Draft Final Summary Overview

Document Title: Use of Genetically Variant Peptides to Statistically Estimate the Genetic Background of Hair Shafts

Award: 2015-DN-BX-K065
PI: Glendon Parker

Products: Accepted peer-reviewed manuscripts only[1-7].

1. Tempest J. Plott*, Noreen Karim, Blythe P. Durbin-Johnson, Dionne P. Swift, R. Scott Youngquist, Michelle Salemi, Brett S. Phinney, David M. Rocke, Michael G. Davis, **Glendon J. Parker**†, Robert H. Rice† Age-Related Changes in Hair Shaft Protein Profiling and Genetically Variant Peptides, **Forensic Science International: Genetics** (2020) 47 (102309):1-8 (https://doi.org/10.1016/j.fsigen.2020.102309) †equal contribution
2. Zachary C. Goecker*, Michelle R. Salemi, Noreen Karim, Brett S. Phinney, Robert H. Rice, **Glendon J. Parker** Optimal Processing for Proteomic Genotyping of Single Human Hairs, **Forensic Science International: Genetics**, (2020) 47 (102314):1-10 (https://doi.org/10.1016/j.fsigen.2020.102314)
3. Franklin RN*, Karim N, Goecker ZC*, Durbin-Johnson BP, Rice RH, and **Parker GJ.** Proteomic Genotyping: Using Mass Spectrometry to Infer SNP Genotypes in Pigmented and Non-Pigmented Hair, **Forensic Science International**, (2020) 110200 (https://doi.org/10.1016/j.forsciint.2020.110200) (1 citation)
4. Trevor Borja*, Noreen Karim*, Zachary Goecker*, Michelle Salemi, Brett Phinney, Muhammad Naeem, Robert Rice, **Glendon Parker.** Proteomic genotyping of fingermark donors with genetically variant peptides, **Forensic Science International: Genetics,** (2019), 42: 21–30 (7 citations) https://doi.org/10.1016/j.fsigen.2019.05.005
5. Milan JA*, Wu P-Wen, Salemi MR, Durbin-Johnson BP, Rocke DM, Phinney BS, Rice RH, **Parker GJ**, Comparison of protein expression levels and proteomically-inferred genotypes using human hair from different body sites, **Forensic Science International: Genetics,** 41 (2019), 19 -23 https://doi.org/10.1016/j.fsigen.2019.03.009 (5 citations)
6. Pei-Wen Wu, Katelyn E. Mason, Blythe P. Durbin-Johnson, Michelle Salemi, Brett S. Phinney, David M. Rocke, **Glendon J. Parker**, Robert H. Rice. Proteomic analysis of hair shafts from monozygotic twins: Expression profiles and genetically variant peptides. **Proteomics, 17**, (13–14), 2017, 1600462 (https://doi.org/10.1002/pmic.201600462) (18 citations)
7. **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, D. J. Fairbanks, A. S. Wilson, R. H. Rice, S. R. Woodward, B. Bothner, B. R. Hart, and M. Leppert. A Demonstration of Protein-Based Human Identification Using the Hair Shaft Proteome **PLOS ONE***,* Sept 7, 2016, *http://dx.plos.org/10.1371/journal.pone.0160653*. (30,567 views and 4,120 downloads as of Jan 22, 2019) (59 citations)

Other products[8]:

8. **Glendon Parker**, Zachary Goecker, Rachel Franklin, Blythe Durbin-Johnson, Jennifer Milan, Christina De Leon, Noreen Karim, Ashleigh Matzoll, Trevor Borja, Robert Rice. Proteomic genotyping: using mass spectrometry to infer SNP genotypes in a forensic context. **Forensic Science International: Genetics, Supplement Series** (2019) 7(1): 664-666

1. Statement of the Problem / Research Question

1.1 The Probative Value of Hair Shafts. *Hair is part of many crime scenes but it is underused forensically[9].*

Each individual sheds about 100 to 200 cranial hairs a day, about one every 10 to 15 min[10]. Hair shafts are

robust and persist in the environment[11-13]. Hair shafts are highly complex tissues and contain high levels of biological information an, their presence can associate a suspect to a victim or a suspect to a crime scene. Before the DNA revolution, hair shaft analysis was a major component of crime scene investigation[10, 14]. DNA based methods are difficult to employ on hair shafts however. The cornification process of hair shafts degrades all nuclear DNA as part of the partial apoptosis involved in terminal differentiation of the hair shaft, resulting in DNA pieces averaging round 30 bp in length[15, 16]. Likewise, the traditional forensic analysis of hair shaft patterns, to associate an individual to evidentiary hair found at a crime scene has been thoroughly discredited and found to be intrinsically subjective in its analysis[17-19]. The dependence of investigators on DNA-based methods, and the discrediting of hair morphological comparisons, have resulted in hair no longer being routinely examined in evidence despite of its ubiquity and potential to associate individuals with a crime scene.

1.2 Proteomic Genotyping and its Applications Protein contains genetic information in the form of single amino acid polymorphisms (SAPs), the result of non-synonymous SNPs[20, 21]. Proteomic genotyping is the detection of genetically variant peptides (GVPs) that contain a SAP and then inferring the presence of the corresponding SNP allele in the individual who produced the protein sample. In aggregate the resulting profile of inferred SNP alleles, like any profile of nucleotide variation, can be used to calculate the statistical association of an individual and a protein sample; the probability that a given combination of inferred SNP alleles would randomly occur in the reference population[7]. Currently almost 500 non-synonymous SNP alleles have been shown to be accurately inferred by genetically variant peptides, many of which are already detailed in the literature[1, 4, 7, 22-24]. Genetically variant peptides will be present in any protein sample. The samples where proteomic genotyping provides maximal benefit for a forensic investigator are samples where obtaining information through DNA-based methods are problematic. This includes telogenic hair shafts, degraded and compromised bones and teeth, fingermarks, and sexual assault evidence. Hair shafts, that are discrete, robust, and easy to obtain, have been the initial focus of the development of proteomic genotyping as a usable and practical method for investigators[7].

1.3 The Scope of Proteomic Genotyping. Proteomic genotyping is a "read" of the nucleotide information contained in message RNA and ultimately exomic DNA, filtered by the translational and transcriptional machinery of the cell. Different tissues, since they have different proteins present, will have a different set of GVPs. As mentioned above, at an individual level the SNP alleles that are reflected in different GVPs will also change as a function of the donor's DNA. At the individual level there are 10 to 12 thousand non-synonymous SNPs are present in any genome, roughly one for every 2 proteins[20]. Globally there are 122 thousand missense SNPs with minor allele frequencies above 0.5% minor allele frequency, the threshold for inclusion in our study, roughly 6 per gene product (www.ncbi.nlm.nih.gov/snp/). Given that up to 10,000 gene products are expressed at some level in the cell, roughly half of all open reading frames, this translates to about 60,000 possible GVP alleles, per cell of which about 5,000 non-reference GVPs may be present in the proteome of any individual cell. This is a significant amount of potentially useful genetic information that can be obtained without the presence of DNA in a sample.

1.4 ***Application of the Product-Rule in Proteomic Genotyping***. A profile of inferred SNP alleles, like any nucleotide profile, can be used to estimate the distribution of the given profile in a reference population[7, 25]. We use the model of random match probability described initially in [7] and refined further in [1]. Basically, we treat each open reading frame as a single locus, with multiple GVPs (if they are present), resulting in multiple effective alleles of different SNP genotype combinations. Random match probabilities can now be estimated as the product of the frequencies of each gene combination. This can be described mathematically as: $RMP = \prod(fGVP_{comb}|population)$ 1). where RMP is the random match probability, 'fGVP$_{comb}$' is the population distribution of the combination of inferred non-synonymous SNP alleles, and 'population' is the respective major reference population in the 1000 Genomes Project.

We have observed in earlier work that there are significant and cumulative differences in the genotype frequency of each GVP-inferred SNP allele in different reference populations. As a result the likelihood ratio of ancestry can vary considerably as a function of donors genetic back ground[26, 27]. These are amplified when combined as a result of the product rule, resulting in different random match probabilities for different populations by many orders of magnitude[4, 7]. These have been exploited to derive likelihood

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. 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.

values for genetic background[1, 7]. These likelihood estimates are derived as the quotient of random match probabilities derived using frequency values for different populations. The resulting likelihood is a value relative relationship between two populations. Earlier work noted that the changes in likelihood ratio of quotient of RMPs from different population values changed over 9 orders of magnitude. Therefore, there was considerable potential in using this method to provide a measure of a hair donor's relative ancestry.

2. Project Design and Methods, Data Analysis, and Findings

2.1 ***Central Hypothesis***: We therefore hypothesize that proteomic genotyping can estimate the ancestry of an individual who donates a hair shaft. This can be obtained by maximizing the efficiency of hair sample processing, followed by applying the optimized sample processing protocol to a sufficient number of subjects (n = 170) from a range of genetic backgrounds to test the hypothesis.

2.2 ***Research Program***: The research program breaks down into 5 components: 1) Optimization of hair shaft sampling methodology, 2) GVP characterization, 3) Estimation of Likelihood Values, and 4) Prediction of Biogeographic Background. In order to present data in a logical manner we change the order in how we present the data. 5) We also include validation studies conducted on real world samples following guidance outlined by SWGDAM[28]

For technical methodology we refer the reader to[1, 7].

2.3 Recruitment of Subjects. See Table 1.

Pre-existing Samples, Glendon Parker UVU	Sample size	Total
European American	13	13
Sorenson Forensics		
African	14	
African American	9	23
Davis, CA		
African	2	
African American / admixed	5	
European American	24	
Hispanic	2	
Asian	15	48
Cohort from Susan Walsh, IUPUI		
Southern European	14	
Northern Europe	14	
Western Europe	18	
Middle Eastern	12	
African	18	
Admixed	14	
Asian	8	98
		182

Table 1. Procurement of Matching Hair and DNA Samples
Samples were obtained in four batches: 1) preexisting samples, from a previous study (Parker *et al* 2016), 2) a contract with Sorensen Forensics (Salt Lake City, UT), 3) local recruitment from Davis, CA, 4) samples obtained through collaboration with Susan Walsh, IUPUI. The first three groups are self identified genetic background. The genetic background from IUPUI are from genetic ancestry analysis. A total of 182 samples were obtained.

2.4 ***Task 1: Optimization of Hair Sample Processing Chemistry***

Early iterations of the hair shaft processing protocol used 10 mg of hair shafts in order to obtain as complex a peptide mixture as possible to allow for discovery of genetically variant peptides (GVPs)[7]. This is not suitable as a forensic method, we therefore sought to optimize the sample processing method so that 20 mm of a single hair shaft could be used. We used as a starting point the protocol developed by

our collaborator Dr. Rice that relied on the harsh detergent sodium dodecyl sulfate, and later the more amenable sodium dodecanoate, and high temperatures in order to maximally solubilize the sample[6, 29].

Results. This research is fully described in Goecker et al. [1]. To summarize: the reference protocol, while resulting in a high level of solubilization also resulted in a high level of deamidation and lower level of peptide identifications. We therefore tried two approaches to maximize the amount of proteomic information obtained and therefore maximizing the amount of genetic information that could be obtained through detection of genetically variant peptides. The first was to use gentle chemistry in processing the sample. By using shorter trypsin digestion time, and lower temperatures, we observed less solubilization but we also saw much greater levels of peptide identifications. The other advance we tried was a dramatic increase in the level of reductants. By increasing DTT levels from the 25 mM level up to the 100 mM level we observed a dramatic increase in GVP detection. Together the sensitivity (TP/(TP+FN)) of GVP detection increased from 11.4% to 33.6%, an almost 3 fold increase[1]. The total identified GVPs increased from 45 to 127 for the optimized processing method, the number of unique peptides from 1585 ± 162 to 2703 ± 230 ($p = 5 \times 10^{-13}$) and the average number of genetically variant peptides detected increased from 20 ± 5 to 73 ± 5 ($p = 1 \times 10^{-13}$). The RMP increased from a maximum of 1 in 1380 and a median value of 1 in 24 for the original processing method to up to 1 in 624 million from a single hair with a median value of 1 in 1.1 million after chemical processing optimization ($p = 4 \times 10^{-7}$). 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}. These increases in RMP also meant that the likelihood of ancestral background increased as well. A likelihood ratio (LR) defined as the RMP calculated from the African population divided by the RMP calculated from the European population. With optimization and increased GVP detection, the likelihood ratio for European samples (eq. 2) 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*.

2.5 Task 2. Characterization of GVPs and optimization of mass spectrometry data acquisition. This research is from a manuscript about to be submitted to Journal of Proteome Research. A total of 497 genetically variant peptides have been discovered in this project. This format is not appropriate for a full cataloging of the genetic and chemical properties of each genetically variant peptide, an information of about 80% of identified GVPs can be located in the supplemental section of [1]. A selected panel of 24 GVPs were synthesized with stable isotope labelling in order to allow for spiking of the sample into the sample. This was then used as in internal standard to evaluate the efficacy of different mass spectrometry platforms. The characterization of this panel's performance (**Figure 1**) demonstrated that targeted mass spectrometry acquisition methods increased detection by roughly 1.3 to 1.6 fold depending on the mode of data acquisition. This is over and above any improvement made using the chemical treatments optimized in Task 1. The effect of these improvement son estimated random match probabilities is dramatic. *An overall 1.6-fold increase in GVP detection would result in 120 GVPs being regularly being detected and an estimated median RMP of 1 in 10^{13}.*

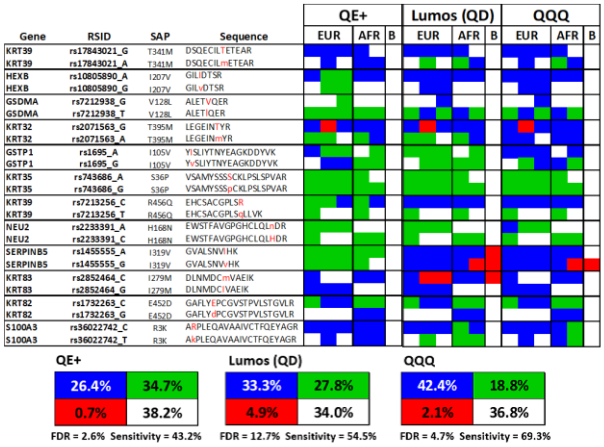


Figure 1. GVP matrix evaluating three analytical methods. Each row is a variant peptide and each column is an accumulated GVP profile from three replicates. QE+, data dependent acquisition on Q Exactive+; Lumos (QD), QuanDirect on Fusion Lumos; QQQ, multiple reaction monitoring on Agilent 6495. There are three European subjects (EUR) and two African subjects (AFR) analyzed in triplicate. Performance evaluation consists of true positives (TP, blue), false negatives (FN, green), false positives (FP, red), and true negatives (TN, white). The false discovery rate (FDR, FDR=TP/(TP+FP)) and sensitivity (Sensitivity = (TP/(TP+FN)) values are annotated.

2.5 ***Task 3 Estimation of Likelihood Ratio Values of Genetic Background.***

As described above the Likelihood ratio for ancestral background follows the following formula: $LR = \frac{Pr(GVP_{profile}|population_1)}{Pr(GVP_{profile}|population_2)}$ where probability of the combined profile of detected genetically variant peptides is calculated for one reference population and divided using the same equation but calculated using population distribution values from a second population. Previous work showed that European and African Populations LR values ranged from 10^4 for African samples to 10^{-4} for European Samples, 8 orders of magnitude. In order to estimate these values from the hair donated in this project, GVP-inferred SNP alleles also require additional levels of characterization at the genetic level. This includes an analysis of F_{ST} values,

Homozygosity, Hardy Weinberg equilibrium and linkage disequilibrium. These parameters provide constraints on proteomic genotyping and use of GVP-inferred SNP alleles to calculate the product rule. As described in Figure 2 there are 8 GVP-inferred SNPs that have FST values above 0.2 and are potentially ancestrally informative. This raises the possibility that some GVPs may be used specifically for ancestral informativeness[30]. These properties help explain some of the dramatic shifts in LR values observed in our data. With enough GVP detections cumulative differences in genotype frequencies will amplify and result in functionally different random match probabilities depending on which major reference population group is used. We showed earlier that the GVP profiles from the hair of African subjects were less frequent in the European population relative to the African population and profiles from European subjects were less frequent in the African population relative to the European population[1, 7]. The effect was significant with LR values spanning 8 orders of magnitude.

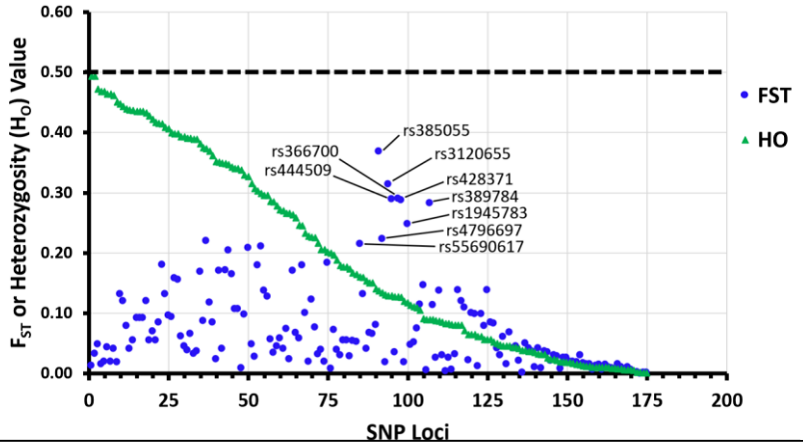


Figure 2. Scatterplot of F_{ST} and observed heterozygosity (Ho). These values are from GVPs that have been observed in our data. Potential ancestry informative markers are labeled.

A total of 170 samples were processed, from 5 population groups: African (AFR, n = 47), Admixed or African American (ADM, n = 21), European (EUR, n = 79), Middle Eastern (ME, n = 12) and Asian (EAS, n = 8). Samples were processed using the optimized hair sampling method described in Task 1 and subjected to shotgun proteomics on a Q Exactive™ Plus Hybrid Quadrupole-Orbitrap™ Mass Spectrometer[31]. The resulting datasets were screened for GVPs using the algorithm GVPFinder (parkerlab.ucdavis.edu) and Random Match Probabilities were calculated using the method described above in the introduction, Task 1 and in Goecker *et al*.[1].

2.6 *Task 4 Prediction of Ancestry as a Function of LR (EUR/AFR) Values*

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. 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.

When LRs were calculated values ranged over 22.5 orders of magnitude, from 10^{-14} to 10^{8} (**Figure 3A**). When absolute LR was plotted as a function of data quality, or in this case as number of detected GVPs, there was a distinct cut-off in data quality of 60 GVPs (**Figure 3B**). The remaining LR values were ranked (ranging from 0 to 1) from each population and plotted (**Figure 3C**). Based off these figures we can make 3 observations: 1) There is no effective difference between the European and Asian and Middle Eastern populations, 2) As can be seen there is a wide range of LR values in different populations and there is considerable overlap, categorical statements about ancestry, at for most subjects, is possible. 3) In spite of this the LR values for each population, African, Admixed, and European, were quite distinctive. Other descriptions about ancestry are therefore possible. *Below a Log_{10}(LR(EUR/AFR)) value of -9 the ancestry of the hair shaft will be 100% African, and above a value of 7 the sample will have a European donor.* These criteria are only met in 18.2% of African samples, and 15.9% of European samples, assuming the that more than 60 detected GVPs could be detected in the hair shaft sample. If the additional assumption is made of an equal prior probability of detecting an African, Admixed or European hair shaft, these conditions would be met a total of 11.4% of the time. It is more useful therefore to describe the probability of ancestry for each range of Log_{10}(LR) values. To reach that point the proportion of each population was plotted as a function binned ranges of

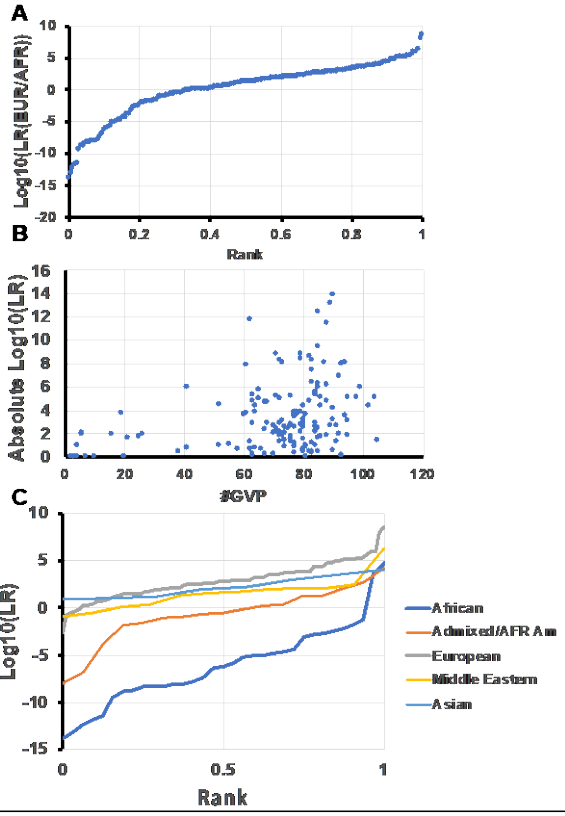
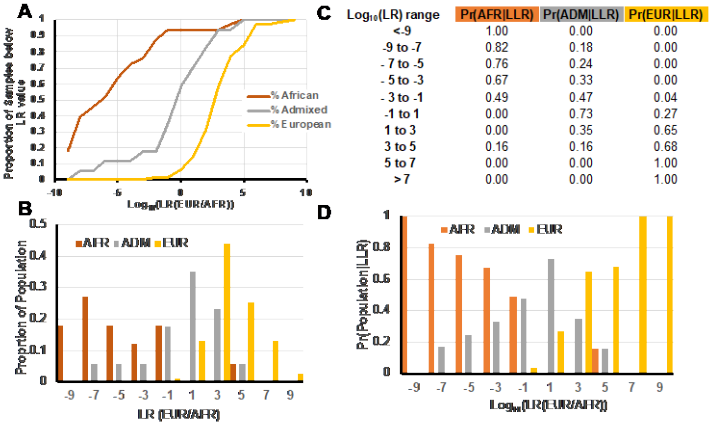


Figure 3. Likelihood Ratios of Hair Samples from subjects from different genetic backgrounds.
A) A total of 170 hair samples were processed, from 5 population groups: African (AFR, n = 47), Admixed or African American (ADM, n = 21), European (EUR, n = 79), Middle Eastern (ME, n = 12) and Asian (EAS, n = 8). Likelihood ratios were estimated as the quotient of the RMP values from the EUR relative to the AFR population and ranked from lowest to highest value. **B)** In order to remove biases from poor quality datasets a plot of absolute LR values was plotted as a function of total GVPs detected in the sample. **C)** After removal of poor-quality datasets (#GVP<60), LR values were replotted as a function of relative rank (0 to 1) for each major population.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. 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.

Log_{10}(LR) values (**Figure 4B**). *These empirical proportions provide a resource for addressing the central research question of this project: determining the probability of a given ancestry for a given likelihood ratio.*



Log_{10}(LR) range	Pr(AFR|LLR)	Pr(ADM|LLR)	Pr(EUR|LLR)
<-9	1.00	0.00	0.00
-9 to -7	0.82	0.18	0.00
- 7 to -5	0.76	0.24	0.00
- 5 to -3	0.67	0.33	0.00
- 3 to -1	0.49	0.47	0.04
-1 to 1	0.00	0.73	0.27
1 to 3	0.00	0.35	0.65
3 to 5	0.16	0.16	0.68
5 to 7	0.00	0.00	1.00
>7	0.00	0.00	1.00

Figure 4. Population distribution of Hair Shaft Derived Log_{10}(LR) Values.
Likelihood ratios were calculated from detected GVP profiles in hair shafts from African (dark orange), Admixed (grey) and European (dark yelloe) subjects. **A)** The likelihood ratios were transformed logarithmically and the proportion of each population that included a given Log_{10}(LR(EUR/AFR)) value or less was plotted as a function of the Log(LR) value. **B)** Log_{10}(LR(EUR/AFR)) values were binned corresponding to two orders of magnitude and the proportion of each population that obtained a measure of each range was measured for each population. Based of the distribution of each population that resulted in a given range of Log_{10}(LR(EUR|AFR)) values, and assuming equal prior probability that an unknown hair shaft could come from either population, the probability of each ancestry for a given range of Log_{10}(LR(EUR/AFR)) (Pr(population|LLR)) values for African (AFR, orange), admixed (ADM, grey) and European (EUR, yellow) subjects was calculated and tabulated **(C)** and plotted **(D)**.

Based off these calculations a GVP-derived Likelihood ratio above 5 orders of magnitude would be 100% likely to be European and between 1 and 5 orders of Magnitude (10 to 100,000) would be 65 to 68% likely to be European. An LR value between 1 x 10^{-1} to 1 x 10^{-5}, would have a probability of 49%, 47% and 4% of being African, Admixed or European respectively. Combining Admixed and African probabilities would result in a 96% probability. Only African and admixed hair shafts can account for an LR value below 1 x 10^{-5}.

2.6 *Additional Validation of Proteomic Genotyping*

SWGDAM guideline stipulate that development of new methodologies also need to be tested on 'real-world' scenarios. In terms of evidentiary analysis of hair shafts located from a crime scene, we anticipated that there are three scenarios that could test the limits of hair shaft analysis: hair from different body sites, greying hair, chemically processed hair (in this case hydrogen peroxide) and hair stored for years and decades at room temperature. In each of these scenarios the central question is whether the resulting GVP profiles would have a systematic bias as a function of potential differences that might occur in a sample. Would the presumed proteomic differences in protein profiling result in a change in GVP profiling. With the exception of hydrogen peroxide treatment all of these studies have been published[2, 3, 5]. In all three published scenarios there was no intrinsic bias imposed by changes in the proteome resulting from differing conditions.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. 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.

2.7 ***Basic Findings*****.** Based on the data presented or otherwise published, we can make the following conclusions. Hair shaft processing methos are better when large amounts of reductant are used and gentle chemical methods are applied, we saw a three-fold improvement in data yields from a single hair shaft. We discovered and characterized 497 GVPs. A description of the physical and chemical properties of many of these is available in the supplemental section of [1]. The genetic characterization of these GVP-inferred SNPs includes those which have significant differences in their FST values and are potentially ancestry informative SNPs. When 170 samples from a range of ancestries were processed and likelihood ratio of genetic background values calculated, values ranged over 22.5 orders of magnitude. When these values were plotted as a function of population there was distinctive patterns as a function of population with admixed samples distributing between European and African samples. When population numbers were binned and then plotted as a function of population proportion as a function of LR value, we were able to then calculate the probability of a given population as a function of LR assuming an equal distribution of African, admixed or European samples. We can state that LR values below 10^{-5} are categorically admixed or African individuals, LR values below 1 in 10^{-7} are categorically African, and LR values above 10^{5} are European. Between these extremes, were most samples are likely to be, Pr(population|LR) values can be calculated and tabulated, for example an LR value between 10^{1} and 10^{5} would be 65% to 68% likely to be European, assuming equal prior likelihood of obtaining a sample from either ancestral background.

3. Implications for criminal justice policy and practice in the United States.

Hair is a ubiquitous forensic substrate that is currently highly underutilized. This project lays the groundwork for efficient processing of hair and the ability to obtain high levels of genetic information that are useful to an investigator. We now routinely obtain random match probabilities above 1 in 100 million and demonstrate the potential for routine values of 1 in 10^{13}. This project demonstrates that information from a forensically relevant hair shaft can now include probabilities of a donor's ancestry for a given likelihood ratio. These methods are rigorous, scientifically and statistically based and as such meet the criteria stipulated in the 2009 NRC report: Strengthening Forensic Science in the United States.

References

[1] Z.C. Goecker, M.R. Salemi, N. Karim, B.S. Phinney, R.H. Rice, G.J. Parker, Optimal Processing for Proteomic Genotyping of Single Human Hairs, Forensic Science International: Genetics 47(102314) (2020). https://doi.org/https://doi.org/10.1016/j.fsigen.2020.102314
[2] T.J. Plott, N. Karim, B.P. Durbin-Johnson, D.P. Swift, R.S. Youngquist, M. Salemi, B.S. Phinney, D.M. Rocke, M.G. Davis, G.J. Parker, R.H. Rice, Age-Related Changes in Hair Shaft Protein Profiling and Genetically Variant Peptides, Forensic Science International: Genetics in press (2020).
[3] 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 science international 310 (2020) 110200. https://doi.org/10.1016/j.forsciint.2020.110200
[4] 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, Forensic science international. Genetics 42 (2019) 21-30. https://doi.org/10.1016/j.fsigen.2019.05.005
[5] 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, Forensic science international. Genetics 41 (2019) 19-23. https://doi.org/https://doi.org/10.1016/j.fsigen.2019.03.009
[6] 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). https://doi.org/10.1002/pmic.201600462
[7] 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, D.J. Fairbanks, A.S. Wilson, R.H. Rice, S.R. Woodward, B. Bothner, B.R. Hart, M. Leppert, Demonstration of Protein-Based Human Identification Using the Hair Shaft Proteome, PLoS One 11(9) (2016) e0160653. https://doi.org/10.1371/journal.pone.0160653
[8] G. Parker, Z. Goecker, R. Franklin, B. Durbin-Johnson, J. Milan, N. Karim, C. De Leon, A. Matzoll, T. Borja, R. Rice, Proteomic genotyping: Using mass spectrometry to infer SNP genotypes in a forensic context, Forensic Science International: Genetics Supplement Series 7 (2019)? ?????664-666.
[9] D.W. Deedrick, S.L. Koch, Microscopy of Hair Part 1: A Practical Guide and Manual for Human Hairs, Forensic Science Communications 6(1) (2004).
[10] D.W. Deedrick, Hairs, Fibers, Crime, and Evidence: Part 1 Hair Evidence, Forensic Science Communications 2(3) (2000).
[11] A.S. Wilson, Taphonomic Alterations to Hair and Nail, in: E.M.J. Schotsmans, N. Márquez-Grant, S.L. Forbes (Eds.), Taphonomy of Human Remains: Forensic Analysis of the Dead and the Depositional Environment, John Wiley & Sons Ltd2017, pp. 81 - 91.
[12] A.S. Wilson, The decomposition of hair in the buried body environment., in: M. Tibbett, D. Carter (Eds.), Soil Analysis in Forensic Taphonomy: Chemical and Biological Effects of Buried Human Remains, CRC Press, Boca Raton, 2008, pp. p 123-151.
[13] P.A. Coulombe, M.B. Omary, 'Hard' and 'soft' principles defining the structure, function and regulation of keratin intermediate filaments, Curr Opin Cell Biol 14(1) (2002) 110-22. https://doi.org/S0955067401003015 [pii]
[14] F.B.o. Investigation, FBI/DOJ Microscopic Hair Comparison Analysis Review, Department of Justice, fbi.gov, 2014.
[15] D. McNevin, L. Wilson-Wilde, J. Robertson, J. Kyd, C. Lennard, Short tandem repeat (STR) genotyping of keratinised hair. Part 1. Review of current status and knowledge gaps, Forensic science international 153(2-3) (2005) 237-46. https://doi.org/10.1016/j.forsciint.2005.05.006
[16] M.D. Brandhagen, O. Loreille, J.A. Irwin, Fragmented Nuclear DNA is the Predominant Genetic Material in Human Hair Shafts, Genes 9(640) (2018) 1-21.
[17] J. Robertson, Managing the forensic examination of human hairs in contemporary forensic practice, Australian Journal of Forensic Sciences 49(3) (2017) 239-260. https://doi.org/DOI: 10.1080/00450618.2017.1279838
[18] President's Council of Advisors on Science and Technology, Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods, in: W.H.O.o.S.a.T. Policy (Ed.) Washington DC, 2016.

[19] The National Research Council, Strengthening Forensic Science in the United States: A Path Forward The National Academy Press, Washington D.C., 2009.
[20] Genomes Project Consortium, A. Auton, L.D. Brooks, R.M. Durbin, E.P. Garrison, H.M. Kang, J.O. Korbel, J.L. Marchini, S. McCarthy, G.A. McVean, G.R. Abecasis, A global reference for human genetic variation, Nature 526(7571) (2015) 68-74. https://doi.org/10.1038/nature15393
[21] G.R. Abecasis, A. Auton, L.D. Brooks, M.A. DePristo, R.M. Durbin, R.E. Handsaker, H.M. Kang, G.T. Marth, G.A. McVean, An integrated map of genetic variation from 1,092 human genomes, Nature 491(7422) (2012) 56-65. https://doi.org/10.1038/nature11632
[22] 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, Forensic Science International: Genetics 43 (2019) 102148. https://doi.org/https://doi.org/10.1016/j.fsigen.2019.102148
[23] K.E. Mason, D. Anex, T. Grey, B. Hart, G. Parker, Protein-based forensic identification using genetically variant peptides in human bone, Forensic science international 288 (2018) 89 - 96.
[24] F. Chu, K.E. Mason, D.S. Anex, A.D. Jones, B.R. Hart, Proteomic Characterization of Damaged Single Hairs Recovered after an Explosion for Protein-Based Human Identification, Journal of proteome research (2020). https://doi.org/10.1021/acs.jproteome.0c00102
[25] A.E. Woerner, F.C. Hewitt, M.W. Gardner, M.A. Freitas, K.Q. Schulte, D.S. LeSassier, M. Baniasad, A.J. Reed, M.E. Powals, A.R. Smith, N.C. Albright, B.C. Ludolph, L. Zhang, L.W. Allen, K. Weber, B. Budowle, An algorithm for random match probability calculation from peptide sequences, Forensic science international. Genetics 47 (2020) 102295. https://doi.org/10.1016/j.fsigen.2020.102295
[26] B. Budowle, A. van Daal, Forensically relevant SNP classes, BioTechniques 44(5) (2008) 603-8, 610. https://doi.org/10.2144/000112806
[27] B.F. Algee-Hewitt, M.D. Edge, J. Kim, J.Z. Li, N.A. Rosenberg, Individual identifiability predicts population identifiability in forensic microsatellite markers, Current Biology 26(7) (2016) 935-942.
[28] SWGDAM, SWGDAM Validation Guidelines for DNA Analysis Methods, in: F.D.o. Justice (Ed.) FBI, Washington DC, 2016.
[29] Y.J. Lee, R.H. Rice, Y.M. Lee, Proteome analysis of human hair shaft: from protein identification to posttranslational modification, Molecular & cellular proteomics : MCP 5(5) (2006) 789-800.
[30] J.R. Kidd, F.R. Friedlaender, W.C. Speed, A.J. Pakstis, F.M. De La Vega, K.K. Kidd, Analyses of a set of 128 ancestry informative single-nucleotide polymorphisms in a global set of 119 population samples, Investigative genetics 2(1) (2011) 1. https://doi.org/10.1186/2041-2223-2-1
[31] N. Nagaraj, N.A. Kulak, J. Cox, N. Neuhauser, K. Mayr, O. Hoerning, O. Vorm, M. Mann, System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top Orbitrap, Molecular & cellular proteomics : MCP 11(3) (2012) M111 013722. https://doi.org/10.1074/mcp.M111.013722