ADVANCEC FORENSIC SAMPLE ANALYSIS UTILIZING GENOME WIDE IDENTIFICATION: COMPLEX MIXTURES AND COPY NUMBER VARIANTS

Kevin C. McElfresh, Nicole Unger, Mary Clair, Kristin Stanford, Ronald Sosnowski
Casework Genetics, 13580 Groupe Dr., Suite 301, Woodbridge, Virginia 22192

Ultra High Density Single Nucleotide Polymorphism arrays (UHDSNP) routinely used in Genome Wide Association Studies (GWAS) are now capable of analyzing from 1 million to over 5 million loci. The net effect of this analytical capability is the sequencing of the molecular differences in the human genome thus providing direct assay of the genetic differences between samples. GWAS analyze the SNPs on the autosomes, the SNPs on the X, Y and mitochondrial genomes, as well as the Copy Number Variant (CNV) loci, all in a single reaction and scan. The biggest advantage in adapting GWAS to Genome Wide Identification (GWID™) is the analysis of complex mixtures. In single-source match vs. non-match, GWID over analyzes the samples since the molecular differences are distinct. However, as the quality of the sample and/or the amount of sample decreases, or the number of contributors increases, or both; the molecular complexity quickly outstrips the analytical capabilities of systems in which the results are based on secondary measures of actual molecular differences (*e.g.* capillary electrophoresis based sieving methods, *i.e.* STRs). The UHDSNP system in use at Casework Genetics (Illumina Omni1-Quad): assays 27 SNPs in the mtDNA genome, 139 CNV loci and 2,184 SNP loci on the Y chromosome, as well as 2,738 CNV loci and 24,756 loci unique to the X chromosome in addition to > 900,000 autosomal loci generating >1.1 million sequence points. The principle metric of GWID mixture analysis is the sum of the B allele frequency differences between the samples compared, providing a measurement of genetic similarity or dissimilarity between samples. The smaller the difference in the B allele frequencies, the more genetically similar the samples are. DNA from a series of test samples was initially analyzed as single source material using GWID. DNA from these samples was then used to create mixtures of 2 or more individuals, analyzed using autosomal GWID, then re-analyzed using the autosomal and Y chromosome CNV loci. Scatter plots of the mixed sample data correlate with the results of the autosomal data analysis and provide additional insight into the analysis of complex mixtures. In this study, the samples for mixtures were also reanalyzed and compared to possible contributors from other studies by reassembling the manifest files and initiating a new analysis file. Table 1 provides representative data. Note that a 2 person mixture can be included or excluded as being a subset of a 3 person mixture.

Comparison	**Σ *f* B allele Differences**
Buccal Swab : Semen (single source match)	16,804
Buccal Swab : Semen (single source exclusion)	255,148
Suspect : Victim + Semen Mixture (2 person mix exclusion)	177,806
Suspect : Victim + Semen Mixture (2 person mix inclusion)	39,149
2 Person Mixture : 3 Person Mixture (exclusion)	121,029
2 Person Mixture : 3 person Mixture (inclusion)	13,751

Table 1. B allele frequency differences between mixed and non-mixed samples.

The scatter plots of all CNV loci, with the Y chromosome cnv SNPs highlighted, mirror the results from Table 1, especially with regard to the analysis of mixtures within mixtures. Figure 1 and 2 are representative of this comparison. The red points are the CNV loci on the Y chromosome. These results will be discussed in light of their application to cases in which challenging samples were processed.

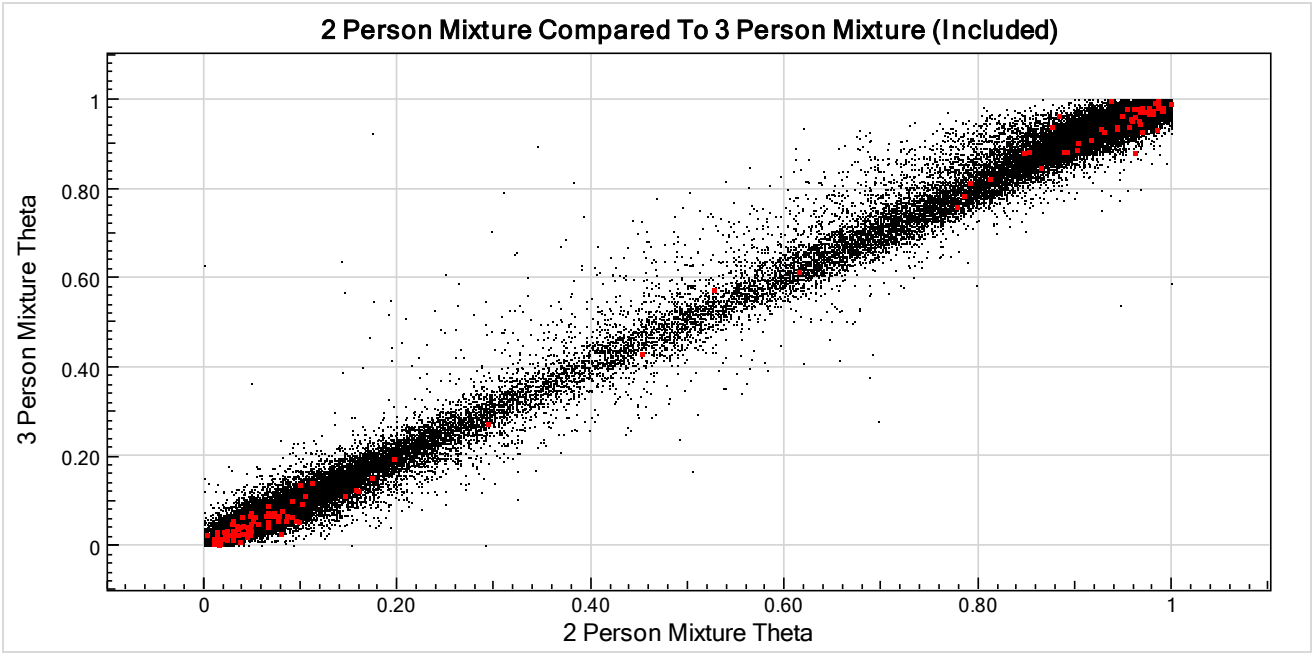

Figure 1. 2 Person Mixture Compared to 3 Person Mixture (Included).

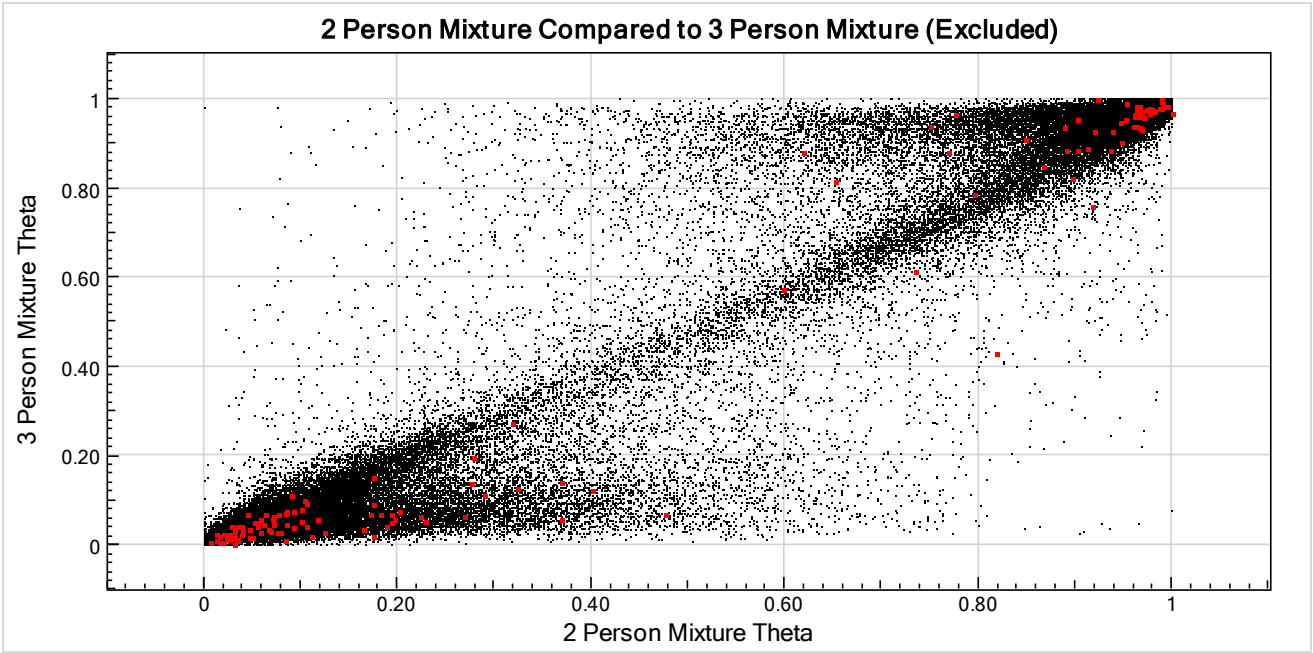

Figure 2. 2 Person Mixture Compared to 3 Person Mixture (Excluded).