## H115 Identification of Non-Synonymous SNPs in Archaeological Hair Protein: Calculation of Measures of Identity and Biogeographic Background

*Jonathan Hilmer, PhD, Montana State University, Dept of Chemistry and Biochemistry, Bozeman, MT 59715; Katie Giddons, BSc, Montana State University, Dept of Chemistry and Biochemisry, Bozeman, MT 59715; Tami Leppert, MSc, University of Utah, Dept of Human Genetics, Salt Lake City, UT 84112; Brian Bothner, PhD, Montana State University, Dept of Chemistry and Biochemistry, Bozeman, MT 59715; Mark Leppert, University of Utah, Dept of Human Genetics, Salt Lake City, UT 84112; Andrew Wilson, PhD, University of Bradford, School of Life Sciences, Bradford, UNITED KINGDOM; and Glendon Parker, PhD*, MS 179, 800 W University Parkway, Orem, UT 84103*

After attending this presentation, attendees will understand how proteomic datasets can be used to obtain forensic and genetic information such as quantifiable measures of identity and biogeographic background.

This presentation will impact the forensic science community by highlighting novel methods that extend the amount of genetic information which can be obtained forensically and anthropologically in the absence of usable DNA, either due to contamination or degradation.

Methodology has been developed to extract identifying genetic information from proteomic datasets. DNA typing has revolutionized forensic practice and jurisprudence; however, DNA often is degraded due to biological, chemical, or environmental factors. Protein is considerably more stable and more abundant than DNA and persists in the environment for a longer period. Protein also contains genetic information in its primary structure, the result of non-synonymous SNPs (nsSNPs) that manifest as Single Amino-Acid Polymorphisms (SAPs). These SAPs-containing peptides are accessible to shotgun tandem mass spectrometry. This study has identified nsSNP-containing peptides from 35 alleles in 26 genes expressed in the forensically informative hair shaft proteome. Complex proteomic datasets from trypsin digests of the hair shafts of 54 validated European American individuals were obtained for this study. Peptides corresponding to nsSNPs expressed in this protein population were identified and collated for each individual. The combined probability of each individual nsSNP profile was calculated using genotypic frequencies of each allelic combination in the European population (1,000 Genomes Project) and the "product-rule." The power of genetic discrimination ranged from 1 in 1.002 to 1 in 9,000. The average power of discrimination was 1 in 280. The power of discrimination increased as a function of proteomic dataset quality ($r^2$=0.624, n=58, p<0.0001). When the power of discrimination is calculated using genotypic frequencies from the African population, increased powers of discrimination are achieved. This is consistent with a decreased likelihood that the samples originate from an African origin. Relative likelihood measurements of European compared to African genetic origin range from 1 to 780 with an average of 50, a median of 18, and a standard deviation of 116. (n=64). Direct validation of the imputed status of each nsSNP allele was achieved with Sanger sequencing. A total of 430 genotype determinations were made from the proteomic data and 426 assignments were confirmed (specificity=99.1%, FPR=0.93%). The overall sensitivity was 31%. Framework has been established for the use of proteomic datasets as a source of identifying genetic information, allowing measures of identity and biogeographic background to be made from forensic or anthropological protein sources, including bone, teeth, preserved soft tissue, and trace evidence such as fingerprints.

**Hair, Proteomics, Bioarchaeology**