BIOGEOGRAPHIC CLASSIFICATION OF EUROPEAN AND AFRICAN HAIR USING GENETICALLY VARIANT PEPTIDES

Zachary C Goecker[1], MPS, Bailey M Wills[2], MS, Michelle R Salemi[3], MS, Brett S Phinney[3], Susan Walsh[2], PhD, Glendon J Parker[1], PhD
[1]Department of Environmental Toxicology, University of California - Davis
[2]Department of Biology, Indiana University - Purdue University Indianapolis
[3]Proteomic Core Facility, Genome Center, University of California - Davis

This research aims to classify hair by biogeographic background from European, African, and admixed individuals using a profile of inferred SNP genotypes obtained by the detection of genetically variant peptides (GVPs). GVPs are the result of non-synonymous SNPs that cause a change in the amino acid sequence of translated proteins. Proteomic LC-MS/MS orbitrap methods allow for the detection of these small mass shifts in tryptic peptides. The SNPs chosen here have differences in genotypic frequency between ancestral groups. This difference can be used to estimate the biogeographic group from which the hair originated. This research strengthens the forensic value of hair shafts by providing objective and standardized protocols for estimating biogeographic background using human hair proteomic genotypes. This adds an additional mode of genetic information to complement forensic analysis of mtDNA haplotypes and anatomical features.

A sample set of 85 African, European, and mixed-descent hair has been used to assess both the ability to target and identify GVPs, and also confirm the presence of the amino acid changes via DNA genotyping. Confirming this expectation is not a simple task, since most SNPs in the hair proteome are similar in allelic frequency between African and European groups. To address this concern, we have identified 113 candidate SNPs in the hair proteome that have at least a 10% difference in minor allele frequency, as obtained from continental regions from the 1000 Genomes Project, between African and European populations. Assessment of these candidates indicate that the difference between the median African likelihood ratio and European likelihood ratio is nine orders of magnitude, with 1 individual exhibiting an incorrect binary classification. We have also identified a subset of ancestry informative markers (AIMs) that correspond to GVPs that may assist in distinguishing between African versus Non-African populations. These AIMs have been filtered using Hardy-Weinberg equilibrium, linkage disequilibrium and F_{st} values. Additional samples are currently being assessed for these candidates. Future developments include statistical analyses such as likelihood ratio and orthogonal transformation methods, where we will demonstrate a developing method used to distinguish between European and African populations based on GVP profiling.