Forensic Science International: Genetics 47 (2020) 102314

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen

Research paper

Optimal processing for proteomic genotyping of single human hairs

Zachary C. Goecker[a], Michelle R. Salemi[b], Noreen Karim[a], Brett S. Phinney[b], Robert H. Rice[a], Glendon J. Parker[a,*]

[a] Department of Environmental Toxicology, University of California Davis, One Shields Ave., Davis, CA 95616, USA
[b] Proteomics Core Facility, University of California, Davis, CA, USA

ARTICLE INFO

Keywords:
Proteomic genotyping
Hair shafts
Hair chemistry
Deminidation
Genetically variant peptides
Proteomics
Human identification

ABSTRACT

The use of hair evidence for human identification is undergoing considerable improvement through the adoption of proteomic genotyping. Unlike traditional microscopic comparisons, protein sequencing provides quantitative and empirically based estimates for random match probability. Non-synonymous SNPs are translated as single amino acid polymorphisms and result in genetically variant peptides. Using high resolution mass spectrometry, these peptides can be detected in hair shaft proteins and used to infer the genotypes of corresponding SNP alleles. We describe experiments to optimize the proteomic genotyping approach to individual identification from a single human scalp hair 2 cm in length (~100 µg). This is a necessary step to develop a protocol that will be useful to forensic investigators. To increase peptide yield from hair, and to maximize genetically variant peptide and ancestral information, we examined the conditions for reduction, alkylation, and protein digestion that specifically address the distinctive chemistry of the hair shaft. Results indicate that optimal conditions for proteomic analysis of a single human hair include 6 h of reduction with 100 mM dithiothreitol at room temperature, alkylation with 200 mM iodoacetamide for 45 min, and 6 h of digestion with two 1:50 (enzyme:protein) additions of stabilized trypsin at room temperature, with stirring incorporated into all three steps. Our final conditions using optimized temperatures and incubation times increased the average number of genetically variant peptides from 20 ± 5 to 73 ± 5 ($p = 1 \times 10^{-13}$), excluding intractable hair samples. Random match probabilities reached up to 1 in 620 million from a single hair with a median value of 1 in 1.1 million, compared to a maximum random match probability of 1 in 1380 and a median value of 1 in 24 for the original hair protein extraction method. Ancestral information was also present in the data. While the number of genetically variant peptides detected were equivalent for both European and African subjects, the estimated random match probabilities for inferred genotypes of European subjects were considerably smaller in African reference populations and *vice versa*, resulting in a difference in likelihood ratios of 6.8 orders of magnitude. This research will assure uniformity in results across different biogeographic backgrounds and enhance the use of novel peptide analysis in forensic science by helping to optimize genetically variant peptide yields and discovery. This work also introduces two algorithms, GVP Finder and GVP Scout, which facilitate searches, calculate random match probabilities, and aid in discovery of genetically variant peptides.

1. Introduction

Hair is a ubiquitous biological material that is shed from the human body at a rate of about 100–150 scalp hair shafts per day [1]. Because hair is a complex biological material, it contains information that can potentially be exploited to provide a link between an individual and a location [2–4]. Forensic hair analysis for identification of individuals, ancestry and species has historically been conducted using morphologic hair comparison, which is now considered controversial [5–12]. Hair shaft protein was recently demonstrated to be a carrier of genetic information in the form of genetically variant peptides (GVPs) [13]. These peptides contain single amino acid polymorphisms, the result of non-synonymous SNPs. Detection of these peptides allows for the inference of the corresponding SNP genotypes [13]. Like any DNA genotype, these can be used to estimate random match probability (RMP) and to statistically associate an individual with a given hair shaft [13].

Abbreviations: ABC, ammonium bicarbonate; DTE, dithioerythritol; DTT, dithiothreitol; GVP, genetically variant peptide; IA, iodoacetamide; KAP, keratin-associated protein; MAF, minor allele frequency; RMP, random match probability; RMT, reductively methylated trypsin; SD, sodium dodecanoate; SNP, single nucleotide polymorphism; TFA, trifluoroacetic acid

* Corresponding author.
E-mail address: gjparker@ucdavis.edu (G.J. Parker).

https://doi.org/10.1016/j.fsigen.2020.102314
Received 11 December 2019; Received in revised form 25 March 2020; Accepted 15 May 2020
Available online 25 May 2020

However, in order to be useful to the forensic science community, several technical issues must be addressed. Primary among these is the need to obtain forensically relevant RMPs from a single human hair [2,13–16].

Hair is a challenging substrate. The bulk of hair consists of highly structured keratin intermediate filaments that are stabilized by a range of covalent bonds that result in a physically robust and chemically resistant tissue [17–19]. These covalent bonds consist of isopeptide bonds, the result of transglutaminase reactions, and particularly high levels of disulfide bonds. Keratins and particularly keratin-associated proteins (KAPs) are cysteine-rich, resulting in a highly cross-linked tissue matrix [18,20]. Hair remains an underutilized forensic substrate that contains important biological information from mitochondrial and fragmented nuclear DNA, proteins, and other, small molecules. Any protocol development would need to balance the chemical fragility of the target molecule against the conditions required to thoroughly decontaminate the hair surface or open up the hair matrix for proteolytic release of internal biomolecules. An ideal processing protocol would efficiently and consistently release informative molecules from the matrix with minimal introduction of analytical biases, regardless of hair biology or human behavior. The starting point for any such protocol should be based on the biochemical and biophysical nature of the hair shaft.

This project is a systematic evaluation of chemical treatments of hair shafts from the scalp to maximize the proteomic yield of GVPs using subjects of European or African ancestry. Present work reaches a counter-intuitive finding that milder conditions result in maximal detection and identification of target GVPs. A significant increase in the amount of DTT reductant, up to 100 mM, maintains the gentle conditions while also opening up the keratin matrix to increase the release of peptides from keratin-associated and other proteins. These optimizations, when applied to single hairs, increase proteolytic release of KAPs and detection of GVPs. A single 2 cm hair shaft resulted in detection of up to 80 GVPs with an RMP of up to 1 in 620 million, a three-fold increase of GVP detection and an average increase in RMP of 4 orders of magnitude compared to earlier findings. Tools have also been developed to more efficiently identify and discover GVPs in proteomic data and are hereby made available to the forensic community.

2. Materials and methods

2.1. Hair collection and preparation

Reference hair and matching DNA were collected from 3 self-described African subjects (Sorenson Forensics LLC, Salt Lake City, UT) and 3 self-described European subjects (Davis, CA) using IRB compliant protocols (IRB# 832776-10). Only two biogeographic groups were studied in this work to demonstrate a proof of concept of the work. These two groups were chosen to represent typical demographic groups in the United States. The average length of hair on the head before cutting was 10 cm. Hair roots were purposefully excluded from the processing. Hairs were collected by cutting a few inches inward from the distal end. Hair shafts were either weighed to give 4 mg of hair per subject per replicate, or cut to 2 cm in length with no regard to distal or proximal orientation.

All hair shafts were washed three times in 1 mL of 2% (w/v) sodium dodecanoate (SD) (Sigma-Aldrich, St. Louis, MO) in 50 mM ammonium bicarbonate (ABC) (Honeywell, Muskegon, MI) to minimize contamination from exogenous materials, such as environmental epidermal corneocytes. Samples were vortexed for 10 s with each wash, and the wash eluent was discarded. For single hair analysis, a 2 cm length was cut into 10 separate 2 mm segments and placed in a protein LoBind tube (Eppendorf, Hamburg, Germany) with the entire hair shaft submerged in solution. Hair samples of 4 mg were left intact and not cut into segments. All reagent solutions were passed through solid-phase extraction filtration with the exception of the reductively methylated trypsin (RMT) [21] and SD, as these would bind to the stationary phase of the cartridge. This step was applied to minimize contamination by exogenous organic material.

2.2. Chemical processing optimization

The starting chemistry for proteomic processing of human hair was obtained from an NCJRS report [22] and related publications [23,24]. This method, referred to as the original processing method, employed overnight incubation at a high temperature for disulfide reduction and 3 days of digestion. In this method, 400 µL of a solution of 2% SD + 50 mM ABC and 50 mM dithioerythritol (DTE) (Sigma-Aldrich, St Louis, MO) was added to the LoBind tube with 4 mg of prepared hair. A cleaned magnetic stir flea (Sigma-Aldrich) was added to the tube and stirred at medium speed for 1 h at room temperature before incubation in an oven with no agitation at 70 °C for 18 h. Samples were again stirred at medium speed at room temperature for 1 h. Free thiols were alkylated with the addition of iodoacetamide (IA) (Sigma-Aldrich) to give a final concentration of 100 mM. The hair-containing solution was stirred in the dark for 45 min. The sample was then acidified (pH ~2) with 8 µL of trifluoroacetic acid (TFA) (ThermoFisher, Chicago, IL) to precipitate the detergent. Detergent extraction was achieved using three consecutive additions of 700 µL ethyl acetate (Sigma-Aldrich). For each extraction, the sample was vortexed and then centrifuged for 3 min at 14,000 relative centrifugal force (rcf). The organic (upper) phase was removed by pipetting with care not to disturb the interphase containing denatured protein and/or fragmented hair. The pH was then adjusted to ~8 using 2.5 µL of ammonium hydroxide (Fisher Scientific) and 25 µL of 1 M ABC. Three 1:50 (enzyme:protein) additions of RMT were added to the sample, with one addition per day for three days [21]. Digests were then centrifuged at 14,000 rcf for 15 min, and the supernatant was collected for mass spectral analysis. The only modifications to this protocol were made during the final optimization comparison, where the volume of reagents was reduced by 75%, for a final volume of ~160 µL, and 2 cm of a hair shaft was used instead of 4 mg.

The resulting chemistry for proteomic processing of human hair, referred to as the optimized processing method, employs a 14 h protocol. In this method, 100 µL of a solution of 2% SD + 50 mM ABC + 100 mM dithiothreitol (DTT) (Invitrogen, Carlsbad, CA) was added to each LoBind tube with 2 cm of prepared and cut hair. A cleaned magnetic stir flea was added to the tube and stirred at medium speed for 6 h at room temperature. Free thiols were then alkylated with the addition of IA to a final concentration of 200 mM, and the solution was stirred in the dark for 45 min. The sample was then acidified to a pH of ~2 using 2 µL of TFA to precipitate the detergent. Detergent extraction was achieved using three consecutive additions of 175 µL of Ethyl acetate. For each extraction, the sample was vortexed and then centrifuged for 3 min at 14,000 rcf to minimize the interphase containing denatured protein and/or fragmented hair before pipetting off the upper organic phase. The pH was then adjusted to ~8 using 6.3 µL of 1 M ABC and 0.6 µL of ammonium hydroxide. Two 1:50 (enzyme:protein) additions of RMT were added to the sample, with one addition every 3 h, for a total digestion time of 6 h. Digests were then centrifuged at 14,000 rcf for 15 min, and the supernatant was collected for mass spectral analysis.

2.3. Peptide quantification

Digestion efficiency was quantified by reaction of insoluble protein with ninhydrin after hydrolysis with 10% sulfuric acid [25,26]. Samples were analyzed based on A570 and compared to a standard curve of hydrolyzed bovine serum albumin. The percentage (w/w) of hair that was in the insoluble fraction was then calculated using the mass of the insoluble pellet divided by the total hair mass, which was usually 4 mg for initial experiments. Before instrumental analysis, solubilized tryptic

peptides were quantified using the Pierce™ Quantitative Fluorometric Peptide Assay (ThermoFisher) after 1:10 dilution. Fluorescence was measured using a Synergy H1 hybrid multi-mode reader (BioTek, Winooski, VT).

2.4. Data acquisition

Samples were analyzed using a ThermoScientific Q-Exactive Plus Orbitrap mass spectrometer with built in Proxeon nanospray and Proxeon Easy-nLC II HPLC. A sample (10 µL) containing 0.75 µg of digested peptide material was loaded on a 100 µm × 25 mm Magic C18 100 Å 5 U reverse phase trap, desalted online and separated over 140 min gradient using a 75 µm ×150 mm Magic C18 200 Å 3 U reverse phase column at 300 nL/min flow rate [27]. The solvent gradient for the elution of peptides began with 5% acetonitrile (ACN) and increased linearly to 20% ACN at 92 min, 32% ACN at 112 min, and 80% ACN at 119 min. The 80% ACN solvent ratio was maintained for 10 min, reduced to 5% at 130 min, and held for 10 min. MS survey was conducted at the *m/z* range of 350–1600, and the 15 most abundant ions from the spectra were subjected to higher-energy C-trap dissociation (HCD) to fragment the precursor peptides and obtain MS/MS spectra [28]. Precursor ions selected in a 1.6 *m/z* isolation mass window were fragmented via 27% normalized collision energy. A 20 s duration was used for dynamic exclusion.

2.5. Data analysis

Raw data files were converted into mzML format using MsConvert GUI software (Proteowizard 2.1, http://proteowizard.sourceforge.net). Files were converted using numpress linear compression and numpress short logged float compression along with peak picking with vendor algorithm for all mass spectrometry levels. These mzML files were then analyzed using GPM Fury software (X!Tandem Alanine (2016.10.15.2)) using the advanced search option. Default search settings were chosen except for exclusion of prokaryotes and viruses in the taxon heading, peptide and protein log(e) score minimum of −1 and −1 respectively, fragment mass error of 20 ppm, parent mass error of ± 100 ppm, and inclusion of point mutations under the refinement specification heading [27]. Post-search filtering based on specific transition levels was manually applied to GVP spectra to account for broad mass error filtering. The output from X!Tandem in the Global Proteome Machine environment included the annotation of single amino acid variants, that were genetic or chemical in origin. These annotations form the basis of subsequent analyses of GVP discovery, detection and post-translational modifications.

A spreadsheet, termed GVP Finder (v1.1), was created to search for GVPs and calculate random match probabilities (RMPs). This spreadsheet can be obtained from the resources menu of (https://parkerlab.ucdavis.edu). In short, previously identified GVPs were searched for by exporting each sample peptide spreadsheet in the GPM Fury software and then were bioinformatically extracted from the list of total identified peptide spectral matches. These GVPs were prescreened to eliminate those that were not unique, defined as sharing the amino acid sequence from another gene product in the human proteome including variants. Unique sequences that correspond to GVPs were searched for, along with chemical modifications or single amino acid polymorphisms. False positive rates, due to errors in peptide spectral matching or errors in software or spreadsheet analysis, were not able to be measured when used in isolation. GVP detection required subsequent validation through DNA genotyping of matching DNA samples. Genotypic frequencies from the European and African reference populations of the 1000 Genomes Consortium were consulted to calculate RMP [29]. When combining datasets from three biological replicates of a sample, presence of a GVP was determined by detection in any of the datasets, with no additional weighting for the second identification. RMPs from combined datasets are reported as averaged and not a cumulative probability with higher discrimination.

2.6. Calculation of random match probability

RMP was calculated using the product rule [13,30] with genotypic frequencies from the 1000 Genomes Project (https://www.internationalgenome.org) from five populations; African, European, East Asian, South Asian, and American [29]. Complete linkage for GVPs shared within an open reading frame was assumed as well as no linkage between open reading frames of different genes. For GVPs that were determined to be genetically linked within an open reading frame, a cumulative genotypic frequency was estimated using summation of all potential diplotype combinations. Sensitivity was calculated as the true positive rate divided by the sum of true positives and false negatives. Homozygosity was not assumed when only one allele was detected from a locus. Instead, the estimated genotype frequency ($gf_p = p2 + 2pq$) from the reference population was substituted [27]. To avoid a null value, each genotypic frequency was expressed as (x + ½)/(n + 1), where 'n' is the sample size and 'x' is the number of individuals with a given SNP, or combination of SNPs, in the sample population [31,32].

2.7. Genetic validation of variant peptides

Matching genomic DNA was extracted from buccal cells and saliva obtained from a mouthwash and isolated using Gentra Puregene Tissue Kit from Qiagen Inc. (European samples) or from buffy coat using an in-house phenol/chloroform protocol by Sorenson Forensics LLC, Salt Lake City, UT (African samples). Exome sequencing data was obtained using the DNA Technologies core and Bioinformatics core facilities in the Genome Center at the University of California, Davis [27]. Barcode-indexed sequencing libraries were generated from genomic DNA samples (1000 ng) sheared on an E220 Focused Ultrasonicator (Covaris, Woburn, MA). The sonicated DNA was size selected with KAPA Pure beads to obtain fragments of about 300bp. Size selected DNA (30 ng) were used for library preparations with the KAPA Hyper DNA library kit, according to the manufacturer's instructions. Ten cycles of PCR were conducted to amplify the libraries. Each library (500 ng) was pooled for exome capture using the IDT xGen® hybridization capture protocol according to the manufacturer's instructions. Seven cycles of PCR were conducted to amplify the library that was analyzed with a Bioanalyzer 2100 instrument (Agilent, Santa Clara, CA), quantified by fluorometry on a Qubit instrument (LifeTechnologies, Carlsbad, CA), and combined in two pools at equimolar ratios. The pools were quantified by qPCR with a Kapa Library Quant kit (Kapa Biosystems-Roche) and each pool was sequenced on one lane of an Illumina Nova Seq (Illumina, San Diego, CA) with paired-end 150 bp reads. Raw Illumina paired-end 151 bp reads were first subjected to quality control. Adapters were removed from the sequencing reads using scythe (https://github.com/vsbuffalo/scythe, version 0.994 beta). Base quality was controlled using a window-based method, sickle (https://github.com/najoshi/sickle, version 1.33), with the cutoff set at 30. Reads less than 30 bp in length were discarded. Reads that passed the quality control were mapped to hg19 reference genome using parameter -M for downstream analysis compatibility [33]. PCR duplicates were removed using Picard tools (http://broadinstitute.github.io/picard/, version 2.18.4). Variants were identified using HaplotypeCaller function in GATK (version 4.0.5.2), followed by variant recalibration using the recommendations from GATK developers [34]. Genotypes for the six subjects used in this research are available in Table S1.

2.8. Discovery of new genetically variant peptides

A spreadsheet, termed GVP Scout (v1.1), was created to search for putative GVPs in proteomic datasets. This spreadsheet can be obtained from the resources menu of (https://parkerlab.ucdavis.edu). In short, identified single amino acid variants from GPM software were screened

and variant peptides with matching common (> 0.5% global minor allele frequency) putative non-synonymous SNP alleles were identified and subsequently filtered manually based on exclusionary characteristics such as unique sequence, minor allele frequency, and mass shift. To prevent the inclusion of peptide with more than one genomic address, all peptide sequences were submitted to PROWL (prowl.rockefeller.edu/prowl/proteininfo) and searched against the IPI human (2010-02- 01) database. Peptides with no match or represented by a single point in the genome were considered unique and included in the study.

The putative list of GVPs was assembled based on hair proteomes using samples from the six individuals in this manuscript (Table S2). Putative GVPs were not held to stringent quality standards and were confirmed using matching mass spectral data. Transitions ideally flanked the single amino acid variant in question. The quality of the whole spectrum was also assessed. However, proteomes that differed based on the processing or analysis methods contained different members in the detected protein population that introduced additional GVPs with MAF > 0.5%. Putative GVPs that were identified in this manner underwent further standards of confirmation steps such as ensuring that the tryptic sequence was unique, the RSID corresponded to a missense mutation, and the mass shift was not due to a chemical modification. Resulting candidate GVPs underwent additional screening via DNA genotyping to become a validated GVP.

2.9. Data reporting and availability

African hair sample A1 (D1.0007) was left out of most calculations and was considered an outlier, due to its chemical intractability. Therefore, results which are reported for African samples only are reported as X ± Y, where X is the average and Y is the variance. All other error values (Y) are reported as standard deviation. Reported P-values also exclude the intractable hair sample. All RAW data files and spreadsheets of detected peptides and proteins from hair digests mentioned in this work, including from the supplemental section, are publicly available on ProteomeXchange (PDX016155) [35]. The folder also includes post-analysis using Global Proteome Machine, such as peptide and protein spreadsheets. See Table S3 for a complete list of data available.

3. Results

3.1. Time and temperature of reduction with detergent treatment

Since proteins undergo chemical modifications when treated with high temperature for long time periods [36], the first optimized parameters for proteomic processing were the duration and temperature for disulfide reduction that was conducted in the presence of detergent. For this experiment, hair samples were reduced for 18 h with 50 mM dithioerythritol (DTE) without agitation at either room temperature or in an oven at 70 °C before three days of digestion. Hair processing was assessed by quantification of the trypsin-insoluble material using ninhydrin as well as proteomic analysis. An initial prediction would be that increased solubilization of hair matrix would result in increased release, and subsequent detection, of hair shaft peptides. Indeed, lower incubation temperatures resulted in more insoluble material (Fig. 1A, S1). Insolubility was especially evident with the African hair sample that exhibited only 35% ± 7% solubilization (65% insoluble material) relative to 67% ± 1% solubilization (33% insoluble material) when treated at 70 °C ($p = 0.03$, Fig. 1A). However, the number of unique peptides actually improved under lower temperatures, increasing from 1840 ± 260 to 2570 ± 60 ($p = 0.02$) (Fig. 1C). This apparent contradiction indicated that solubilization alone is not a reliable indicator of peptide release and identification from the hair matrix. An insight into the chemical mechanisms at play in the heated sample was provided by deamidation data. Reduction at room temperature decreased the deamidation ratio, defined as the number of peptides containing deamidation divided by the total number of peptides, from 0.19 ± 0.06 to 0.05 ± 0.01 ($p = 0.007$) (Fig. 1B). This demonstrated that higher temperatures were increasing conformational mobility of the peptide and facilitating chemical modifications that change the peptide mass and result in a dilution of the initially-released peptide.

The reduction time was then assessed by comparing the 18 h 70 °C static reduction with a 6 h 23 °C reduction that incorporated stirring at medium speed (Fig. 1D). The reduction with stirring, shorter incubation time, and lower incubation temperature yielded an increase in the number of unique peptides from 2060 ± 50 to 2830 ± 70 ($p = 4 \times 10^{-4}$), compared to samples that were held static for 18 h at 70 °C. This suggests that shorter durations of reduction at room temperature are beneficial for proteome coverage and maximizing useful peptides for GVP analysis.

3.2. Trypsin time-course

The second parameter to be optimized was the time required for trypsin proteolysis. The initial condition was for three days with one 1:50 addition each day. A time-course experiment was conducted, where a single 1:50 addition of reductively methylated trypsin (RMT) was made to 4 mg of hair for one subject of European ancestry and one subject of African ancestry. Digestion was stopped by freezing at either 1, 3, 6, or 24 h. Fig. 2 demonstrates the effect digestion had on the number of unique peptides and the number of genetically variant peptides (GVPs) detected. After 6 h of digestion, both European and African hair values reached a plateau. However, the African hair samples yielded fewer unique peptides (2590 ± 10 compared to 2890 ± 50, $p = 0.01$) and fewer GVPs (38 ± 1 compared to 46 ± 3, $p = 0.02$) compared to the European samples at 6 h of digestion. This difference is primarily due to the concentration of reducing agent, as mentioned in the next section. The data suggested that there was no advantage in longer incubation times beyond the 6 h digestion period. Likewise, there were no advantages in terms of time of digestion for the detection of proteins of interest such as keratin associated proteins (KAPs) (Figure S2A).

3.3. Concentration of reducing agent

Hair shafts have high levels of disulfide bonds that result in extensive protein-to-protein cross-linking and subsequent tissue rigidity and robustness. This makes disulfide bonds an attractive target for opening up the keratin matrix to increase access to internal biomolecules in a way that avoids harsh chemistries. Accordingly, a European and an African hair sample were reduced using DTE concentrations of 25 mM, 50 mM, 75 mM, and 100 mM in biological triplicates (Fig. 3, S3, & Table S4). After trypsin digestion and proteomic mass spectrometry, resulting datasets were analyzed for protein coverage (Fig. 3A, Table S4). Higher levels of DTE increased coverage of detected proteins. At 100 mM DTE, protein coverage improved to the point that 37 of the 427 proteins had 100% coverage and 76 proteins had 50% or more coverage, compared to that at 25 mM DTE, which had 6 of the 656 proteins at 100% coverage and 53 proteins with over 50% coverage. The initial processing conditions for hair processing used 50 mM reductant [22–24], and at this level only 8 of 475 proteins had full coverage and 50 had 50% coverage or greater.

Part of the increase in protein coverage can be attributed to an increased number of identified KAPs in both the European and African hair samples (Figure S2B). This diverse family of small proteins can contain up to 36% of their amino acids as cysteine [37]. In terms of KAPs, the African hair increased from 8 ± 1 to 38 ± 2 ($p = 7 \times 10^{-4}$) and the European hair increased from 31 ± 3 to 47 ± 2 ($p = 0.009$) going from 25 to 100 mM DTE. There was also an increase in the number of detected KAPs for the European sample after reducing time and temperature during reduction and also reducing

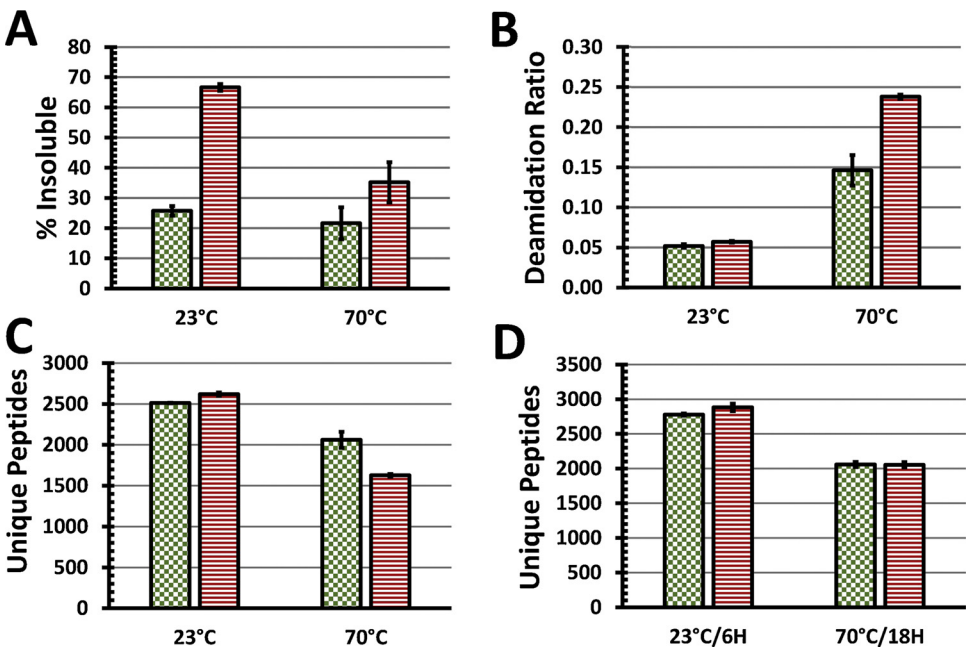

Fig. 1. Effect of temperature and time during disulfide reduction. Hair samples (4 mg) from European (green) and African (red) subjects. (A) % Protein (w/w) remaining insoluble after digestion of samples reduced at room temperature using the original processing method or at 70 °C. (B) Deamidation ratio (number of deamidations divided by the total number of peptides) as a function of incubation temperature. Conditions are the same as Fig. 1A. (C) The numbers of unique peptides from the original processing method. (D) Numbers of unique peptides compiled from the original processing method (70 °C/18 H) or at 23 °C for 6 h (23 °C/6 H). (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

digestion time (Fig. 3B). The numbers of KAPs detected were similar between the modified method (M + 100) and a previously reported urea-based method (p + 100). 47 KAPs were detected using the reduction-optimized method and 48 KAPs were detected using the urea-based method for the European sample, and 38 versus 36 KAPs for the African sample. This increase was not observed for the African hair sample until modifying the concentrations of reducing agent.

With higher levels of reductant, access to the relaxed keratin matrix facilitates the release of genetically variant peptides from other proteins (Fig. 3C). The African hair increased in GVP number from 50 ± 1 to 70 ± 8 (p = 0.02) and the European hair increased from 66 ± 2 to 83 ± 4 (p = 0.01) going from 25 to 100 mM DTE. Some GVPs were identified more frequently in non-KAP proteins when using higher concentrations of reducing agent such as those derived from SNPs rs9916724, rs9916484, and rs9916475 in KRT37. Both groups yielded the most GVPs at 100 mM DTE, which was taken as the optimum for subsequent analysis.

3.4. Comparing the finalized and original chemistries

A comparison was made between the original processing chemistry and the optimized processing chemistry for 2 cm of reference hair from six subjects (Figs. 4 and 5). Three subjects were of African ancestry and three subjects were of European ancestry. All subjects had three replicates for each condition (original and optimized) that were separately digested and analyzed. The resulting profiles of detected GVPs, as illustrated in the insert for Fig. 4 (Gene, rsID, SAP and sequence), gave inferred profile of non-synonymous SNP alleles that were directly compared with whole exome sequencing from the same individuals. Four performance outcomes for each inference (TP, true positive, blue; FP, false positive, red; TN, true negative, white; FN, false negative, green) were indicated for each broad protein class in hair shafts, keratins, KAPs and other proteins. The rate (%) of each outcome is indicated. The most noticeable improvement in true positive inference is the detection of GVPs in KAPs. The intractable hair sample was especially lacking in this protein class with only 1 GVP identified, a clear outlier. Because of this we did not include results from this sample in overall comparisons outlined below. This is primarily due to an overall loss in KAPs from family 4, 5, and 9 (Table S5). Overall sensitivity of the analysis (TP/(TP + FN)) improved 3-fold from 11% to 34%, without altering instrumental parameters. The improved sensitivity was attributed mostly to GVPs in KAPs, increasing from 0 to 49. However, more GVPs were identified and detected in all protein categories, indicating that cleavage of disulfide bonds resulted in opening up the keratin matrix and increased overall protein digestion and release of peptides from the matrix. The total identified GVPs increased from 45 to 127 for the optimized processing method (Fig. 4 & S4). The false positive rate (TP/(TP + FP)) did not change with the use of optimized chemistry.

Results indicate that the optimized processing method

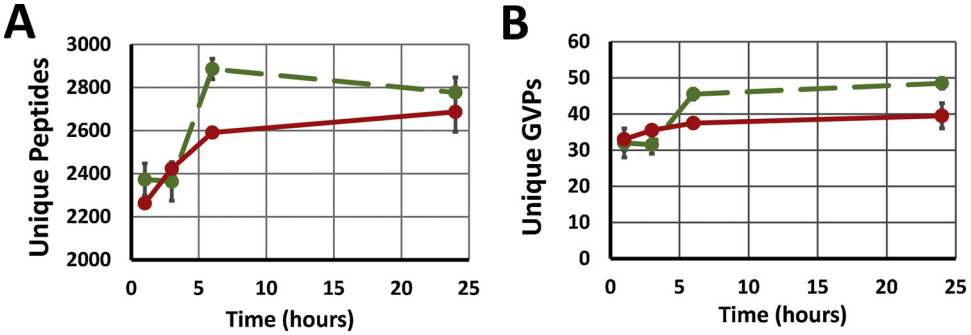

Fig. 2. Time course of hair protein digestion. Production of unique peptides (A) and unique GVPs (B) after a single 1:50 (enzyme:protein) addition of trypsin in samples from European (green-dashed line) and African (red-full line) subjects.

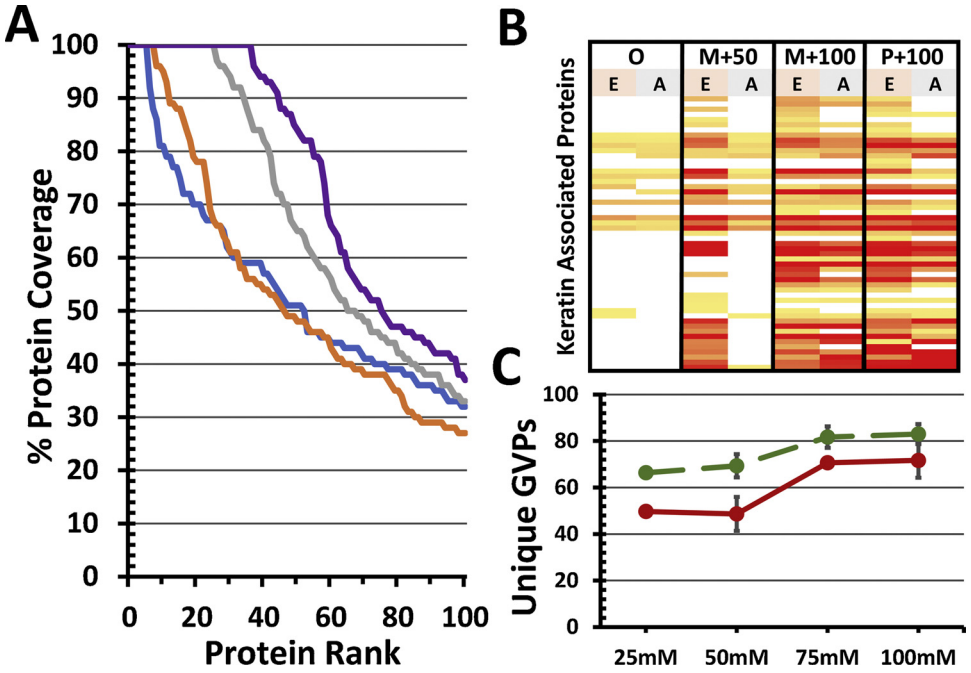

Fig. 3. Concentration of reducing agent using 4 mg of hair. (A) Protein coverage from an African sample with different concentrations of reducing agent. Blue represents 25 mM DTE, orange represents 50 mM DTE, grey represents 75 mM DTE, and purple represents 100 mM DTE. Proteins are ranked based on coverage and only 100 proteins of the highest coverage are included. See Table S4 for more details. (B) A heatmap of keratin associated proteins comparing a subject of European (E) and African (A) ancestry. White denotes no protein detected and red indicates a high level of protein detected (over 100 peptides). The abbreviation "O" indicates original method while "M" indicates use of the optimized method, "+50″ and "+100″ indicate using 50 mM and 100 mM DTE, respectively. The abbreviation "*p* + 100″ indicates a method of hair processing described by Parker et al. [13] where Protease-Max and urea were used. (C) Unique GVPs detected in samples from a subject of European (green) and African (red) ancestries processed using the optimized processing method. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

outperformed the original processing method except with an intractable hair sample from one subject (A1) (Fig. 5). Optimization of processing increased the number of unique peptides 1.7-fold from 1590 ± 160 to 2700 ± 230 ($p = 5 \times 10^{-13}$) (Fig. 5A). The average number of genetically variant peptides detected increased 3.7-fold from 20 ± 5 to 73 ± 5 ($p = 1 \times 10^{-13}$) after optimization (Fig. 5B). RMP increased from a maximum of 1 in 1400 and a median value of 1 in 24 for the original processing method to up to 1 in 620 million from a single hair with a median value of 1 in 1.1 million after chemical processing optimization ($p = 4 \times 10^{-7}$) (Fig. 5C). Likewise, median RMPs for the African samples increased from 1 in 5.1×10^1 to 1 in 1.5×10^8, and European samples increased from 1 in 1.3×10^1 to 1 in 2.2×10^3. While the numbers of unique peptides and GVPs were similar between the European and African subjects, calculated RMPs were higher (1.5×10^8 vs 2.2×10^3) in African subjects due to the differences in the genotype frequency of inferred loci in each reference population.

RMPs calculated using genotype frequencies from different reference populations (1000 Genomes Project) were compared using a likelihood ratio (LR) defined as the RMP calculated from the African population divided by the RMP calculated from the European

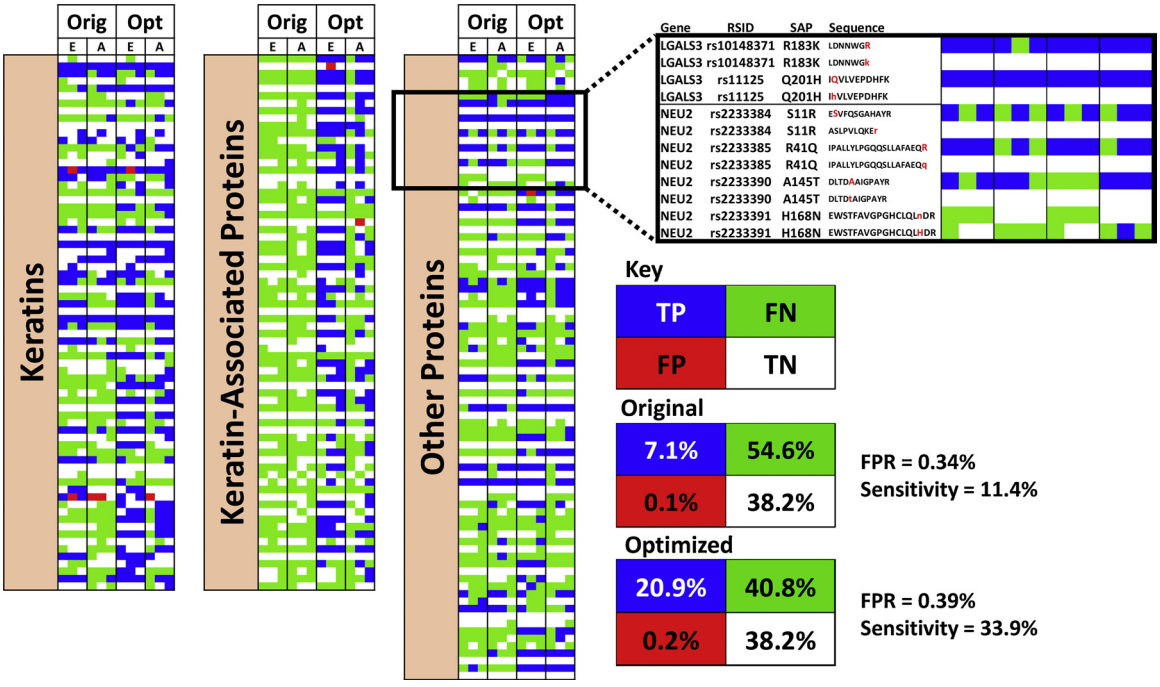

Fig. 4. GVP matrix comparing original and optimized processing methods from single hairs. This matrix represents GVPs that have been verified via whole exome sequencing. As indicated by the zoomed-in insert in the top right corner, each row is a variant peptide. Each column is an accumulated GVP profile from three replicates. Orig, original processing method; Opt, optimized processing method; E, three European subjects; A, three African subjects; TP, true positive; FN, false negative; FP, false negative; TN, true negative; FPR, false positive rate. See figure S4 for more details.

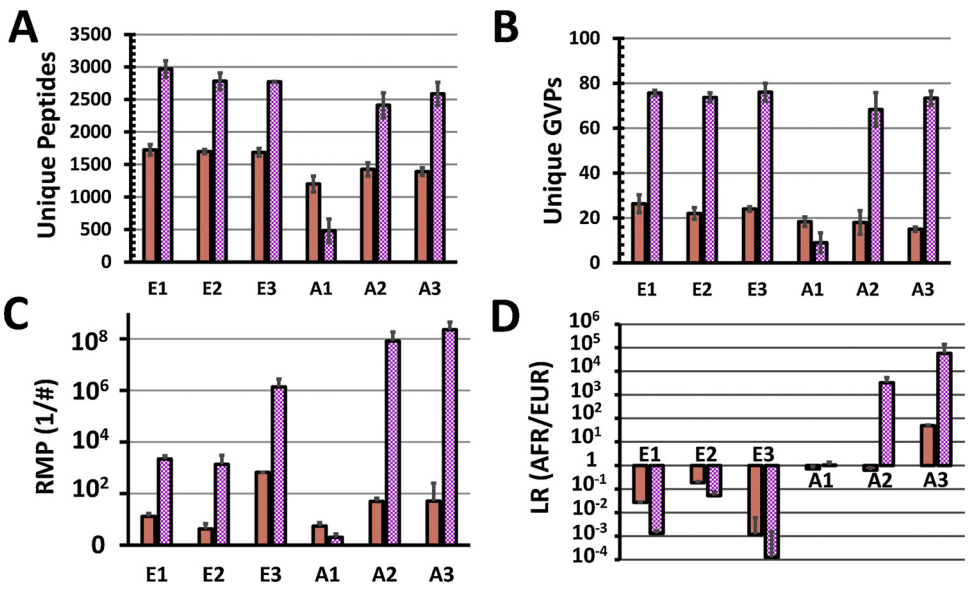

Fig. 5. Results from single hairs. Comparisons of original (salmon) and optimized (purple) methods of hair processing are shown. (A) Numbers of unique peptides; (B) Numbers of GVPs; (C) Random match probabilities; (D) Likelihood ratios from three subjects of European (E) ancestry and three subjects of African (A) ancestry. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

population (LR = Pr(GVP profile|AFR) / Pr(GVP profile|EUR)) (Fig. 5D). With optimization and increased GVP detection, the likelihood ratio for European samples decreased by 0.94 ± 0.39 orders of magnitude ($p = 1 \times 10^{-4}$), while the African samples increased by 3.90 ± 0.32 orders of magnitude ($p = 5 \times 10^{-4}$). The GVP profiles from African subjects were therefore considerably less frequent in European populations than in African ones and *vice versa*. Final likelihood ratio estimates averaged 4.1 ± 0.6 orders of magnitude for the two tractable African samples, and negative 2.7 ± 1.3 orders of magnitude (average ± standard deviation, of log transformed values) for the European samples ($p = 0.008$, using log transformed values) a difference of 6.8 orders of magnitude. These effects reflect differences in the structure of the respective reference populations. The use of LR values for ancestral characterization may be further explored with a larger cohort of Europeans and African samples.

3.5. Newly discovered genetically variant peptides

In summary, using the discovery protocols described in the Methods section, a total of 125 non-synonymous SNP loci were discovered and 152 GVPs confirmed proteomically and subsequently validated by direct comparison with DNA sequenced genotypes (Tables S1, S6, and S7). To make these discoveries, the GVP Scout spreadsheet was used and the peptides filtered for uniqueness. Non-synonymous SNP loci were identified in the genes, described in more detail in Tables S6 and S2. Of the 125 SNPs, 59 have not been reported in other forensic proteomic literature. Of these 59, six are in KRT genes and 19 are in KRTAP genes. Of particular interest are common SNPs that have a global minor allele frequency above 0.30 (rs58001094, rs2037912, rs4818950, rs2074285, rs688906, rs537301040, rs9897031, and rs238239). These loci are expected to be observed as heterozygote genotypes more frequently resulting in higher discriminatory power. A comprehensive description of the chemical and genetic properties of all GVPs used in this study is included in the Supplemental section (Tables S1 and S7).

4. Discussion

Forensically-applicable proteomic genotyping requires sample workflows to be developed that are sensitive enough to extract the necessary genetic information from the minimum of material, in this case a fraction of a single hair shaft. This development project optimized the sensitivity of hair proteomic genotyping by focusing on two factors: milder chemical conditions and sulfur chemistry. The milder conditions were assisted by the use of sodium dodecanoate that is strongly amphipathic and an effective denaturant, while also being relatively easy to remove through brief acidification and organic extraction [38]. Mild chemistries, such as lower temperatures and shorter incubation times, decreased the soluble fraction after digestion and yet increased the number of unique peptides, most likely due to the reduced level of pre-digestion peptide modification. The modification that best illustrates this is deamidation (Fig. 1B), but other modifications would also be present (data not shown). Therefore, an increase in solubilization of hair protein did not necessarily equate to better proteomic data. The overall result of using mild processing chemistries is an improvement in digestion efficiency that increased the number of unique peptides, genetically variant peptides (GVPs), and resulting random match probabilities (RMPs) from human hair. The data from 2 cm of a hair shaft is now equivalent in yield to that previously obtained from 4 mg [39] or even 10 mg [13] of hair tested. The focus on mild chemistries has the additional benefit of reduced processing times, that are currently only 14 h.

Hair has distinctively high levels of disulfide chemistry and so higher levels of reductant allowed the keratin matrix to open up further to promote hair protein proteolysis and release keratin-associated and other proteins for subsequent analysis. To optimize detection, a target peptide needs to have a maximal concentration in a sample and have minimal modifications so that signal was focused into a single mass. This requires a balance between the release of a peptide into the sample from the keratin matrix with a reduction in subsequent down-stream chemistries that will change the mass of the peptide through chemical modification, or miscleavage [40]. The chemistry required to maximize the release of target peptides from the keratin matrix also acts to modify the peptides and spread the signal across a range of masses resulting in a lower yield of unique peptides and GVPs with a single mass [26,41]. This project shifts the balance point between these two opposing factors by using high levels of reductant, as much as 100 mM dithiothreitol (DTT), and a strong detergent that opens up the keratin matrix releasing proteins and peptides without resorting to harsher chemistries. The evidence of this is the increased presence of keratin associated proteins (KAPs) in the samples, along with their GVPs (Figure 3 & 4). Increased levels of reductant have previously been shown to be critical to releasing KAPs in wool and textiles [20,42].

Earlier reports on forensic proteomics that focused on hair shaft protein used large amounts of hair, 4 or 10 mg (Table S8), since they were focused on either basic science questions, such as protein profiles, or discovery of genetically variant peptides for proteomic genotyping

[13,39,43]. Naturally, development of a forensically useful hair proteomic protocol would focus on a method that required only a fraction of a single hair shaft that would be the limit of material obtained through casework [44–47]. This study has been an open part of this process [48–51]. Over that time period other single hair methods for proteomics and proteomic genotyping have also been reported, and like this study also demonstrate high levels of protein detection and/or discrimination with 1 mm to 20 or 25 mm of a single hair shaft [44–46,52,53]. Some of the chemistry in this project is similar to that reported, but not fully documented, by other protocols [44].

Other hair processing protocols take different approaches. At one extreme a recently published method using heavily alkaline conditions was used to quickly extract around 50% of hair shaft protein [54]. These harsh conditions resulted in poor protein and peptide yields, and presumably would result in chemical degradation within the hair. One of the most widely used protocols, the Shindai method, uses 2.6 M thiourea, 5 M urea and 5% beta-mercaptoethanol at high temperatures (50 °C) for 24–72 h at pH 8.5 [54–56]. This and related commonly used methods using 8 M urea have the advantage of not relying on detergent that can be difficult to remove prior to mass spectrometry [42,45,47,56,57]. These often resulted in similar levels of protein and peptide yields [47]. Other research groups remove detergent and desalt using in-gel digestion that has the advantage of further denaturing protein and increasing fractionation [42,53]. However, in-gel digestion protocols result in sample loss since they do not use insoluble material that are a potentially rich source of proteomic material and are time and resource intensive [39]. The chemistry employed in the initial GVP-demonstration paper used urea and a mass spectrometry-compatible surfactant, along with 100 mM DTT [13]. We did not pursue development of this method, although it also achieves rich proteomic datasets for large quantities of hair, because of the chemical fragility and milder amphipathic character of the acid-labile surfactant [58].

There are still some chemistries that may be incorporated into hair sample processing. We find that 15–20% of hair mass is left insoluble after digestion. We hypothesize that this is due to covalent linkages that would not change when solubilizing in SD instead of ABC (data not shown) [15]. Improvements in the protocol may focus on stronger detergents, combined use of urea and thio-urea as used in the Shindai method. Other buffers, detergents, enzymes, and alkylating agents could still be tested to further optimize proteomic processing. Further optimization of the timing and combination of the steps employed in this project is still possible.

Intractable hair samples in our hands comprised about 3% of both African and European samples (data not shown). About 50% of intractable hair samples have undergone hair-straightening treatment. In our analysis of intractable hair, many methods were tested to aid in solubilization. Sonication, high temperatures, freeze-thawing, organic extraction, and increasing the concentration of DTT were all tested, without success. Intractable hair samples were slightly more digested using the original processing method compared to the optimized method. However, intractable hair samples still yield less than 20% of the unique peptides and unique GVPs compared to normal hair samples. The major proteomic difference between normal and intractable hair samples is that they lack peptides from KAPs that are high in cysteine content (Table S5). More effort will be invested in future research to diagnose and mitigate the problems seen with intractable hair samples.

Proteomic datasets should ideally be equivalent in terms of protein, unique peptide, and GVP number between different biogeographic groups, color, and age. Datasets differing in these characteristics may yield a systematic bias in the GVP profiles and in resulting statistical analyses between these groups. For instance, the original processing method had on average 1.4x more GVPs in the European cohort than in the African cohort. This may indicate that certain groups would hold higher evidentiary value of proteomic data. Present research, aiming to reduce statistical bias between a European and African cohort, has decreased the difference in GVP number down to 1.1x between Europeans and Africans. However, RMP calculations will still benefit from the variety and intrinsic distribution of SNPs in the African population that result from its deeper evolutionary history [29].

Future research for the study of genetically variant peptides in human hair may well involve targeted proteomics, ancestral classification, automation in sample processing, scouting and identification of novel GVPs, and developing a genotyping kit for confirmed and validated GVPs. However, the method proposed here is a significant advance and demonstrates a three-fold increase in sensitivity of GVP detection and a three orders of magnitude increase in RMP. This foundation, in addition to being a resource for the field, also allows us now to investigate other areas of development necessary for implementation as a forensic tool. These include investigating different casework scenarios that would affect data yields or introduce statistical bias into the analysis [43,59]. Our improvements also provide a foundation for further refinement of downstream mass spectrometry data acquisition and bioinformatics processing protocols.

5. Conclusion

In forensic science it is essential to maximize the extraction of the target biological material. An effective use of human hair in forensic proteomics requires sensitive and efficient sample processing protocols that can be used on a single hair shaft. Maximization of peptide production and minimization of additional chemistries is required to increase the detection of informative peptides. Harsher chemistries are especially problematic because they chemically modify peptides and further dilute the mass signatures. In this research, we combine milder digestion conditions with an increase in reductive compounds, up to 100 mM DTT, to cleave the high levels of disulfide bonds and open up the keratin and keratin-associated protein matrix. This approach should also work for those investigating other chemically fragile biomolecules in the hair shaft, such as mitochondrial DNA and chemically labile small molecules. This optimized method produces more unique peptides, genetically variant peptides, and more discriminatory random match probabilities, particularly through the release of keratin-associated proteins. Random match probability has also improved to over 1 in 600 million for a single hair. The method outlined here produces a similar number of genetically variant peptides between European and African hair digests, and significantly improves the evidentiary value of 2 cm of hair.

Funding

This project was supported by Award No. 2015-DN-BX-K065, awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect those of the Department of Justice.

CRediT authorship contribution statement

Zachary C. Goecker: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Software, Visualization, Writing - original draft, Writing - review & editing. **Michelle R. Salemi:** Investigation, Writing - review & editing. **Noreen Karim:** Validation, Writing - review & editing. **Brett S. Phinney:** Supervision, Writing - review & editing. **Robert H. Rice:** Conceptualization, Funding acquisition, Methodology, Resources, Supervision, Writing - original draft, Writing - review & editing. **Glendon J. Parker:** Conceptualization, Methodology, Project administration, Resources, Writing - original draft, Writing - review & editing.

Declaration of competing interest

The authors have declared no conflict of interest, with the exception of GJP who has a patent based on the use of genetically variant peptides for human identification (US 8877455 B2, Australian Patent 2011229918, Canadian Patent CA 2794248, and European Patent EP11759843.3). The patent is owned by Parker Proteomics LLC. Protein-Based Identification Technologies LLC (PBIT) has an exclusive license to develop the intellectual property and is co-owned by Utah Valley University and GJP. This ownership of PBIT and associated intellectual property does not alter policies on sharing data and materials. These financial conflicts of interest are administered by the Research Integrity and Compliance Office, Office of Research at the University of California, Davis to ensure compliance with University of California Policy.

Acknowledgments

The authors thank Dr. Blythe Durbin-Johnson from UC Davis and Dr. Susan Walsh of Indiana University – Purdue University Institute, for their advice. Sorenson Forensics LLC also played a crucial role for their assistance in tissue procurement. This publication was made possible, in part, with support from the UC Davis Genome Center Bioinformatics Core Facility. The sequencing was carried out at the DNA Technologies and Expression Analysis Cores at the UC Davis Genome Center, supported by NIH Shared Instrumentation Grant 1S10OD010786-01. We specifically acknowledge the assistance of Jie Li, Emily Kumimoto, Siranoosh Ashtari, Vanessa K Rashbrook, and Lutz Froenicke.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.fsigen.2020.102314.

References

[1] A.M. Kligman, Pathologic dynamics of human hair loss: I. Telogen effluvium, Arch. Dermatol. 83 (2) (1961) 175–198.
[2] N.R. Barthélemy, A. Bednarczyk, C. Schaeffer-Reiss, D. Jullien, A. Van Dorsselaer, N. Cavusoglu, Proteomic tools for the investigation of human hair structural proteins and evidence of weakness sites on hair keratin coil segments, Anal. Biochem. 421 (1) (2012) 43–55.
[3] F. Desiere, E.W. Deutsch, A.I. Nesvizhskii, P. Mallick, N.L. King, J.K. Eng, A. Aderem, R. Boyle, E. Brunner, S. Donohoe, N. Fausto, Integration with the human genome of peptide sequences obtained by high-throughput mass spectrometry, Genome Biol. 6 (1) (2005) 9.
[4] C.S.R. Kadiyala, S.E. Tomechko, M. Miyagi, Perfluorooctanoic acid for shotgun proteomics, PLoS One 5 (12) (2010) e15332.
[5] H. Edwards, C. Gotsonis, Strengthening Forensic Science in the United States: A Path Forward, Statement before the United State Senate Committee on the Judiciary, 2009.
[6] M.M. Houck, B. Budowle, Correlation of microscopic and mitochondrial DNA hair comparisons, J. For. Sci. 47 (5) (2002) 1–4.
[7] B.J. Kim, J.I. Na, W.S. Park, H.C. Eun, O.S. Kwon, Hair cuticle differences between Asian and Caucasian females, Int. J. Dermatol. 45 (12) (2006) 1435–1437.
[8] J. Ogle, M.J. Fox, Atlas of Human Hair: Microscopic Characteristics. Crc Press, 2017.
[9] C.L. Brace, Region does not mean "race"—reality versus convention in forensic anthropology, J. For. Sci. 40 (2) (1995) 171–175.
[10] D.W. Deedrick, S.L. Koch, Microscopy of hair part 1: a practical guide and manual for human hairs, For. Sci. Commun. 6 (1) (2004).
[11] S.L. Koch, M.D. Shriver, N.G. Jablonski, Variation in human hair ultrastructure among three biogeographic populations, J. Struct. Biol. 205 (1) (2019) 60–66.
[12] J.W. Hicks, Microscopy of Hairs: A Practical Guide and Manual, Federal Bureau of Investigation, FBI Laboratory, 1977.
[13] G.J. Parker, T. Leppert, D.S. Anex, J.K. Hilmer, N. Matsunami, L. Baird, J. Stevens, K. Parsawar, B.P. Durbin-Johnson, D.M. Rocke, C. Nelson, Demonstration of protein-based human identification using the hair shaft proteome, PLoS One 11 (9) (2016) .e0160653.
[14] M.T.P. Gilbert, A.S. Wilson, M. Bunce, A.J. Hansen, E. Willerslev, B. Shapiro, T.F. Higham, M.P. Richards, T.C. O'Connell, D.J. Tobin, R.C. Janaway, Ancient mitochondrial DNA from hair, Curr. Biol. 14 (12) (2004) 463–R464.
[15] R.H. Rice, Proteomic analysis of hair shaft and nail plate, J. Cosmet. Sci. 62 (2) (2011) 229.
[16] R.H. Rice, K.M. Bradshaw, B.P. Durbin-Johnson, D.M. Rocke, R.A. Eigenheer, B.S. Phinney, J.P. Sundberg, Differentiating inbred mouse strains from each other and those with single gene mutations using hair proteomics, PLoS One 7 (12) (2012) e51956.
[17] C.C. Chou, M.J. Buehler, Structure and mechanical properties of human trichocyte keratin intermediate filament protein, Biomacromolecules 13 (11) (2012) 3522–3532.
[18] J.E. Plowman, D.P. Harland, S. Deb-Choudhury (Eds.), The Hair Fibre: Proteins, Structure and Development, Vol. 1054 Springer, 2018.
[19] C.H. Lee, M.S. Kim, B.M. Chung, D.J. Leahy, P.A. Coulombe, Structural basis for heteromeric assembly and perinuclear organization of keratin filaments, Nat. Struct. Mol. Biol. 19 (7) (2012) 707.
[20] C. Solazzo, J.M. Dyer, S. Clerens, J. Plowman, E.E. Peacock, M.J. Collins, Proteomic evaluation of the biodegradation of wool fabrics in experimental burials, Int. Biodeterior. Biodegrad. 80 (2013) 48–59.
[21] R.H. Rice, G.E. Means, W.D. Brown, Stabilization of bovine trypsin by reductive methylation, Biochim. Biophys. Acta (BBA)-Protein Struct. 492 (2) (1977) 316–321.
[22] R.H. Rice, P.W. Wu, S.M. Mann, Human Hair Proteomics-Improved Evidence Discrimination, (2015) https://www.ncjrs.gov/pdffiles1/nij/grants/249102.pdf.
[23] C.N. Laatsch, B.P. Durbin-Johnson, D.M. Rocke, S. Mukwana, A.B. Newland, M.J. Flagler, M.G. Davis, R.A. Eigenheer, B.S. Phinney, R.H. Rice, Human hair shaft proteomic profiling: individual differences, site specificity and cuticle analysis, PeerJ 2 (2014) e506.
[24] P.W. Wu, K.E. Mason, B.P. Durbin-Johnson, M. Salemi, B.S. Phinney, D.M. Rocke, G.J. Parker, R.H. Rice, Proteomic analysis of hair shafts from monozygotic twins: expression profiles and genetically variant peptides, Proteomics 17 (13–14) (2017) 1600462.
[25] S. Moore, W.H. Stein, A modified ninhydrin reagent for the photometric determination of amino acids and related compounds, J. Biol. Chem. 211 (2) (1954) 907–913.
[26] G. Schiffman, Immunological methods for characterizing polysaccharides, Method Enzymol. 8 (1966) 79–85.
[27] T. Borja, N. Karim, Z. Goecker, M. Salemi, B. Phinney, M. Naeem, R. Rice, G. Parker, Proteomic genotyping of fingermark donors with genetically variant peptides, For. Sci. Int. Genet. (2019), https://doi.org/10.1016/j.fsigen.2019.05.005.
[28] E.L. de Graaf, A.M. Altelaar, B. van Breukelen, S. Mohammed, A.J. Heck, Improving SRM assay development: a global comparison between triple quadrupole, ion trap, and higher energy CID peptide fragmentation spectra, J. Proteome Res. 10 (9) (2011) 4334–4341.
[29] 1000 Genomes Project Consortium, A global reference for human genetic variation, Nature 526 (7571) (2015) 68.
[30] I.W. Evett, B.S. Weir, Interpreting DNA Evidence: Statistical Genetics for Forensic Scientists, Sinauer Associates, Sunderland MA, 1998.
[31] H. Jeffreys, An invariant form for the prior probability in estimation problems, *Proc. R. Soc. Lond. Ser. A. Math. Phys. Sci.* 186 (1946) 453–461 1007.
[32] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, D.B. Rubin, Bayesian Data Analysis, Chapman and Hall/CRC, 2013.
[33] H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM, arXiv preprint arXiv (2013), pp. 1303–3997.
[34] R. Poplin, V. Ruano-Rubio, M.A. DePristo, T.J. Fennell, M.O. Carneiro, G.A. Van der Auwera, D.E. Kling, L.D. Gauthier, A. Levy-Moonshine, D. Roazen, K. Shakir, Scaling accurate genetic variant discovery to tens of thousands of samples, BioRxiv (2018) 201178.
[35] Y. Perez-Riverol, Q.W. Xu, R. Wang, J. Uszkoreit, J. Griss, A. Sanchez, F. Reisinger, A. Csordas, T. Ternent, N. del-Toro, J.A. Dianes, PRIDE inspector toolsuite: moving toward a universal visualization tool for proteomics data standard formats and quality assessment of ProteomeXchange datasets, Mol. Cell. Proteom. 15 (1) (2016) 305–317.
[36] P. Hao, Y. Ren, A.J. Alpert, S.K. Sze, Detection, evaluation and minimization of nonenzymatic deamidation in proteomic sample preparation, Mol. Cell. Proteom. 10 (10) (2011) O111–009381.
[37] M.A. Rogers, L. Langbein, S. Praetzel-Wunder, H. Winter, J. Schweizer, Human hair keratin-associated proteins (KAPs), Int. Rev. Cytol. 251 (2006) 209–263.
[38] Y. Lin, L. Huo, Z. Liu, J. Li, Y. Liu, Q. He, X. Wang, S. Liang, Sodium laurate, a novel protease-and mass spectrometry-compatible detergent for mass spectrometry-based membrane proteomics, PLoS One 8 (3) (2013) e59779.
[39] Y.J. Lee, R.H. Rice, Y.M. Lee, Proteome analysis of human hair shaft: from protein identification to posttranslational modification, Mol. Cell. Proteom. 5 (5) (2006) 789–800.
[40] T. Šlechtová, M. Gilar, K. Kalíková, E. Tesařová, Insight into trypsin miscleavage: comparison of kinetic constants of problematic peptide sequences, Anal. Chem. 87 (15) (2015) 7636–7643.
[41] N.E. Robinson, A.B. Robinson, Deamidation of human proteins, Proc. Natl. Acad. Sci. 98 (22) (2001) 12409–12413.
[42] J.E. Plowman, S. Deb-Choudhury, A. Thomas, S. Clerens, C.D. Cornellison, A.J. Grosvenor, J.M. Dyer, Characterisation of low abundance wool proteins through novel differential extraction techniques, Electrophoresis 31 (12) (2010) 1937–1946.
[43] J.A. Milan, P.W. Wu, M.R. Salemi, B.P. Durbin-Johnson, D.M. Rocke, B.S. Phinney, R.H. Rice, G.J. Parker, Comparison of protein expression levels and proteomically-inferred genotypes using human hair from different body sites, For. Sci. Int. Genet. 41 (2019) 19–23.
[44] K.E. Mason, P.H. Paul, F. Chu, D.S. Anex, B.R. Hart, Development of a protein-based human identification capability from a single hair, J. For. Sci. (2019), https://doi.org/10.1111/1556-4029.13995.
[45] T.L. Carlson, M. Moini, B.A. Eckenrode, B.M. Allred, J. Donfack, Protein extraction

from human anagen head hairs 1-millimeter or less in total length, BioTechniques 64 (4) (2018) 170–176.
[46] F. Lei, J. Li, L. Shan-Fei, Z. Jian, L. Hai-Bo, J. An-Quan, Y. Jian, W. Gui-Qiang, L. Cai-Xia, Development and validation of protein-based forensic ancestry inference method using hair shaft proteome, Prog. Biochem. Biophys. 46 (1) (2019) 81–88.
[47] L.A. Catlin, R.M. Chou, Z.C. Goecker, L.A. Mullins, D.S.B.S.S. Silva, R.R. Spurbeck, G.J. Parker, C.M. Bartling, Demonstration of a mitochondrial DNA-compatible workflow for genetically variant peptide identification from human hair samples, For. Sci. Int. Genet. 43 (2019) 102148.
[48] Z.C. Goecker, S.R. Salemi, B.S. Phinney, R.H. Rice, G.J. Parker, Optimization of Peptide Generation From Human Hair for Proteomic Analysis, Proceedings of the 65th Annual American Society for Mass Spectrometry (2017) 4–8 JuneIndianapolis, IN..
[49] Z.C. Goecker, S.R. Salemi, B.S. Phinney, R.H. Rice, G.J. Parker, Optimization of Human Hair Proteomic Processing to Maximize Total and Genetically-variant Peptides, Proceedings of the 28th Annual International Symposium on Human Identification, October Seattle, WA, Poster #19, 2017, pp. 2–5.
[50] Z.C. Goecker, S.R. Salemi, B.S. Phinney, R.H. Rice, G.J. Parker, The Optimization of Human Hair Proteomic Processing for Single Hair and Ancestral Analysis, Proceedings of the 70th Annual American Academy of Forensic Sciences (2018) 19–24 FebruarySeattle, WA, Poster B57.
[51] G.J. Parker, Interplay Between Proteomic and Genomic Datasets: Merging Both Information Flows in a Forensic Context, Proceedings of the Gordon Research Conference, Forensic Analysis of Human DNA (2018) 17–22 JuneNewry, ME.
[52] K.F. Jones, T.L. Carlson, B.A. Eckenrode, J. Donfack, Assessing protein sequencing in human single hair shafts of decreasing lengths, For.Sci. Int. Genet. 44 (2020) 102145.
[53] Z. Zhang, M.C. Burke, W.E. Wallace, Y. Liang, S.L. Sheetlin, Y.A. Mirokhin, D.V. Tchekhovskoi, S.E. Stein, Sensitive method for the confident identification of genetically variant peptides in human hair keratin, J. For. Sci. (2019), https://doi.org/10.1111/1556-4029.14229.
[54] S.Y. Wong, C.C. Lee, A. Ashrafzadeh, S.M. Junit, N. Abrahim, O.H. Hashim, A high-yield two-hour protocol for extraction of human hair shaft proteins, PLoS One 11 (10) (2016) e0164993.
[55] T. Fujii, D. Li, Preparation and properties of protein films and particles from chicken feather, J. Biomacromol. 8 (2008) 48–55.
[56] A. Nakamura, M. Arimoto, K. Takeuchi, T. Fujii, A rapid extraction procedure of human hair proteins and identification of phosphorylated species, Biol. Pharm. Bull. 25 (5) (2002) 569–572.
[57] N. Araki, M. Moini, Age estimation of museum wool textiles from Ovis aries using deamidation rates utilizing matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, Rapid Commun. Mass Spectrom. 25 (22) (2011) 3396–3400.
[58] P.E. Hellberg, K. Bergström, K. Holmberg, Cleavable surfactants, J. Surfactants Deterg. 3 (1) (2000) 81–91.
[59] R.N. Franklin, N. Karim, Z.C. Goecker, B.P. Durbin-Johnson, R.H. Rice, G.J. Parker, Proteomic genotyping: using mass spectrometry to infer SNP genotypes in pigmented and non-pigmented hair, Forensic Sci. Int. 310 (2020) 110200.