B97 Identification of Contributors in Complex DNA Mixtures Utilizing High-Density SNP Arrays: Influence of Sample Ancestry, Ancestry SNPs, and Reference Population

Brittin McMahon, MS, The Center for Advanced Forensic DNA Analysis, 2305 Executive Circle, Greenville, NC 27834; Thomas J. Cavanagh, MS, The Center for Advanced Forensic DNA Analysis, 2305 Executive Circle, Greenville, NC 27834; Mary L. Clair, BS, The Center for Advanced Forensic DNA Analysis, 2305 Executive Circle, Greenville, NC 27834; Jodi R. Bailey, MS, 2305-102 Executive Circle, Greenville, NC 27834; Lucy A. Davis, BHS, The Center for Advanced Forensic DNA Analysis, 2305-102 Executive Circle, Greenville, NC 27834; Christiana H. Shoopman, BA, 2305 Executive Circle, Greenville, NC 27834; and Sandra L. Close, PhD, The Center for Advanced Forensic DNA Analysis, 2305 Executive Circle, Greenville, NC 27834*

After attending this presentation, attendees will better understand the utilization of Single Nucleotide Polymorphism (SNP) genotyping to identify individual contributors to a complex mixture DNA sample and the effect of contributor ancestry, SNPs utilized during analyses, and reference population selection upon results interpretation.

This presentation will impact the forensic science community by demonstrating a practical alternative to Short Tandem Repeat (STR) analysis using capillary electrophoresis for evaluating mixed DNA samples and will also present the factors which could affect evaluation, such as reference population and ancestry of the contributor(s).

An increasing number of samples evaluated by forensic laboratories include DNA from multiple contributors, such as those routinely observed in sexual assaults and touch DNA. Advances in genotyping methodologies, routinely employed by clinical and research laboratories, provide alternatives to the use of capillary electrophoresis and STR testing. In a single assay, high density SNP Arrays can generate data from thousands to millions of genetic markers including ancestry, phenotype, X, Y, mitochondrial, and identity markers. This substantial increase in genetic data may be expected to improve the power with which contributors to a DNA mixture may be identified.

The use of SNPs to determine inclusion or exclusion of a contributor to a mixture involves comparison of the mixture with a reference population. Studies were conducted to evaluate the impact of sample ancestry, ancestry alleles used in the analyses, reference population selection on the determination of inclusion or exclusion, and on the statistics supporting the identification of contributors in sample sets utilizing the QSNP Informatics Software™. This software employs forensically relevant algorithms utilizing genetic markers to estimate the genetic distance between a sample of interest, an evidence sample, and a reference population as proposed in Homer.[1]

Two methods were utilized to generate multiple reference populations: Experimental and Bioinformatic. In the Experimental method, reference populations were created by analyzing human Coriell ancestry panel samples with the Illumina® Infinium® Assay and HumanOmniExpressExome-8 BeadChip array. Allele frequencies were developed from the cumulative results of approximately 100 genomes per Coriell ancestry panel samples (AA48 and Cau50). The Bioinformatics Methodology utilized publicly available data from the 1,000 Genomes Project (http://www.1000genomes.org). SNPs identified by the 1,000 Genomes Project were filtered to include only those also found on the Exome8 array. Frequencies of SNPs from approximately 1,500 samples from the African and European Ancestries (AFR and EUR) were utilized for comparative purposes.

Four reference populations were compared for their impact on calculated Distance (D) and T-statistic (T) values in matched and mismatched samples of Caucasian and African ancestry. Pairwise comparisons of a matching Caucasian DNA sample (Table 1) show that D and T values are positive, indicating a match. The values for matching Caucasian samples are numerically higher when calculated against reference populations of African ancestry than when calculated against European reference populations. The Bioinformatics-derived reference populations generally provide more positive values than their corresponding experimental reference populations.

Table 1. Summary Statistics for Matched and Mismatched Caucasian Samples

	AA48		Cau50		Pooled AA48 & Cau50		AFR		EUR		Pooled AFR & EUR	
	D	T	D	T	D	T	D	T	D	T	D	T
Mean Matched	0.195		0.150	815	0.168	889	0.260	924	0.208	878	0.226	960
St. Dev.	0.002	10	0.002	12	0.002	11	0.002	10	0.002	13	0.003	14

Mean Mismatched	0.01	37		-136		-68	0.021	51				-76
St. Dev.	0.005	16	0.003	17	0.004	20	0.008	19	0.006	20	0.007	22

The impact of these factors on inclusion and exclusion thresholds will be presented for both major and minor contributors in simulated forensic case scenarios, using both Experimental and Bioinformatics reference populations.

Reference:

1. Homer N, Szelinger S, Redman M, Duggan D, Tembe W, et al. (2008) Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays. *PLoS Genet* 4(8): e1000167. doi:10.1371/journal.pgen.1000167

SNPs, DNA Mixtures, Population Genetics