## B75 Massively Parallel DNA Sequencing Applications for Forensic Mixture Analysis

*Christine H. Baker, MS*, Battelle Memorial Institute, 1001 Research Park Boulevard, Town Center Two, Ste 400, Charlottesville, VA 22911; Jocelyn M. Bush, MS, Battelle Memorial Institute, 505 King Avenue, Columbus, OH 43201; Lindsay Catlin, MS, Battelle Memorial Institute, 505 King Avenue, Columbus, OH 43201; Esley M. Heizer, Jr., PhD, Battelle Memorial Institute, 505 King Avenue, Columbus, OH 43201; Adriana N. Swatzell, 1400 Lascassass Pike, Apt H45, Murfreesboro, TN 37132; Richard A. Guerrieri, MS, Battelle, 1 Corduroy Court, Stafford, VA 22554; Mark Wilson, PhD, Battelle, 2001 Jefferson Davis Highway, Ste 1203, Arlington, VA 22202; and Michael L. Dickens, PhD, Battelle, 505 King Avenue, Columbus, OH 43201*

The goal of this presentation is to characterize massively parallel DNA sequencing data for the analysis of mixture samples.

This presentation will impact the forensic science community by making a powerful case for the promise of Massively Parallel Sequencing (MPS) technology to enhance current capabilities for mixture deconvolution.

The introduction of MPS technology for forensic DNA analysis offers many advantages over legacy Capillary Electrophoresis (CE) typing due to the elucidation of sequence information in addition to fragment size, as well as the ability to multiplex many forensically relevant markers. The information gleaned from DNA sequence data relative to CE fragment size may provide increased discriminatory power among individuals, which could prove valuable for the deconvolution of mixture samples. Despite the clear theoretical advantages of MPS data for mixture analysis, mixture deconvolution software tools have not been developed to utilize the added information obtained from sequence data.

To characterize MPS data for mixtures samples, more than 100 artificial mixture samples were established, ranging from two to five contributors (male and female) in varying ratios to include minor contributor input at 3% of the total mixture quantity. Samples were generated using eight single-source contributors (four males and four females) previously characterized using CE and MPS technologies and selected for size- and sequence-based allele diversity. A dilution series of all single-source samples was created to establish profiles at various input levels from 500pg to 15pg. These samples were processed using the Promega® PowerSeq™ Auto/Y System (targeting 22 autosomal Short Tandem Repeats (STRs), 23 Y-chromosomal Short Tandem Repeats (Y-STRs), and amelogenin) across five sequencing runs on the Illumina® MiSeq®. Data were analyzed with Battelle ExactID® software. Results were further characterized using custom software applications to identify allele and stutter sequences and to recognize contributor profiles among the mixture data.

Data from the single-source samples were evaluated for genotype accuracy and patterns in stutter and drop out across the dilution series. Single-source samples generated with full DNA input (500pg) were accurate and complete relative to data generated with CE and other MPS platforms. Across the 8 contributors, 8 out of 46 loci exhibited at least one isoallele, which are identical by length but differ in sequence. Loci D2S1338 and D12S391 displayed the largest gains in resolution due to sequence differences among same-size alleles. Patterns in sample- and locus-specific drop out were characterized across the dilution series. Stutter patterns were largely consistent across dilution series and reflected similar patterns to those observed in CE data.

Mixture samples were queried for each contributor profile and were compared with expectations based on single-source sample performance. Minor contributors were detected even when providing as little 3% (approximately 15pg) of the template DNA in the mixture. The relative representation of each contributor in the MPS data was evaluated by comparing read counts for alleles present only in a single contributor, as well as by comparing read counts attributed to Y-chromosome versus autosomal STRs to compare male versus female contributions. By all measures, the relative read counts attributed to each contributor were highly consistent with the true proportions of each sample contributing to the mixtures.

The current data set provides novel insight into the value of MPS data for mixture analysis. The results support a growing body of evidence suggesting that MPS technology provides accurate and reliable data for forensic STR markers, with an improved power of discrimination for loci exhibiting isoalleles. Moreover, the current demonstration of the sensitivity of MPS data to capture minor contributors, together with the strong reflection of the true contributor ratios in the relative read counts, makes a powerful case for the promise of MPS technology to enhance current capabilities for mixture deconvolution.

**Massively Parallel Sequencing, Mixture Deconvolution, Forensic DNA Analysis**



*Presenting Author