AN INITIAL EVALUATION OF NEXT-GENERATION SEQUENCING TECHNOLOGY FOR STR-BASED MIXTURE DECONVOLUTION USING PCR-ENRICHED SAMPLES

Nancy McMillan, Jared Schuetter, Dan Bornman, Aaron Sander, David Rigsby, Robert Carnell, Nicholas Linn, Angela Minard-Smith, Wayne Robbins, Seth A. Faith, Battalle

Several recent studies have demonstrated the ability to perform STR genotyping using next generation sequencing (NGS) technologies. The key to success of these studies was the implementation of algorithms that could infer length based STR alleles and determine single nucleotide polymorphisms (SNPs) that capillary electrophoresis (CE) typically will not detect. In this study, we demonstrate refinements to a custom designed set of algorithms to perform both single reference and mixture analysis using NGS. DNA samples were first prepared using a front-end PCR enrichment (STR 18-plex including the CODIS loci) and then sequenced on an Illumina MiSeq instrument. Initial results demonstrated that 13 SNPs were present in a study cohort of 13 individuals. We identified preliminary analytic and stochastic thresholds resulting from both the STR length and sequence variants observed within the sample population. To optimize for minimal allele drop-in (<1%) and low allele drop-out (<1%), receiver operating characteristic (ROC) curve analysis was used. For the next phase of the study, 2-person mixture samples were evaluated at ratios ranging between 1:3 and 1:19. Next, least squared deconvolution algorithms were implemented to provide probability based genotype profiles from the mixtures. We compare these results to CE based mixture analysis (GeneMapper IDX). We conclude that NGS coupled with advanced algorithms show potential for mixture deconvolution. A few points of optimization were also identified in this study, including enhancements to the PCR enrichment methods to reduce sequence error and stutter. Future efforts should include characterization of SNP frequencies within STRs in the human population.