ACCURACY AND RELIABILITY OF ANCESTRY PREDICTIONS FROM THE PRECISION ID ANCESTRY PANEL FOR PURE NATIVE AND FORENSICALLY RELEVANT SAMPLES

Ashley M. Cooley[1,2], Kelly A. Meiklejohn[1], Tracey Dawson Cruz[2], James M. Robertson[3]
[1]Visiting Scientist Program, Counterterrorism and Forensic Science Research Unit, Federal Bureau of Investigation Laboratory Division
[2]Department of Forensic Science, Virginia Commonwealth University
[3]Counterterrorism and Forensic Science Research Unit, Federal Bureau of Investigation Laboratory Division

Human individualization is typically accomplished by analyzing short tandem repeats (STRs). However, in cases where only a partial or incomplete STR profile is obtained, single nucleotide polymorphisms (SNPs) could provide valuable information on phenotypic characteristics (i.e. eye and hair color) as well as biogeographic ancestry. Thermo Fisher Scientific, which developed the high throughput Ion Torrent™ PGM™ sequencer, released the Precision ID Ancestry Panel, a 165-SNP panel for forensic ancestry prediction.

This study was aimed at assessing the accuracy, reproducibility and sensitivity of this panel, along with the ability to provide accurate ancestry predictions, for: 1) seven high quality DNA samples which represent the three major ancestries of forensic interest in the United States (Hispanic, Caucasian and African American); and 2) forensically relevant samples, such as a toothbrush, bone, hair, shaving razor, cigarette butt, and nail clipping (n, 10). Libraries were prepared in triplicate using 0.2 ng, 0.5 ng and 1.0 ng DNA as input for the high quality DNA samples (n, 63), and in duplicate where possible for the forensically relevant samples using 0.05 ng – 1.0 ng of DNA (n, 39).

Data was analyzed using the manufacturer's HID SNP Genotyper plugin (v.4.3.1) as well as CLC Genomics Workbench (Qiagen). Only 2% of all possible quality control (QC) flags were raised for the high-quality DNAs by the plugin QC filter and CLC; 59% of these flags were due to the major allele frequency being outside the manufacturer's thresholds. A total of 9.8% of all possible flags were raised for the forensically relevant samples by the plugin QC filter and CLC; 45% of the flags were due to having no genotype call. The impact of DNA degradation on the reliability and accuracy of ancestry predictions was artificially simulated using subsets of SNPs based on amplicon lengths (i.e. SNPs with amplicon lengths <50 bp, <75 bp, <100 bp, and <200 bp) for the high quality samples prepared using 1.0 ng. Ancestry predictions were not always consistent across SNP subsets, indicating that predictions from degraded samples may not always be reliable. Even though the forensically relevant samples had more flagged SNPs than the high-quality DNAs, 72% of samples still had concordant ancestry predictions between replicates, showing that this panel has the potential to be used in forensic casework with further testing.