ASSESSMENT OF LOW-LEVEL ERROR IN MASSIVELY-PARALLEL SEQUENCING (MPS) DATA SETS GENERATED USING THE ILLUMINA® MISEQ® PLATFORM AND SYNTHESIZED HUMAN MITOCHONDRIAL DNA OLIGONUCLEOTIDES

Brittania J. Bintz, M.S.[1], Timothy Driscoll, Ph.D.[2], Mark R. Wilson, Ph.D.[1]

[1]Forensic Science Program, Western Carolina University
[2]Department of Biology, West Virginia University, Morgantown, WV

Massively-parallel sequencing (MPS) methods are proving to be particularly well-suited for mitochondrial DNA analysis, and may provide forensic analysts with a powerful tool that enables deconvolution of mtDNA mixtures, or accurate quantitation of low-level heteroplasmy. However, some effort remains in validating the systems for such analyses. Several MPS platforms are commercially available, each with a unique library preparation strategy and sequencing chemistry that may give rise to method-specific errors. Furthermore, since many alignment and variant calling algorithms are available, there is limited consistency in the use of data analysis methods employed. Finally, no studies have been performed to determine what depth of coverage is required to confidently call a true biological low-level variant above the level of method-generated noise. Here, we describe a study that aims to identify error rates associated with each step in the Illumina® MiSeq® MPS workflow. Initially, synthetic oligonucleotides with sequences matching the rCRS hypervariable (HV) regions I and II of the human mtDNA genome were purchased from Life Technologies. Each oligonucleotide was designed to contain Illumina® sequencing primers, flow cell adapters and multiplexing indices on either end to enable direct sequencing without additional preparation. The oligonucleotides were also designed to contain restriction enzyme cut sites between the target sequence and Illumina® modifications. This design allowed for removal of Illumina modifications so the same sample could be prepared for sequencing using recommended library preparation strategies. Each synthetic oligonucleotide was sequenced a) directly with no additional preparation, b) after Illumina® Nextera® XT library preparation, and c) after triplicate PCR amplification with target specific primers followed by Nextera® XT library preparation. Samples prepared with treatments B and C were sequenced in duplicate to enable assessment of intra-run variation. Sequences were generated on the Illumina® MiSeq® with a v2 300 cycle run kit. Resulting sequence data was aligned to the rCRS using bwa-mem. Variant calling was performed with SAMtools 0.1.19 using the consensus-caller and a maximum depth of 1000. Error rates obtained from all sample treatments were compared to identify differences at each step in the library preparation workflow. Ultimately, this experimentation sets the groundwork for validation of the Illumina® MiSeq® MPS system for mtDNA analysis in forensic casework.