CONSIDERATIONS FOR ASSESSING PRECISION OF SEQUENCE-GENERATED DATA: "FEASIBILITY AND GUIDANCE STUDY OF NEXT GENERATION SEQUENCING FOR FORENSIC DNA APPLICATIONS"

Christine H. Baker, Jocelyn M. Bush, Richard A. Guerrieri, Mark R. Wilson, Angela Minard-Smith and Michael L. Dickens, Battelle Memorial Institute

Forensic DNA analysis through capillary electrophoresis (CE) based typing of short tandem repeat (STR) markers is a well-established technology with widespread legal and technical acceptance and is integral for the generation of over 15 million DNA profiles registered within the FBI's National DNA Index System (NDIS). The emergence of massively parallel sequencing (also referred to as next generation sequencing) presents opportunities for potential success beyond conventional CE-based techniques, specifically with respect to degraded specimens (*e.g.,* missing persons) and complex mixture samples. More significantly, MPS provides a broader scope of informative and discriminating data through multiplexing of additional marker types, such as single nucleotide polymorphisms (SNPs) associated with identity, physical appearance, ancestry and kindred relationships. Consistent with the introduction of such a technology, it is necessary to first perform a comprehensive validation of the process to thoroughly assess reliability of the testing and gain a full appreciation of potential limitations. A critical element of the validation process includes an evaluation of the technology's precision by characterizing the degree of agreement in results across a series of repeated measurements. Unlike CE technology, which requires precision in length measurements for accurate fragment size assignment relative to a standard, MPS-generated data is not directly dependent upon such parameters. As part of a two-year study funded by the National Institute of Justice to assess the technical readiness and feasibility of the MPS technology for forensic applications, Battelle describes an approach for MPS precision evaluation based upon MPS read (or sequence) composition at each locus. Specifically, all reads were identified as alleles, stutter or non-allelic products, and the relative abundance of each category was defined as a weighted proportion of the total reads assigned to the locus. A workflow consisting of the prototype Promega PowerSeq Auto/Y System (24 autosomal STRs/23 Y-STRs) and Battelle's ExactID® analysis software was used to evaluate the degree of precision attained across seven partner U.S. laboratories, incorporating both intra- and inter-laboratory generated sequence data from standard reference materials provided by the National Institute of Standards and Technology (NIST). The test design, resulting data and corresponding interpretations and conclusions from this study will be conveyed within this poster presentation.