MASSIVELY PARALLEL SEQUENCING DATA ANALYSIS – METHODS TO IDENTIFY LOCUS-SPECIFIC ANALYTICAL THRESHOLDS AND STUTTER PERCENTAGES

Adam M Garver[1]; MFS, Elizabeth A Montano[2]; M.S, Jocelyn M Bush[2]; M.S, Max M Larijani[1]; M.S, Steven M Wiechman[1], Esley M Heizer[2]; PhD
[1]Ohio Attorney General's Office Bureau of Criminal Investigation
[2]Battelle Memorial Institute

By generating large amounts of complex data, massively parallel sequencing provides new and interesting challenges for forensic laboratories. Setting an analytical threshold and determining applicable stutter percentages may be applied to validation studies focused on this new technology. Sequencing data from a dilution series of Component C from NIST Standard Reference Material® 2391c, a set of buccal swabs from five different individuals, and a set of mock casework-type samples were analyzed to determine a method of setting a potential analytical threshold and calculating stutter percentages. Amplicons were created via a manual and an automated pathway using the Promega PowerSeq™ Auto/Y amplification kit, and libraries were prepared with the TruSeq® DNA PCR-Free Library Preparation Kit. These libraries were then sequenced on a MiSeq FGx™ Forensic Genomics System running in RUO mode and the data was analyzed with Battelle's ExactID® software and a custom analysis program written specifically for this application. An analytical threshold was calculated by dividing the read counts of all non-allelic reads by the total read count at each locus for each sample. The calculated analytical thresholds, based on these percentages, were consistent across both sample types and concentrations, while varying across loci. This suggests that a dynamic method for setting an analytical threshold may be desirable. Stutter calculations for the massively parallel sequencing data sets were performed by applying a computer program to the data that allowed for quick and efficient determination and analysis of stutter across all locations present in the amplification kit (both autosomal and Y-STRs). This stutter analysis suggests that a per-locus stutter percentage may be applied by the software when interpreting sequencing data in a forensic context.