Assessing the Reliability of SNP Micro-Array Data for Forensic Genealogy

Robert Bever, Ph.D., Jon Davoren M. Sc., Erin Sweeney, M. Sc, Teresa Vreeland, B.S., and Mike Cariola, MFS.
Bode Technology, 10430 Furnace Road Suite 107, Lorton, VA 22079

After attending the presentation, attendees will understand how the source of tissue, quantity of template DNA, and degree of degradation affects high-density Single Nucleotide Polymorphism (SNP) micro-array results. Additionally, the audience will learn how these variables affect the ability to identify related individuals through searching of genealogical databases consisting of high-density SNP genotypes.

The data presented will provide forensic scientists with basic information regarding the effect of DNA quality and quantity on the accuracy and call rate of high-density SNP genotype profiles. The data presented can be used to determine whether DNA extracted from forensic samples will be suitable for SNP micro-array processing used in forensic genealogical applications.

Forensic Genealogy (FG) is a powerful investigative technique to aid in the identification of victims and suspects of unsolved crimes. By establishing the capabilities and limitations of FG, the community will be able to address the inevitable court challenges and more effectively use this technique to provide investigative leads associated with unsolved murders, sexual assaults and missing person cases.

Forensic Genealogy is a three-step process. Step 1 is the extraction and quantitation of DNA from forensic evidence or human remains. Step 2 is the generation of SNP genotypes using either high-density SNP micro-arrays or whole genome sequencing. These steps serve as the genetic foundation of subsequent downstream genealogical investigation. The final step is comparing the SNP data to genealogical databases by a professional genealogist.

In this study, the quality of high-density SNP micro-array data using varying amounts of DNA from blood and post-coital samples was evaluated by sending the same samples to three laboratories for either SNP chip or whole genome analysis. DNA extracts (2 to 200 ng) purified from blood or post coital samples were analyzed using the Illumina Infinium Global Screening Array, the Infinium CytoSNP-850K Bead chip, and Whole Genome Sequencing based on the Illumina NovaSeq™ technology. The SNP micro-array genotypes and sequence results were evaluated on call rates, accuracy (as compared to the whole genome sequence), and reproducibility within and between chips.

DNA extracted from blood and analyzed on the CytoSNP or the GSA chip had SNP call rates that ranged from 88.9% (2ng) to 99.78% (200 ng). SNP call rates associated with post coital DNA and analyzed with the GSA chip ranged from 50% (2 ng) to 76% (50 ng), whereas the same DNA processed on the CytoSNP array had call rates from 64% (2 ng) to 91% (50 ng). Intra-chip reproducibility was greater than 99% for samples with > 10 ng and as low as 97.8% for 2 ng samples. Only 104,385 SNPs were common to both chips; concordant SNP genotypes between the two chips at these SNPS ranged from 82% to 89% with blood or post coital DNA samples that had at least a 70% call rate.

When SNP genotype calls were compared to Whole Genome Sequencing data, the degree of concordance varied based on the amount of DNA, the degree of degradation, and the stringency level of the GENCALL Score using the Illumina Genome Studio software. All the non-degraded DNA from blood and the post-coital samples with at least 10 Ng of DNA could be used to reliably match samples entered into genealogical databases. Whereas, highly degraded DNA with low template amount could not match samples in the genealogical databases.

Key words: Forensic Genealogy, SNP Micro-array, Degraded DNA.