A124 SNPs, Forensic DNA Mixtures, and Population Genetics

Kevin C. McElfresh, PhD, 13580 Groupe Dr., Ste 301, Woodbridge, VA 22192*

After attending this presentation, attendees will have a better understanding of analyzing forensic DNA samples containing mixtures
of DNAs.

This presentation will impact the forensic science community by providing advanced analytical capabilities for DNA mixtures that cannot be analyzed using STRs.

The essence of forensic DNA analysis is the comparison of known differences in the human DNA of one sample with the differences of an unknown sample. Match or non-match is based on the sum of the differences between the two samples. If a difference is characterized in binary code of one for complete difference and 0 for no difference, then in a comparison of single source forensic DNA samples, if that sum is zero, the samples match, but if that sum is non-zero the samples do not. The problem lies in the comparison of mixtures of DNAs. In this case, there is the distinct possibility that there is an inclusion hidden beneath the non-zero answer. In the case of STRs, mixtures beyond two individuals virtually negate the ability to attempt an analysis of the zero answer within the non-zero aggregate. Ultra-High Density Single Nucleotide Polymorphism (UHD-SNP) arrays analyzing >5 million loci across the human genome provide an alternative forensic DNA mixture analysis capability due to the smaller number of alleles per locus that are compensated for by the larger number of loci. Thus, the smaller numbers of alleles (two) at an SNP locus allows for the assessment and aggregation of small per locus differences over a large data set and determine the effect of the zero-based inclusion within a non-zero aggregate. The objective of this project was to determine the fundamental effects of reference population and number of SNP loci necessary to correctly assess the contributors to a forensic DNA mixture.

Computer simulations of SNP data done by Homer and Jacobs, using different methods of calculation, both addressed the possible analytical capabilities for mixtures of DNAs within a medical and forensic context (Homer) and in a medical Genome Wide Association context (Jacobs).[1,2] In the computer simulations, the percent contribution to the mixture was simulated and tested using prepared mixtures. In addition to the mixture proportion, there was a clear component of reference population to the analysis. This study has examined those calculations using samples from forensic mixtures utilizing samples from typical forensic crime scenes, e.g., blood/semen mixtures and also complex mixtures of more than two individuals.

A baseline of fundamental characteristics of SNP results was developed using single source match and non-match DNAs taken from blood and semen. These included serial dilution of DNAs and subsequent comparison of the number of loci and the effect of the population analysis on the results. From these studies a clear baseline of Loss of Detection (LOD) and sample analysis criteria were drawn. Reference populations were built for Caucasian, African, Hispanic and Asian populations. The population specific results were then compared using the Illumina provided SNP cluster file and population specific cluster files using these criteria. The same analysis was then done using mixtures of samples, again taken from forensically relevant sample types. Comparisons of numbers of loci and the effect of the reference populations were analyzed within this context. Specifically, the analysis was geared toward the effect of partial results on the ability to correctly analyze inclusion versus exclusion and also to assess the reference population effects on those answers. As expected, the number of loci, either as a function of sample quality or as a function of DNA concentration, had the largest effect on the ability to analyze the results while the population effects were secondary. In either case, utilization of more than approximately 50,000 SNP loci overcame the localized effects of population structure. The results were then used to determine a set of analytical parameters for the interpretation of UHD-SNP forensic DNA results.

References:

1. Homer N, Szelinger S, Redman M, Duggan D, Tembe W, *et al*. Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. PLoS Genet 2008;4(8):e1000167. doi:10.1371/journal.pgen.1000167
2. Jacobs K, Yeager M, et.al. (2009) A new statistic and its power to infer membership in a genome-wide association study using genotype frequencies Nature Genetics doi:10.1038/ng.455

SNP, Mixture, Population Genetics

 ** Presenting Author*