Targeted Mitochondrial DNA Extraction and Enrichment Using the SageHLS System

sage science

Application Note: SageHLS

Chris Boles and Nathan Houde, Sage Science, Inc., Beverly MA Correspondence: chris.boles@sagescience.com

A novel approach for enriching mitochondrial DNA for sequence analysis without requiring PCR or gradient centrifugation.

Introduction: Mitochondrial DNA Sequencing

Mutations in mitochondrial DNA (mtDNA) cause a variety of heritable conditions related to energy deficiency in brain and muscle tissues (reviewed in 1). Diagnosis of mtDNA mutations remains a challenging task for several reasons:

1. Despite the fact that there are several hundred copies of mtDNA per mammalian cell (estimated 300-1000 mtDNAs per cell), the mtDNA genome is very small (a 16,659 bp circle) and comprises only about 0.2% of total cellular DNA.

2. Unlike mutations in nuclear genes for mitochondrial proteins which usually are found in 1 or 2 copies per cell (heterozygous or homozygous), mtDNA mutations may only be present in limited subset of mtDNA molecules. Moreover, mtDNA mutations associated with human disease often have a tissue-specific distribution.

3. There are several nuclear pseudogenes of mitochondrial genes that have high sequence homology to their mtDNA-encoded counterparts. Sequencing reads from these pseudogenes complicates identification of mtDNA variants.

To ensure detection of low abundance mtDNA mutations without interference from nuclear pseudogenes, many researchers enrich mitochondria using differential gradient centrifugation, or enrich specific mitochondrial DNA sequences by PCR, before NGS sequence analysis (reviewed in 2). In this note, we demonstrate a novel approach to mtDNA isolation which provides enrichment levels comparable to gradient centrifugation, but in a small fraction of the hands-on time.

Overview of mtDNA Enrichment Strategy on the SageHLS Platform

The SageHLS system (**Figure 1**) is an integrated platform for extraction and enzymatic processing of high molecular weight (HMW) DNA. HLS extraction utilizes a proprietary agarose gel cassette with two loading wells. Intact cells are loaded into a sample well, and an SDS-based lysis reagent is loaded into an adjacent reagent well. Extraction is performed by electrophoresing the SDS lysis reagent through the sample well. During this process, the cells are rapidly lysed without any mixing or viscous shear force, and the nuclear DNA remains substantially chromosome length.

Because of its large size, the HMW nuclear DNA becomes entangled and immobilized in the agarose as it enters the wall of the sample well. Proteins, RNA, lipids, and other SDS-coated cellular components are electrophoresed rapidly out of the sample well and down the gel column. In addition, cellular DNA molecules less than ~2Mb in size, such as mtDNA, are mobile under the extraction electrophoresis conditions, and move into the agarose gel behind the SDS plug. In this manner, the mtDNA is efficiently separated from nuclear chromosomal DNA, which remains in the sample well.

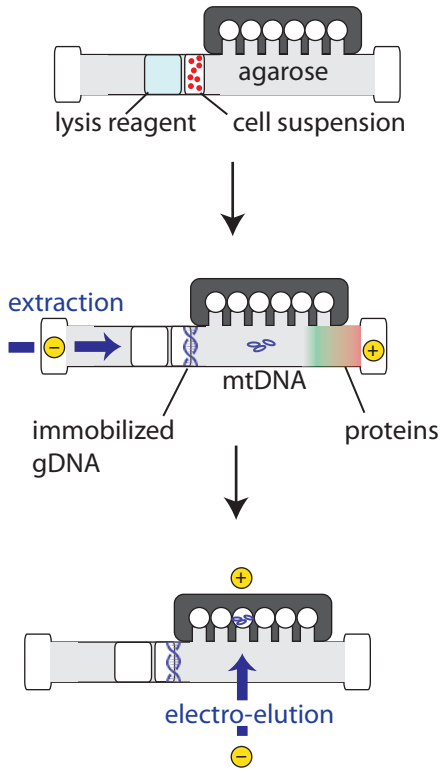

Figure 1. A Schematic of mtDNA extraction process on the SageHLS gel cassette

Experimental Methods

Human white blood cells (WBC) were isolated from ACD whole blood samples by three centrifugal washes in HLS Red Blood Cell (RBC) Lysis Buffer (Sage Science). WBCs were quantified from genomic DNA content using a rapid SDS lysis procedure (Sage Science) followed by DNA quantification using the Qubit dsDNA HS kit (Thermo Fisher Scientific). For mtDNA extraction, two HLS lanes were loaded with 1.2 million WBCs per lane. The HLS workflow included 1.25 hours of extraction and size selection electrophoresis followed by 1.5 hours of electroelution to collect the mtDNA into the elution modules of the HLS cassette.

The elution position of the mtDNA product was determined by qPCR (Thermo Life ABI Taqman Gene Expression Assay ID: Hs0259874-g1 for gene MT-ND2, ABI QuantStudio 3 instrument). In the HLS lane used for Oxford Minion library prep, the mtDNA peak was equally distributed between elution modules 3 and 4. Elution fraction 4, which contained 0.67 ng total DNA (by Qubit HS assay), was used for the library prep. In the HLS lane used for Illumina Miseq library prep, elution fraction 4 contained most of the mtDNA and was used for the library prep. For both libraries, the HLS elution fraction (~80ul total volume) was concentrated by ethanol precipitation. Illumina sequencing libraries were generated with Nextera Flex kits and sequenced using the Miseq 2X150 bp paired end protocol. Oxford Nanopore Minion sequencing was carried out using the Rapid Library kit (RAD004) with a Minion R9.4.1 flow cell. For Minion library construction, 50 ng of HMW E. coli genomic DNA was added to the mtDNA-enriched HLS elution product to act as a carrier during library construction.

Illumina short read data was aligned to the hg38 reference genome(ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/001/405/GCA_000001405.15_GRCh38/seqs_for_alignment_pipelines.ucsc_ids/GCA_000001405.15_GRCh38_no_alt_analysis_set.fna.gz) by BWA-MEM (vs. 0.7.17-r1188, https://github.com/lh3/bwa), and sorted and deduplicated using Samtools (vs. 1.9, https://github.com/samtools/samtools). Coverage over mtDNA and nuclear pseudogenes was evaluated with bedtools (v2.27.1, https://github.com/arq5x/bedtools2) and IGV (Linux vs. 2.6.2, https://software.broadinstitute.org/software/igv/).

Oxford Nanopore Minion data were base-called in real time using the MinIT processing unit (Oxford Nanopore), and the resulting fastq data file was aligned to the hg38 reference using minimap2 (vs. 2.17_x64-linux, https://github.com/lh3/minimap2) and sorted with Samtools. Read length distributions were analyzed using NanoPlot (vs. 1.24.0, https://github.com/wdecoster/NanoPlot), and coverage was evaluated visually using IGV.

Results

The Miseq run yielded approximately 16,000-fold coverage over most of the mtDNA reference (**Figure 1, lower**), with coverage dipping down occasionally to 1000-fold over GC-rich homopolymer regions, as expected for PCR-amplified libraries such as those generated by Nextera Flex chemistry. Approximately 53 single nucleotide polymorphisms (SNPs) were noted and 2 probable indels. 40 of the observed sequence polymorphisms were homogeneous, while 15 positions had mixed base calls with minor base frequencies at 1-2%, suggesting possible low levels of heteroplasmy in the blood donor.

The Minion data (**Figure 2, upper**) showed approximately 60x coverage over the mitochondrial genome, and coverage was significantly more even over the entire mtDNA genome, as expected from a non-amplified long-read library. The transposase-generated Rapid library had an N50 read length of 15,940 bp, and approximately 50% of the reads were nearly full-length ~16,500 reads, apparently resulting from a single transposase insertion into an intact mtDNA circle (**Figure 3**). Base calling was significantly noisier in the Minion run than in the Miseq (**Figure 2, upper vs lower**). Whereas most of the SNPs identified in the Miseq data could be seen in the Minion data, the Minion data had significantly more potential SNPs, and more sites with multiple base calls.

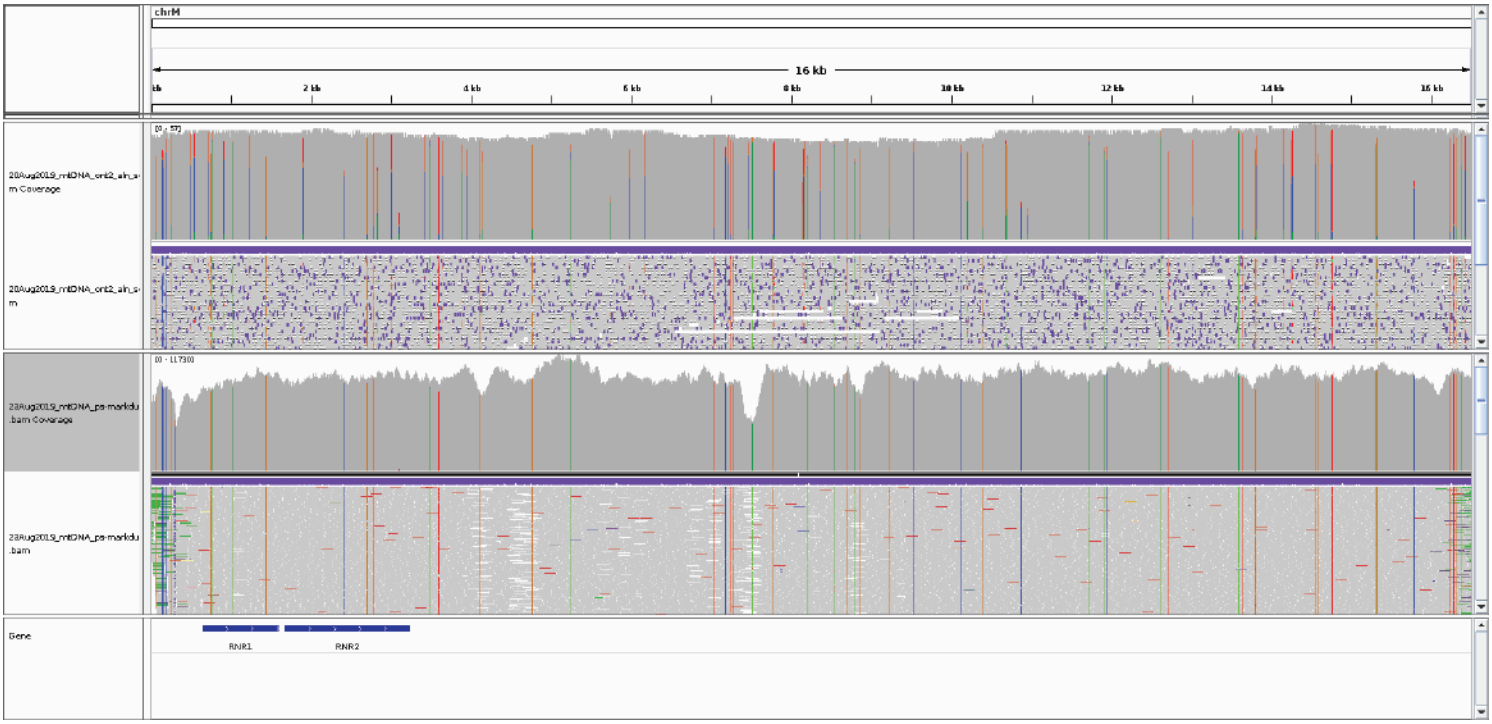

Figure 2. IGV view of sequence data from Minion (upper half) and Miseq (lower half) obtained from HLS-enriched mtDNA sample. In each half window, the upper panel shows coverage and the lower panel shows line graphs for the individual sequencing reads. Coverage for Minion sequencing was approximately 60-fold, and coverage for the Miseq runs was approximately ~16,000-fold.

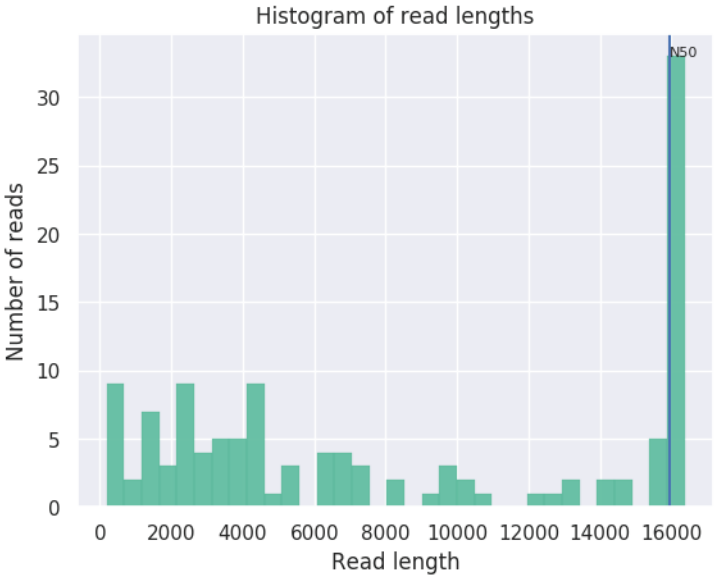

Figure 3. Length distribution of reads from Minion sequencing of HLS-enriched mtDNA. Overall coverage was about 60-fold. The histogram shows that 30 of the reads were full-length 16.5kb reads of the entire mtDNA molecule.

We examined alignments to the total nuclear genome to assess overall enrichment of the mtDNA over the nuclear DNA fraction. As shown in **Table 1** the average coverage per bp over the entire nuclear genome was 0.0090 (reads/bp) whereas over the mtDNA the average coverage of the Miseq data was 60 reads/bp, leading to an estimated enrichment of 6700-fold. This enrichment results from the HLS enrichment process and the fact that there are multiple mtDNA molecules per cell. Assuming an average mtDNA content for WBCs of 350/cell (3,4), the enrichment due to the HLS procedure is estimated to be around 20-fold. The fraction of on-target reads was about 3.5%, which is comparable to values obtained from extraction procedures employing density gradient enrichment of mitochondria prior to DNA extraction (reference 1, 5.7%).

Mitochondrial or PCR enrichment is performed to reduce the number from sequencing reads from the nuclear mitochondrial pseudogenes, which would otherwise complicate mtDNA variant calls. There are 6 closely related pseudogene loci in the human genome as shown in **Figure 4**.

To estimate the nuclear sequence contamination from these pseudogenes in our mtDNA data, we filtered the Miseq data for high mapping quality (saving reads with MQ > 30) and evaluated the number of reads mapping to the nuclear pseudogene positions relative to the homologous mtDNA positions. These data are shown in Table 2. The most highly conserved mitochondrial pseudogene in the nucleus resides on chromosome 1 and covers the central third of the mitochondrial genome (see yellow bar in Figure 3). Table 2 shows that up to 1.3% of the aligned reads could potentially be derived from the chr1 pseudogene under our workflow conditions. Potential contamination from the other 5 nuclear pseudogenes much less likely, estimated to be less than 0.03%.

chr	Size (bp)	Alignments	Alignments/bp
chr1	248956422	2119353	0.0085
chr2	242193529	2307836	0.0095
chr3	198295559	1969860	0.0099
chr4	190214555	1908979	0.01
chr5	181538259	1763171	0.0097
chr6	170805979	1658041	0.0097
chr7	159345973	1504623	0.0094
chr8	145138636	1385232	0.0095
chr9	138394717	1077552	0.0077
chr10	133797422	1250016	0.0093
chr11	135086622	1258004	0.0093
chr12	133275309	1282657	0.0096
chr13	114364328	983565	0.0086
chr14	107043718	844709	0.0079
chr15	101991189	731450	0.0072
chr16	90338345	740060	0.0082
chr17	83257441	716095	0.0086
chr18	80373285	736548	0.0092
chr19	58617616	528799	0.009
chr20	64444167	570862	0.0089
chr21	46709983	365059	0.0078
chr22	50818468	310253	0.0061
chrX	156040895	1415058	0.0091
chrY	57227415	10555	0.0002
chrM	**16569**	**1002069**	**60.4785**

Table 1. Summary of Miseq alignment data from HLS-enriched mtDNA sample.

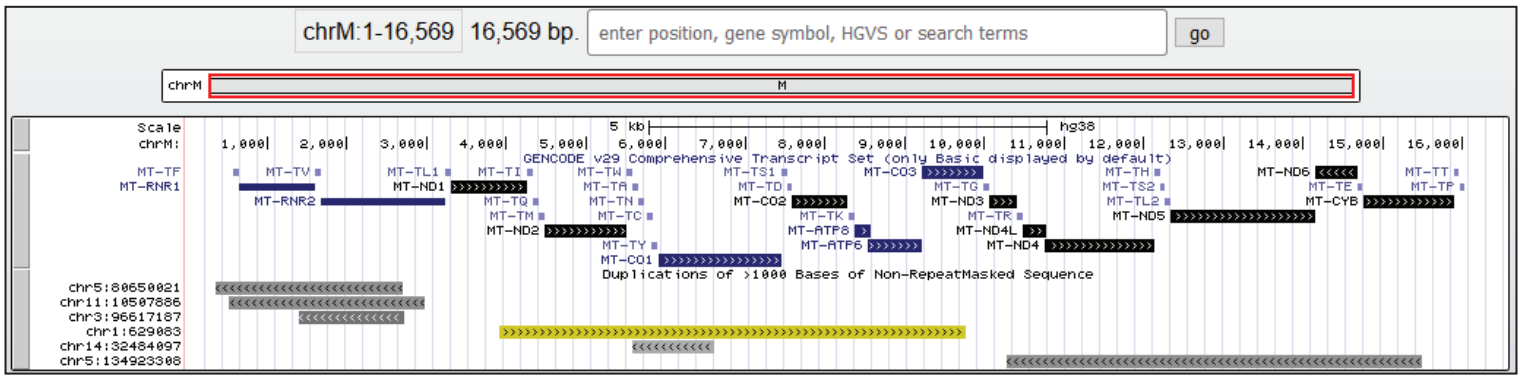

Figure 4. UCSC browser image of hg38 chrM map, showing position of mtDNA genes (upper portion) and major nuclear pseudogenes (lower).

Chromosome (All values for hg38)	Start pos	End pos	Reads in interval	Interval size	Fraction interval covered	**Ratio aligns numts/chrM**
chr1	629084	634924	4096	5840	0.86	
chr3	96617188	96618510	24	1322	0.96	
chr5	80650022	80652368	36	2346	0.82	
chr5	134923309	134928527	103	5218	0.74	
chr11	10507887	10510336	28	2449	0.54	
chr14	32484098	32485118	0	1020	0	
chrM (all)	1	16569	1002069	16569	1	
chrM (chr1 numts)	3914	9755	319891	5841	1	**1.28%**
chrM (chr3 numts)	1396	2718	92177	1322	1	**0.03%**
chrM (chr5/80Mb numts)	341	2697	146972	2356	1	**0.02%**
chrM (chr5/134Mb numts)	10269	15487	358557	5218	1	**0.03%**
chrM (chr11 numts)	521	2972	158076	2451	1	**0.02%**
chrM (chr14 numts)	5583	6606	64890	1023	1	**0.00%**

Table 2. Sequencing coverage of the major nuclear pseudogenes (numts) compared with coverage of the homologous mtDNA regions in the Miseq run using HLS-enriched mtDNA.

Conclusions

Our results demonstrate that the SageHLS system can be used to enrich mtDNA from human cells to purity levels comparable to those offered by much more laborious "gold-standard" methods utilizing density gradient isolation of mitochondria prior to DNA isolation. Our method requires a suspension of 1-1.5m intact human (or other mammalian) cells. In this study we have used an "easy" sample of isolated white blood cells as the input sample. However, there are many procedures for dispersing tissues into cell preparations for cytometry and single-cell sequencing that could be adapted to provide cellular input material for our HLS workflow. Extraction and size selection on the HLS instrument are completely automated after sample loading. The concentration and purity of HLS-enriched mtDNA is compatible with mtDNA sequencing by Illumina platforms, and should enable detection of rare heteroplasmic mtDNA variants that are present at low single digit percentages. In addition, as input requirements for single molecule platforms decrease (Oxford Nanopore and PacBio), our method should be useful for preparing mtDNA inputs for those platforms, which will facilitate detection of larger indels and rearrangements.

References

1. McCormick EM, Muraresku CC, and Falk MJ, (2018) Curr Genet Med Rep. 6(2): 52–61. doi:10.1007/s40142-018-0137-x.

2. Gould MP, Bosworth CM, McMahon S, Grandhi S, Grimerg BT, LaFramboise T (2015) PCR-Free Enrichment of Mitochondrial DNA from Human Blood and Cell Lines for High Quality Next-Generation DNA Sequencing. PLoS ONE 10(10): e0139253. doi:10.1371/journal.pone.0139253

3. Gahan ME, et al. (2001) J Clin Virol. 22 :241-247. doi:10.1016/S1386-6532(01)00195-0.

4. Xia C-Y, et al., (2017) Chin Med J. 130:2435-40. doi: 10.4103/0366-6999.216395.