WHOLE MITOCHONDRIAL GENOME TYPING ON ION TORRENT™ PGM™ PLATFORM

Xiangpei Zeng, Seung Bum Seo, Bobby LaRue, Jonathan King, Bruce Budowle, Institute of Applied Genetics, Department of Forensic and Investigative Genetics, University of North Texas Health Science Center

Mitochondrial DNA (mtDNA) is a valuable genetic marker for human identification where evidence contains too little or no nuclear DNA, such as a hair shaft without root or a fingernail. Because there are many more copies of mtDNA than nuclear DNA in a cell, mtDNA tying also can be useful for challenged samples (e.g., old bone) and for population studies, such as human migration and ancestry estimation.

Sanger sequencing is the gold standard method for mtDNA typing in forensic applications, but the methodology has some limitations with throughput, scalability, speed, and resolution. Massively parallel sequencing technology (MPS) provides a platform for more comprehensive coverage of genetic markers per sample analyzed than currently is possible with Sanger sequencing. MPS technologies sequence DNA with high coverage and throughput of specified targets. Moreover, a number of different samples, which can be distinguished by barcoding, may be sequenced simultaneously. One of the available personal sequencers is the Ion Torrent Personal Genome Machine (PGM™). The PGM is a platform based on non-optical sequencing on CMOS integrated circuits that detects small changes in pH, due to release of H+ during addition of a nucleotide to the growing strand. For the PGM, as many as 96 samples can be sequenced at one time within a 2 hour run time.

Sequencing of the entire mitochondrial genome can provide higher resolution and discrimination power than is currently possible with only sequencing portions of the non-coding region of the mitochondrial genome. For whole mtDNA sequencing analyses, 24 samples were analyzed with the PGM platform and Ion 314™ Chip. In this study, sequencing of DNA was assessed for throughput, coverage, concordance of results with other sequencing platforms (i.e., Miseq system, Illumina; and Sanger sequencing by capillary electrophoresis), resolution of heteroplasmy and interpretation of homopolymeric stretch regions. Average coverage of the 24 samples was at least 490X in the reactions. Most variants were concordant between two different NGS platforms and with non-coding region data generated by Sanger sequencing. High quality scores (>89%) were obtained for the typing results from all samples with Haplogrep (Kloss-Brandstaetter A. et al., http://www.haplogrep.uibk.ac.at). While final calls were corrected, some regions were problematic. Most of these regions were located at homopolymer regions due to base position shift. In addition, extreme read strand bias was observed at several locations. For example, sequencing in one direction (i.e., the positive strand) showed the correct bases; while on the other strand a base presented as a deletion.

Overall, the PGM generated good quality sequence data rapidly with relatively high coverage. The progress on these studies will be presented to provide insight to the forensic science community on the near term applications and long term potential utility of MPS for forensic applications.