Utilizing 300,000 or more Single Nucleotide Polymorphisms for the Identification of Individuals Within a Forensic Mixture

Kevin C. McElfresh[1], Ron Sosnowski[1], Christie Abbott[1], Kristin Stanford[1] and David Craig[2]. 1) Casework Genetics, Richmond VA, 2) Translational Genomics Research Institute, Phoenix AZ

In 20 years the amount of DNA required for a DNA test has decreased from micrograms to nanograms, the time required has decreased from weeks to hours, and the acceptance of the technology from skeptical to fundamental. However one of the remaining persistent problems is the interpretation of mixtures, that is, multiple individuals within a single sample. In part the problem is inherent in the limitations of current STR technology and in part it is due to the genetics of the systems. None-the-less, the issue remains that while utilizing length measures of DNA, no matter how well resolved, a person can be identified uniquely from a single source sample, yet that same individual cannot be anything more than generally included in a mixture of three or more individuals.

Studies investigating the genetic basis of common complex diseases, simultaneously genotype hundreds of thousands of Single Nucleotide Polymorphisms (SNPs) at low cost and with high reproducibility. These high-density SNP genotyping arrays are frequently used in genome-wide association studies (GWAs) and cover greater than 80% of common human genetic variation. With GWAs it is necessary to type many individuals in order to determine genetic associations with a complex disease. This is accomplished by pooling the DNA from multiple individuals and assaying them as a single sample, i.e. creating a mixture then typing the individuals within the sample. Our technology takes the analytical underpinnings of GWAs and applies them in a forensic context. In effect, a GWA study is a variation of a forensic mixture. In forensics the mixture is generated from unknown sources in unknown ratios rather than created for a specific research purpose from known samples. This has significant implications for the analysis of the results. We show that by applying GWAs based analysis methods to forensic samples it is possible to resolve individuals within complex mixtures, as well as address other challenging forensic sample types such as in highly degraded samples. Our technology utilizes the cumulative statistical power from genotyping hundreds of thousands of SNPs to analyze single source forensic samples as well as complex mixtures. The data show that it is possible to parse a mixture of a few individuals using as few as several thousand SNPs and mixtures of hundreds of individuals using 300,000 or more SNPs. Developmental validation of the high density SNP approach will be reviewed as part of the demonstration of the forensic application of the technology.

Work presented is proprietary and patent pending.