PERFORMANCE OF A LARGE, MASSIVELY PARALLEL SEQUENCING SNP ASSAY UTILIZING UNIQUE MOLECULE INDICES FOR MISSING PERSONS IDENTIFICATIONS

Michelle Peck[1], Felix Bittner[1], Šejla Idrizbegović[1], Ana Bilić[1], René Huel[1], Christopher Phillips[2], Andreas Tillmar[3,4], and Thomas Parsons[1]
[1]International Commission on Missing Persons
[2]Institute of Forensic Sciences, University of Santiago de Compostela
[3]Department of Forensic Genetics and Forensic Toxicology, National Board of Forensic Medicine
[4]Department of Clinical and Experimental Medicine, Linköping University

A large SNP panel ("MPSplex") utilizing Unique Molecular Indices (UMIs) was designed to address two of the biggest challenges with missing persons cases – highly degraded DNA and the need for kinship matching based on single and/or distant relatives. The panel consists of over 1200 SNP targets, focusing on highly heterozygous, tri-allelic SNPs. Library preparation utilizes QIAGEN's QIAseq chemistry, which tags the original DNA molecules with UMIs prior to any amplification steps to reduce PCR and sequencing errors and assist in data interpretation. The chemistry also has the advantage of using single target-specific primers, which is particularly beneficial to degraded DNA. Initial sensitivity studies with Genome in a Bottle (GIAB) samples, which have known genotypes, and bone extracts indicated accurate and sensitive detection of SNPs down to 31 pg of DNA input. Average UMI sample coverage was strongly correlated to the initial DNA input. Further, while lower multiplexing levels resulted in increased SNP detection, the average UMI coverage remained in the expected range. Testing with degraded bone samples and artificially degraded GIAB samples with DNA fragment sizes ~150bp or less, resulted in over 1000 SNPs typed. Critical to the interpretation of these SNP profiles is setting appropriate thresholds to deal with drop-in and drop-out. These thresholds are being evaluated in respect to UMI coverage and other sequencing metrics and assessing distinct homozygote and heterozygote thresholds. The advantages of UMIs, particularly in respect to thresholds, will be discussed. Additionally, appropriate matching strategies are necessary to account for such a large panel with linked loci. Expected ranges of likelihood ratios for different relationships based on statistical simulations will also be presented. First application of the MPSplex panel has been with targeted cases with distant relatives and challenging bone extracts. The ICMP has reported highly significant DNA matches with first cousins as references. This represents a significant expansion in the capability of missing persons identifications and further enhancements to processing throughput and analysis strategies will increase its application.