CONFIRMATION OF MUSCLE PROTEIN POLYMORPHISMS USEFUL FOR HUMAN IDENTIFICATION

Tatiana Perez[1], Emmanuel Chang[2], Heyi Yang[1], Erin Butler[1], Amy Mundorff[3], David Fenyö[4], Donald Siegel[1]
[1]Department of Forensic Biology, NYC Office of Chief Medical Examiner
[2]Department of Chemistry, York College, City University of New York
[3]Department of Anthropology, University of Tennessee
[4]Center for Health Informatics and Bioinformatics, New York University

The predominant method for identification of human remains is through genomic DNA (gDNA) analysis of short tandem repeats (STRs) or single nucleotide polymorphisms (SNPs). However, occasions arise when gDNA is not present in sufficient quantities or quality due to natural taphonomic decay, or accidental or intentional physical processes and/or chemical contamination (e.g. fire, lye, automotive/aviation fuel, etc.). Mitochondrial DNA (mtDNA) analysis is often the second choice for identification, but it too can be degraded and it does not have the same discriminatory power of gDNA. Nucleotide polymorphisms make g- and mtDNA useful for individual identification. Amino acid polymorphisms serve the same function in proteins. Proteins have the additional advantages that they are less labile than nucleic acids and can offer greater discriminatory powered than mtDNA.

To evaluate the use of protein polymorphisms for the identification of human remains, total protein from muscle samples of 14 descendants (Forensic Anthropology Center, University of Tennessee at Knoxville, IRB approved) were examined by mass spectrometry. An initial pilot study of five individuals identified hundreds of proteins. The equivalent protein sequences, including all splice variants, were retrieved from the RefSeq database to establish a custom *in silico* muscle protein polymorphism database for further mass spectrometric database searching. This custom database contained amino acid sequences representing all identified muscle proteins with all single amino acid polymorphisms (one polymorphism per sequence) extracted from the Go Exome database, a collection of high quality, deep sequenced exomes from over 6,500 individuals

Mass spectra from all individuals were searched against the created database for matches with polymorphic peptides with population frequencies greater than 0.05. A total of 13 polymorphisms were identified in eight different genes. DNA from all 14 individuals was sequenced at all 13 loci to confirm genotype status. Analysis, comparing gene population frequencies at each locus of each individual are currently underway to determine the power of the 13 loci for individual identification.

Supported in part by National Institute of Justice Grant 2014-DN-BX-K014.