B51 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, MSc, 111 Memorial Drive, NSB 231, Cullowhee, NC 28723; Timothy Driscoll, PhD, West Virginia University, Dept of Biology, Life Sciences Building, PO Box 6057, Morgantown, WV 26506; and Mark R. Wilson, PhD, Western Carolina University, Dept of Chemistry/Physics, Forensic Science, Cullowhee, NC 28723*

After attending this presentation, attendees will better understand ongoing validation studies designed to assess the accuracy, precision, and reproducibility of the Illumina® MiSeq® sample preparation workflow and sequencing chemistry in human mitochondrial DNA (mtDNA) analysis.

This presentation will impact the forensic science by illustrating that the use of MPS technologies in forensic casework is eminent. Therefore, dissemination of information related to MPS validation studies is an important step in showing general acceptance of the method by the scientific community and, ultimately, the courts.

MPS sequencing methods are proving to be particularly well-suited for mtDNA 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.[1-5] 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.

This study seeks to identify error rates associated with each step in the Illumina® MiSeq® MPS workflow. Initially, synthetic oligonucleotides with sequences matching the revised Cambridge Reference Sequence (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 the recommended library preparation strategies. Each synthetic oligonucleotide was sequenced: (1) directly with no additional preparation; (2) after Illumina® Nextera® XT library preparation; and, (3) 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 alignment algorithm.[6] Variant calling was performed with SAMtools 0.1.19 using the consensus-caller and a maximum depth of 1,000.[7]

Error rates obtained from all sample treatments were compared to identify differences at each step in the library preparation workflow. Overall, data quality decreases as the sequencing progresses with late cycle basecalls being lower in quality than early cycle calls. Additionally, reverse reads tend to have lower average quality scores than forward reads. Overall, there does not appear to be a discernable effect from sequencing chemistry or PCR amplification during sample preparation. Ultimately, this experimentation sets the groundwork for validation of the Illumina® MiSeq® MPS system for mtDNA analysis in forensic casework.

*Presenting Author

Reference(s):

1. McElhoe J.A., Holland M.M., Makova K.D., Su M.S., Paul I.M., Baker C.H., et al. Development and assessment of an optimized next-generation DNA sequencing approach for the mtgenome using the Illumina MiSeq. *Forensic Sci. Int*. Genet. 2014:13:20-29.
2. Templeton J.E.L., Brotherton P.M., Llamas B., Soubrier J., Haak W., Cooper A., et al. DNA capture and next-generation sequencing can recover whole mitochondrial genomes from highly degraded samples for human identification. *Investig. Genet*. 2013:4:26-2223-4-26.
3. King J.L., LaRue B.L., Novroski N.M., Stoljarova M., Seo S.B., Zeng X., et al. High-quality and high-throughput massively parallel sequencing of the human mitochondrial genome using the Illumina MiSeq. *Forensic Sci. Int. Genet*. 2014:12:128-135.
4. Parson W., Strobl C., Huber G., Zimmermann B., Gomes S.M., Souto L., et al. Evaluation of next generation mtGenome sequencing using the Ion Torrent Personal Genome Machine (PGM). *Forensic Sci. Int. Genet*. 2013:7:543-549.
5. Parson W., Huber G., Moreno L., Madel M.B., Brandhagen M.D., Nagl S., et al. Massively parallel sequencing of complete mitochondrial genomes from hair shaft samples. *Forensic Sci. Int. Genet*. 2015:15:8-15.
6. Li H. Aligning sequence reads, clone sequences, and assembly contigs with BWA-MEM. *arXiv*:1303.3997 [q-bio.GN]
7. Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009:25(16):2078-2079.

---

mtDNA, Massively Parallel Sequencing, Validation Study