (22%) of the total significant differences.

Figure 1. Weighted spectral counts for 9 proteins differentially expressed in hair shafts from 6 Caucasian subjects. Samples were analyzed separately from 3 locations on the scalp.



	CA	KE	KO
AA	8	6	7
CA		4	3
KE			4

Table 2. Pairwise comparisons of proteins in hair from subjects from different ethnic groups. Values are significant differences in weighted spectral counts from hair from African-American (AA), Caucasian (CA), Kenyan (KE) and Korean (KO) subjects.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

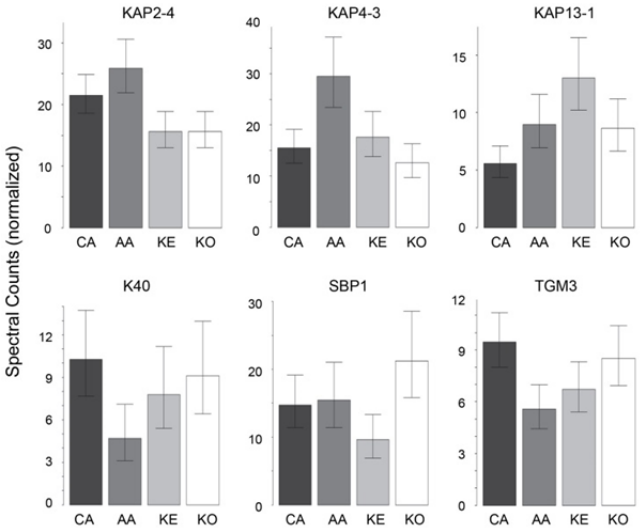


Figure 2. Weighted spectral counts for 6 proteins differentially expressed in hair from subjects of different ethnic origin. Each bar represents the aggregate for Caucasian (CA), African-American (AA), Kenyan (KE) and Korean (KO) samples.

3. Distinguishing hair shafts from different body sites

Hair samples from three body sites (axillary, facial and public regions) were analyzed and compared to those from scalp. As seen in Figure 3 below, pairwise comparisons using 92 proteins permitted distinguishing among them. Scalp hair displayed the most differences from the others, while axillary and pubic hair displayed the fewest differences from each other. In these comparisons, KRTs and KAPs together and in equal amounts accounted for only one-third of the significant differences. A variety of enzymes and structural proteins contributed to the observed differences as illustrated in Figure 4 below.

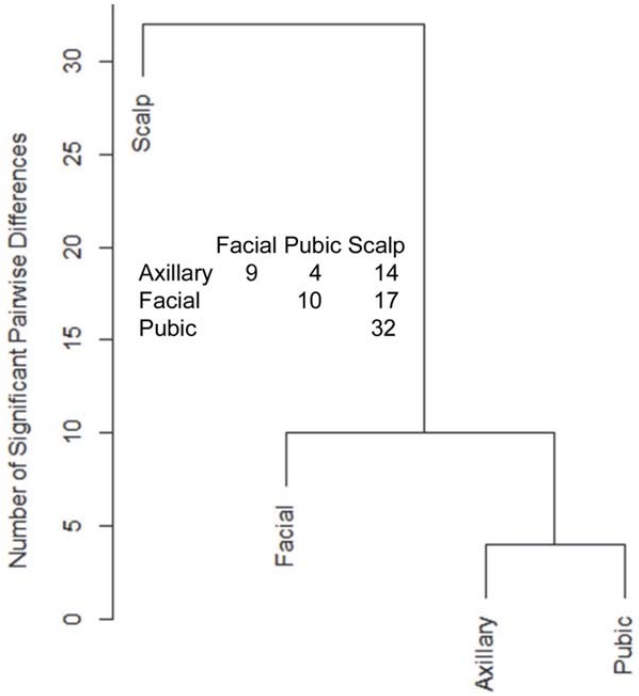


Figure 3. Pairwise comparisons of proteins (92 analyzed) in hair obtained from different body sites of the same four individuals. The values shown for significantly different weighted spectral counts were used for hierarchical clustering by relatedness.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

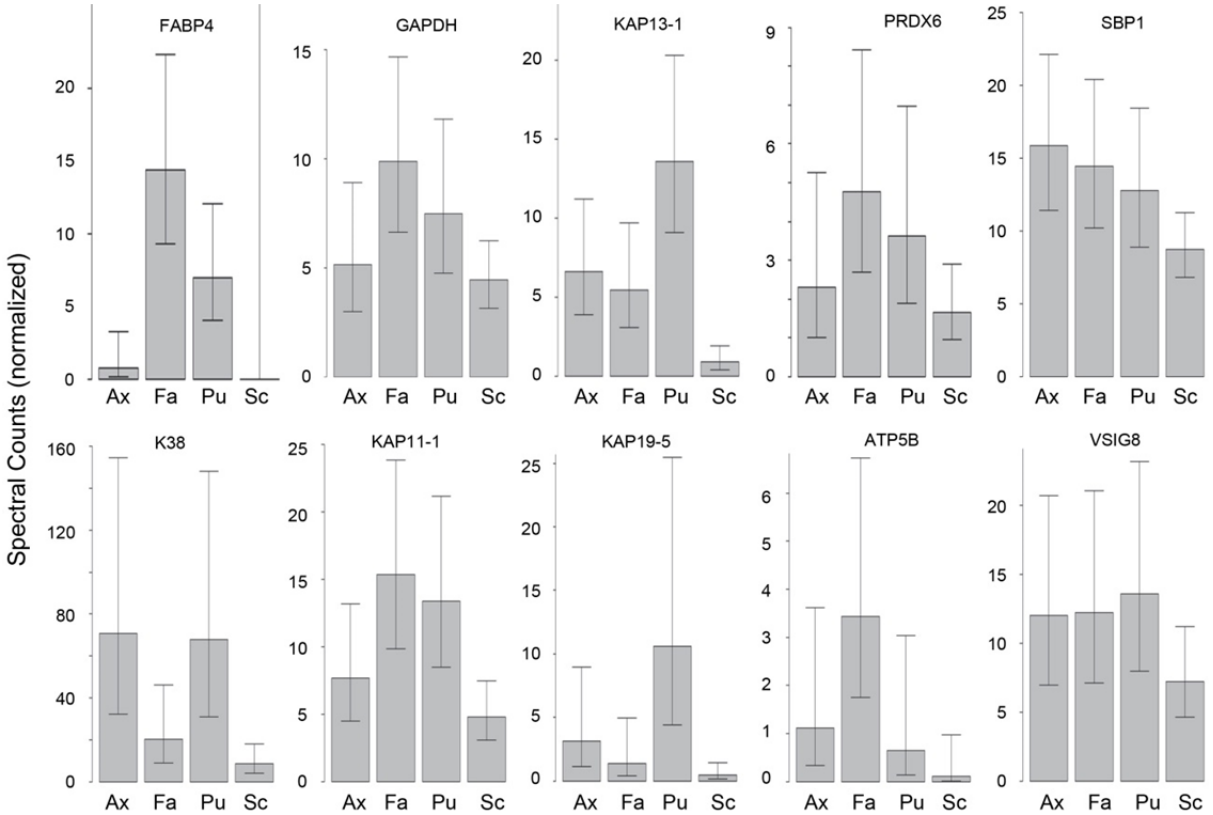


Figure 4. Weighted spectral counts for 10 proteins differentially expressed in hair from different body sites of Caucasian subjects. Illustrated are composite values for single samples of Ax (axillary), Fa (facial), and Pu (pubic) hair from each of 4 Caucasian subjects and for triplicate values of Sc (scalp) hair from the same four subjects.

B. Targeted Phase - Reanalyze the samples using a small set of sentinel proteins that are the most diagnostic for differences among individuals. The original grant proposal envisioned selecting sentinel proteins and reanalyzing the hair digests for the proteotypic peptides in them. Since these peptides are specific and reproducibly detected in digests in high yield, this approach would considerably simplify the analysis and provide higher sensitivity.

Inspection of results from the shotgun analysis in the first phase did reveal potential sentinel proteins that were highly variable among individuals, suggesting that proteotypic peptides could be identified. Before embarking on the expensive task of obtaining proteotypic peptides labeled with heavy isotopes, we proceeded to verify that the proposed multiple reaction monitoring would be feasible in parallel experiments in another project using samples of human nail. Our original experiments giving acceptable yields were obtained with the insoluble fraction of nail plate after extensive detergent extraction (Rice, et al, 2010). However, subsequent work indicated that much better discrimination among individual mouse strains was obtained with total hair rather than with the insoluble material after extraction (Rice, et al, 2012). We realized that the anticipated approach, involving removal of the large majority of the keratin fraction, would have removed many of the proteotypic peptides. When the total extract of nail was employed to demonstrate feasibility of using this fraction for multiple reaction monitoring, the analysis was unexpectedly impeded by interference from the high content of keratin peptides with proper separation of peptides in general. Numerous attempts to circumvent this technical problem over the course of a year by reducing amounts injected in the assay, altering conditions of chromatography and repeated intensive column cleanup were unsuccessful. The quantitation obtained was inconsistent, and the lack of reproducibility prevented subsequent analysis. Since

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

nail plate and hair shaft are equivalent in keratin content, proceeding with the planned analysis of hair shaft was clearly futile. The interference by keratin in such analysis is well-known.

One way to avoid this problem would be to decrease the content of irrelevant peptides. Simplification of the digest by removal of unwanted keratin peptides would be expected to permit quantitative analysis as originally envisioned. In general, such a goal is not feasible by routine biochemical methods. However, a feasible method could be to isolate the proteotypic peptides of interest with specific antibodies. This approach has been demonstrated to be successful with human serum proteins (Krastins, et al, 2013). However, it involves derivation of antibodies against proteotypic peptides, an expensive and laborious process beyond the remaining time frame (probably two years) and budget of this grant. Since several other aspects of using proteomic hair analysis in forensic investigation remained to be evaluated before this approach can be adopted, we elected to pursue two of considerable importance, envisioned in the original proposal for future work, that were within the time frame and budget of this grant.

C. Dependence of human hair profile on individual genetics

In the course of collecting hair for analysis, we obtained samples from a donor and a twin sibling. The two individuals were originally believed to be fraternal, but their nearly identical appearances led them to obtain genetic testing commercially from 23 and Me (https://www.23andme.com/ancestry/), and to conclude that they are indeed identical twins. However, their hair profiles were readily distinguishable. This unexpected finding led to extending our analysis to hair samples from identical twins. Samples were obtained from the Twins Research Registry maintained by SRI International Biosciences Division, which focuses on twins in the San Francisco Bay Area. Hair samples from 9 sets of identical twins were received through SRI, processed and analyzed. Preliminary statistical analysis indicated the protein profiles within each pair were quite similar despite the clearly distinguishable profiles comparing the pairs to each other. Subsequent more stringent analysis permitted the quantitation of differences illustrated in the Table 3. The data showed that twins can be distinguished, although not as readily as unrelated individuals. In addition, these results permitted us to test the hypothesis that the hair profiles would diverge with age if differences were due to epigenetic changes occurring by a random process. A standard regression analysis showed a 0 slope of changes with age and a correlation coefficient of 0. Although changes in profile could occur with age, we infer that such changes would not be random but occur largely in the same direction within the twin pairs.

Age	Protein Differences
19	2
19	3
20	2
25	15
28	5
59	9
66	1
73	5
76	2
84	8

Table 3. Protein differences for twin pairs sorted by age. Shown are results of pairwise comparisons within twin pairs giving a mean of 5.2 and a range of 1-15 significant differences.

The protein profile differences within twin pairs were much smaller than the differences among unrelated individuals. For example, fifty random pairings of unrelated individuals within this twin population gave an average of 52 differentially expressed proteins compared to 3 found in actual twin pairings. In addition to collecting the subjects' ages, we also surveyed their geographic locations, use of medications and alcohol, health status, grooming (hair care), diet, sleep habits, exercise, work environment, stress and hair length. None of these factors were correlated with the number of significant protein differences. The data are consistent with constraint by the donor's genetics with at most a weak influence of environmental conditions.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

However, inasmuch as these twins were raised together, an effect of early environment would not be evident from the survey.

D. Optimization of hair processing for small sample sizes

The high sensitivity of mass spectrometry can be exploited with improved sample processing. Since a typical shotgun analysis requires only 1 µg of protein, the 4 mg samples exceed the minimal sample amount by at least 1000 fold. Thus, a single hair strand 1 cm in length (30-50 µg) should easily suffice for analysis. The large sample sizes routinely employed to date have the advantage that they are much less subject to contamination by ambient protein. This is a serious consideration even with scrupulous use of gloves, since a large fraction of household dust is exfoliated epidermis and hair. However, in principle, contamination presents a surmountable problem following established principles in laboratories dedicated to forensic DNA analysis using PCR.

1. Analysis of supernatant from ethanol precipitation

A critical consideration in processing small hair samples is the yield. Probably the most problematic step in the current protocol is precipitation by ethanol, required to remove SDS from the sample so it will not interfere with subsequent mass spectrometric analysis. Flocculent precipitation with 4 mg of hair is unlikely to occur with 40 µg, however. Further impetus to this initiative came with the observation that the yield of protein from precipitation even with 4 mg of hair is not 100%. Although difficult to measure accurately, our best estimate is that several percent of the total protein remain in the clear ethanol supernatant after centrifugation.

We investigated whether the protein remaining in the ethanol supernatant constituted a specific fraction that might have been overlooked in the analysis due to its lack of precipitation. Such a fraction could aid in distinguishing samples. Evaporation of ethanol from the supernatant permitted visualization of the concentrated remaining protein by gel electrophoresis from aliquots of both human and mouse hair samples (Figure 5). The protein bands with greatest staining migrated in the vicinity of 35-50 kDa as might be expected of keratins, while a small amount of material (possibly protein) migrated near the dye front. For further analysis, the protein was freed of small molecules by gel filtration using Sephadex G-25 maintained in 0.05% SDS. The void volume material was dialyzed against several changes of 67% ethanol to remove residual SDS (Rice, 1974) and then 0.05 M ammonium bicarbonate before digestion and mass spectrometric analysis. The protein profile was essentially the same as that obtained from the rest of the protein precipitated by ethanol. We conclude that our standard procedure is satisfactory but could be improved for small sample amounts.

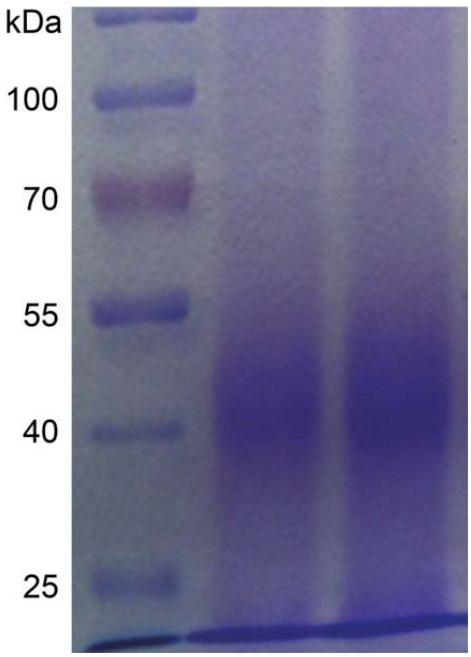


Figure 5. Electrophoretic mobility of protein in the supernatant after precipitation of mouse hair protein with ethanol. The majority of protein had mobility corresponding to 35-50 kDa. Duplicate samples are shown. Results with human hair were equivalent.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

2. Replacement of SDS by detergent removable by organic solvent extraction

Our initial efforts to find a replacement for SDS focused on two commercially available mass spectrometry-compatible detergents (ProteaseMax and RapiGest) and sodium deoxycholate. In the presence of DTE, SDS addition leads to visible swelling of the hair shaft after two hr at room temperature (Rice, 2011), evidence that it can disrupt the hair shaft structure. The three detergents above did not induce obvious hair shaft swelling even with longer treatment at higher temperatures. In parallel experiments, digestion of hair shafts reduced and alkylated in ProteaseMax or RapiGest gave <10% the yield of hair treated with SDS, while deoxycholate-treated hair shafts had roughly one-third the yield.

We reasoned that a detergent with a long alkyl chain such as dodecyl provides a hydrophobic moiety sufficient to disrupt the hair shaft, but the other detergents above that were tried are not sufficiently hydrophobic. Replacing the sulfate of SDS with carboxylate (i.e., dodecanoate) would retain the hydrophobicity but permit the detergent to be removed by organic solvent extraction at low pH. Another detergent with similar hydrophobicity is perfluoro-octanoate, which has the added advantage that it is volatile when acidified. Both have been used for preparation of peptide samples for mass spectrometry (Kadiyala, et al, 2010; Lin Y, 2013). The organic solvent used in the extraction, ethyl acetate, is partially soluble in water (estimated 8% by volume), which is compatible with trypsin digestion after adjusting the pH to 8.

Since our first studies of hair (Lee, et al, 2006), the reduced and alkylated hair shafts have been digested with stabilized trypsin (1% by mass, added daily) for 3 days at room temperature to insure maximal digestion. The extent of digestion is likely augmented over that time period by the low level of inherent chymotryptic activity present in trypsin and reflected in some 20% of the peptides being half-tryptic (Lee, et al, 2006). As a measure of the efficiency of hair shaft digestion, the amount of hair present in the insoluble residue after trypsinization was measured. With our standard procedure, the residual protein was 13 $\pm$ 4% of the initial amount. In contrast to deoxycholate or the two commercially available detergents compatible with mass spectrometry, which permitted considerably lower degrees of digestion, sodium dodecanoate and ammonium perfluoro-octanoate each gave ≈15% residual protein (Figure 6 below). Mass spectrometric analyses of the solubilized peptides in the digest were essentially the same as with SDS. On this basis, the hair samples from identical twins were processed using sodium dodecanoate instead of sodium dodecyl sulfate to demonstrate the utility of this procedural modification (details in Methods section).

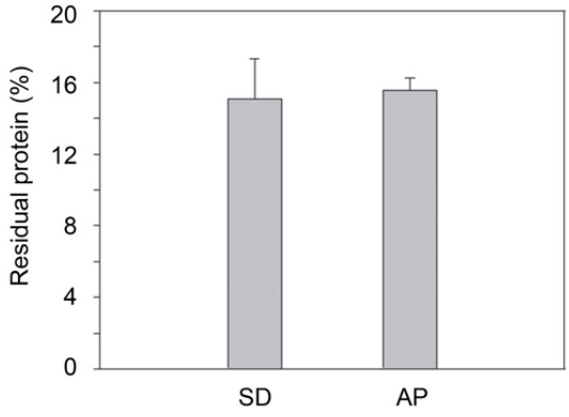


Figure 6. Residual protein (not solubilized) after trypsinization of human hair. The hair was reduced and alkylated in 2% sodium dodecanoate (SD) or ammonium perfluoro-octanoate (AP) instead of sodium dodecyl sulfate.

IV. Conclusions

1. Discussion of findings

● Present experiments verify the potential utility of hair profiling. Individual differences in hair shaft proteomic profile are clearly evident, confirming predictions based on analysis of hair from

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

inbred mouse strains. Based on pairwise comparisons of individual proteins, the differences are substantial. The results so far suggest this method may have discrimination capability similar (but complementary) to that of mitochondrial DNA.

● Differences between hair samples from different ethnic origin are detectable, but the degrees of difference are smaller than among individuals. Among various protein classes, the keratin associated proteins appear most useful in such discrimination and merit further examination.

● Protein profiles of hair from different body sites are distinguishable. A worthwhile goal for the future, entailing finding the variance of profiles from various sites in an individual, would be to determine whether hair from the scalp and from other body sites of a single individual can be predicted from the profile of one of them.

● Processing of small hair samples is now possible by modification of our standard treatment procedure. This observation raises the likelihood that small samples will have sufficient material to be divided for both protein profiling and mtDNA analysis.

As genetic analysis has proceeded in diverse populations using advanced technology (e.g., SNP chips), it has become possible to trace lineages from the first migrations leaving Africa as modern humans spread across the world (Li, et al, 2008). Since ancestry informative markers are evident at the genetic level, distinctive protein profiles may also exist. However, we envisioned that if individual variation is much larger than inter-ethnic variation, as some work has suggested (Storey, et al, 2007) and is now manifest in the present work, ancestry informative markers may not be numerous. The relative paucity of proteins distinguishing hair from distinct ethnic origins in the present work indicates that further effort will be required to learn whether the differences have practical utility. The major hair shaft components, KRTs and KAPs, were both useful in discriminating among individuals, but the latter appeared to be more useful among ethnic groups. A previous study also pointed to differences in KAPs among subjects of African, Jamaican and African-American origin (Porter, et al, 2009). These findings, and the frequent length polymorphisms they display (Fujikawa, et al, 2012), often reflecting amino acid repeats in the coding region, may impel further focus on KAPs for this purpose. The results also indicate that some visible differences (e.g., curliness) are not bestowed by major structural proteins. Moreover, the importance of lipid processing for features such as combability is now appreciated (Shimomura, et al, 2009).

● As mass spectrometry instrumentation continues to increase in sensitivity, more proteins will be amenable to inclusion in the analyses, increasing the power of comparisons among samples from individuals. For example, the newly introduced Thermo Q-Exactive instrument has several fold the sensitivity of the LTQ instrument used in most of the present work and would likely improve the resolution among samples.

● The results of our twin study support the hypothesis that hair protein profiling could be useful in suspect identification. The profiles appear to be under strong genetic control, and environmental factors in adulthood appear weak influences at most. Whether a profile is stable over a lifetime or changes with age remains to be established, but at least changes appear to be the same in twins. This factor could be investigated in human individuals or, more easily, in inbred strains of mice. The possibility also remains that the time at which an embryo splits in gestation to yield identical twins could influence the distinguishability of the hair protein profiles. This would be analogous to development of fingerprints, which occurs well after splitting into twins occurs, but whether a process exists to fix hair protein profiles and, if so, when it takes place is unknown.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

2. Implications for policy and practice

The present results establish that differences in hair profiles exist in the human population, and these differences are substantial enough to distinguish among individuals. However, we are not yet in a position to employ proteomic analysis in practice. For example, the dependence of an individual's hair protein profile on age needs further study. In addition, to put such an approach into practice would involve establishing a database and developing a statistical basis for searching it. This work provides a foundation for such efforts.

3. Implications for further research

A strong genetic component appears to be involved in gene expression differences observed in many species. This likely reflects variation in transcription factor binding affinity for chromatin that can even have epigenetic consequences (Kasowski, et al, 2013; Kilpinen, et al, 2013; McVicker, et al, 2013). However, indications that the hair protein profile in humans could depend on age (Giesen, et al, 2011) and the possible influence of other factors including environment and diet (Almeida, et al, 2014) cannot be completely discounted. In addition, how well cuticle cells remain attached to the shaft could influence the profile (Laatsch et al, 2014). The latter could depend on hair length, weathering, and cuticle cell stability, variables among inbred mouse strains (Rice, et al, 1999a) and thus likely among individual humans. These factors would need clarification in determining the value a database of individual hair profiles would have for forensic science.

Identifying proteins that are differentially expressed between hair samples from different individuals and among different ethnic groups is an important first step in establishing whether development of a classification system would be useful. Whether we can classify hair into ethnic groups in a rigorous fashion yielding p-values remains to be seen. Beyond ethnic classification, if we can demonstrate the value of hair profile comparisons, establishing a searchable database would be possible. However, correctly matching a given hair sample with an entry in a database is a much more difficult problem than identifying differentially expressed individual proteins. Useful scenarios include one in which an unidentified hair is compared against a database of multiple samples each from multiple individuals, and one in which an unidentified hair from a crime scene is compared against multiple samples from a known individual. Although the discrimination power of hair profiles remains to be established, we speculate it will be similar but complementary to that of mtDNA analysis. Since the scale of sampling from individuals depends on the observed variance, the data we have obtained will help guide statistical developments.

V. References

Almeida AM, Plowman JE, Harland DP, Thomas A, Kilminster T, Scanlon T, Milton J, Greeff J, Oldham C, Clerens S (2014) Influence of feed restriction on the wool proteome: A combined iTRAQ and fiber structural study. J Proteomics 103:170-177

Bryc K, Auton A, Nelson MR, Oksenberg JR, Hauser SL, Williams S, Froment A, Bodo J-M, Wambebe C, Tishkoff SA, Bustamante CD (2010) Genome-wide patterns of population structure and admixture in West Africans and African Americans. Proc Natl Acad Sci USA 107:786-791

Fujikawa H, Fujimoto A, Farooq M, Ito M, Shimomura Y (2012) Characterization of the human hair keratin-associated protein 2 (KRTAP2) gene family. J Invest Dermatol 132:1806-1813

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Gaudette BD (1999) Evidential value of hair examination. In: Forensic Examination of Hair (Robertson J, ed), Taylor & Francis, London, 243-260

Giesen M, Gruedl S, Holtkoetter O, Fuhrmann G, Koerner A, Petersohn D (2011) Ageing processes influence keratin and KAP expression in human hair follicles. Exp Dermatol 20:759-761

Kadiyala CS, Tomechko SE, Miyagi M (2010) Perfluorooctanoic acid for shotgun proteomics. PLoS One 5(12):e15332

Kasowski M, Kyriazopoulou-Panagiotopoulou S, Grubert F, Zaugg JB, Kundaje A, Liu Y, Boyle AP, Zhang QC, Zakharia F, Spacek DV, Li J, Xie D, Olarerin-George A, Steinmetz LM, Hogenesch JB, Kellis M, Batzoglou S, Snyder M (2013) Extensive variation in chromatin states across humans. Science 342:750-752

Keller A, Nesvizhskii AI, Kolker E, Aebersold R (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Analyt Chem 74:5383-5392

Kilpinen H, Waszak SM, Gschwind AR, Raghav SK, Witwicki RM, Orioli A, Migliavacca E, Wiederkehr M, Gutierrez-Arcelus M, Panousis NI, Yurovsky A, Lappalainen T, Romano-Palumbo L, Planchon A, Bielser D, Bryois J, Padioleau I, Udin G, Thurnheer S, Hacker D, Core LJ, Lis JT, Hernandez N, Reymond A, Deplancke B, Dermitzakis ET (2013) Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. Science 342:744-747

Kim SH, Kim KC, Shin DJ, Jin HJ, Kwak KD, Han MS, Song JM, Kim W, Kim W (2011) High frequencies of Y-chromosome haplogroup O2b-SRY465 lineages in Korea: a genetic perspective on the peopling of Korea. Invest Genet 2:in press

Krastins B, Prakash A, Sarracino DA, Nedelkov D, Niederkofler EE, Kiernan UA, Nelson R, Vogelsang MS, Vadali G, Garces A, Sutton JN, Peterman S, Byram G, Darbouret B, Pérusse JR, Seidah NG, Coulombe B, Gobom J, Portelius E, Pannee J, Blennow K, Kulasingam V, Couchman L, Moniz C, Lopez MF (2013) Rapid development of sensitive, high-throughput, quantitative and highly selective mass spectrometric targeted immunoassays for clinically important proteins in human plasma and serum. Clin Biochem 46:399-410

Lappalainen T, Dermitzakis ET (2010) Evolutionary history of regulatory variation in human populations. Hum Molec Genet 19:R197-R203

Lee YJ, Rice RH, Lee YM (2006) Proteome analysis of human hair shaft: From protein identification to posttranslational modification. Molec Cell Proteom 5:789-800

Li JZ, Absher DM, Tang H, Southwick A, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL, Myers RM (2008) Worldwide human relationships inferred from genome-wide patterns of variation. Science 319:1100-1104

Lin Y HL, Liu Z, Li J, Liu Y, He Q, Wang X, Liang S (2013) Sodium laurate, a novel protease- and mass spectrometry-compatible detergent for mass spectrometry-based membrane proteomics. PLoS One 8(3):e59779

Manning CD, Raghavan P, Schütze H (2008) Introduction to Information Retrieval. In: ed), Cambridge University Press,

McVicker G, van de Geijn B, Degner JF, Cain CE, Banovich NE, Raj A, Lewellen N, Myrthil M, Gilad Y, Pritchard JK (2013) Identification of genetic variants that affect histone modifications in human cells. Science 342:747-749

National Research Council (2009) Strengthening Forensic Science in the United States: A Path Forward. National Academies Press, Washington, DC, pp 157-159

Nesvizhskii AI, Keller A, Kolker E, Aebersold R (2003) A statistical model for identifying proteins by tandem mass spectrometry. Analyt Chem 75:4646-4658

Petkov PM, Ding Y, Cassell MA, Zhang W, Wagner W, Sargent EE, Asquith S, Crew V, Johnson KA, Robinson P, Scott VE, Wiles MV (2004) An efficient SNP system for mouse genome scanning and elucidating strain relationships. Genome Res 14:1806-1811

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Porter CE, Dixon F, Khine CC, Pistorio B, Bryant H, de la Mettrie R (2009) The behavior of hair from different countries. J Cosmet Sci 60:97-109

Rice RH (1974) Minor protein components in cowpea chlorotic mottle virus and satellite tobacco necrosis virus. Virology 61:249-255

Rice RH (2011) Proteomic analysis of hair shaft and nail plate. Int J Cosmetic Sci 62:229-236

Rice RH, Bradshaw KM, Durbin-Johnson BP, Rocke DM, Eigenheer RA, Phinney BS, Sundberg JP (2012) Differentiating inbred mouse strains from each other and those with single gene mutations using hair proteomics. PLoS One 7:e51956

Rice RH, Crumrine D, Hohl D, Munro CS, Elias PM (2003) Cross-linked envelopes in nail plate in lamellar ichthyosis. Br J Dermatol 149:1050-1054

Rice RH, Crumrine D, Uchida Y, Gruber R, Elias PM (2005) Structural changes in epidermal scale and appendages as indicators of defective TGM1 activity. Arch Dermatol Res 297:127-133

Rice RH, Green H (1977) The cornified envelope of terminally differentiated human epidermal keratinocytes consists of cross-linked protein. Cell 11:417-422

Rice RH, Means GE, Brown WD (1977) Stabilization of bovine trypsin by reductive methylation. Biochim Biophys Acta 492:316-321

Rice RH, Rocke DM, Tsai H-S, Lee YJ, Silva KA, Sundberg JP (2009) Distinguishing mouse strains by proteomic analysis of pelage hair. J Invest Dermatol 129:2120-2125

Rice RH, Wong VJ, Pinkerton KE (1994) Ultrastructural visualization of cross-linked protein features in epidermal appendages. J Cell Sci 107:1985-1992

Rice RH, Wong VJ, Price VH, Hohl D, Pinkerton KE (1996) Cuticle cell defects in lamellar ichthyosis hair and anomalous hair shaft syndromes visualized after detergent extraction. Anatomic Rec 246:433-440

Rice RH, Xia Y, Alvarado RJ, Phinney BS (2010) Proteomic analysis of human nail plate. J Proteome Res 9:6752-6758

Robertson J (1999) Forensic and microscopic examination of human hair. In: Forensic Examination of Hair (Robertson J, ed), Taylor & Francis, London, 79-154

Robertson KS, McNevin D, Robertson J (2007) STR genotyping of exogenous hair shaft DNA. Australian J Forensic Sci 39:107-122

Rowe WF (2001) The current status of microscopical hair comparisons. Sci World J 1:868-878

Shimomura Y, Wajid M, Zlotogorskic A, Lee YJ, Rice RH, Christiano AM (2009) Founder mutations in the lipase H (LIPH) gene in families with autosomal recessive woolly hair/hypotrichosis. J Invest Dermatol 129:1927-1934

Skelly DA, Ronald J, Akey JM (2009) Inherited variation in gene expression. Ann Rev Genomics Hum Genet 10:313-332

Storey JD, Madeoy J, Strout JL, Wurfel M, Ronald J, Akey JM (2007) Gene-expression variation within and among human populations. Am J Hum Genet 80:502-509

Szabo S, Jaeger K, Fischer H, Tschachler E, Parson W, Eckhart L (2012) In situ labeling of DNA reveals interindividual variation in nuclear DNA breakdown in hair and may be useful to predict success of forensic genotyping of hair. Int J Legal Med 126:63-70

Taupin JM (2004) Forensic hair morphology comparison - a dying art or junk science? Science & Justice 44:95-100

Wu B, Potter CS, Silva KA, Liang Y, Reinholdt LG, Alley LM, Rowe LB, Roopenian DC, Awgulewitsch A, Sundberg JP (2010) Mutations in sterol O-acyltransferase 1 (*Soat1*) result in hair interior defects in AKR/J mice. J Invest Dermatol 130:2666-2668

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

VI. Dissemination of Research Findings

1. Publications

Laatsch CN, Durbin-Johnson BP, Rocke DM, Mukwana S, Newland AB, Flagler MJ, Davis MG, Eigenheer RA, Phinney BS, Rice RH (2014) Human hair shaft proteomic profiling: individual differences, site specificity and cuticle analysis. PeerJ 2:e506; DOI 10.7717/peerj.506 (open access)

Laatsch CN (2013) The Analysis of Human Hair Proteomics for Improved Evidence Discrimination. Thesis submitted in partial fulfillment of the requirements for the M.S. degree in Forensic Science from the University of California, Davis.

Wu P-W (2014) Are Monozygotic Twins Distinguishable By Hair Protein Profiles? Thesis submitted in partial fulfillment of the requirements for the M.S. degree in Forensic Science from the University of California, Davis.

2. Presentations

Rice RH (6/19/2012) Proteomic analysis of hair: Improved evidence discrimination. NIJ Conference 2012, Arlington, VA (poster giving background and rationale)

Laatsch C, Wu P-W, Rice RH (2/18/2014) Hair shaft proteomic analysis - Improved evidence discrimination? NIJ R&D Grantees Meeting during the 2014 AAFS Meeting in Seattle, WA (platform talk by P-W Wu); webinars of the talk repeated March 25, 26, 27 (30 min each)
Archival registration website:
https://www.forensiced.org/training/viewcourse.cfm?moduleid=930084CA-82F3-42B9-9232-967C0494707C

• Wu P-W, Rice RH (10/23/2014) A Pilot Study of Hair Protein Profiles in Monozygotic Twins-Improved Evidence Discrimination? Joint CAC & NWAFS Meeting in Rohnert Park, CA.

• Wu P-W, Rice RH (12/5/2014) A Pilot Study of Hair Protein Profiles in Monozygotic Twins-Improved Evidence Discrimination? An update of the twins study results presented at the CAC DNA Study Group in Richmond, CA.

3. Planned

Wu P-W, Rice RH - Results of experiments on hair from identical twins:

• The results comparing profiles among identical twins are anticipated to be submitted soon for publication in the peer reviewed scientific literature.

Our findings have stimulated a collaborative effort with a laboratory affiliated with the Forensic Science Center at the Lawrence Livermore National Laboratory nearby in California to improve the analysis of genetic polymorphisms in hair protein revealed through peptide analysis. A grant proposal was submitted (unsuccessfully) to the NIJ for this purpose and is in preparation for resubmission.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.