Rapid Sequencing Of The Entire Human mtDNA Using Next-Generation Sequencing By Ion Torrent PGM

Simon Koren[1], Mojca Tajnik[2], Nataša Toplak[1], Minka Kovač[1], Damjan Glavač[2]

[1]Omega, d.o.o., Dolinškova 8, SI-1000 Ljubljana, Slovenia; [2]University of Ljubljana, Faculty of Medicine, Department of Molecular Genetics, Korytkova 2, SI-1000 Ljubljana, Slovenia; **simon.koren@omega.si**

Introduction

Mitochondria are subcellular organelles that function as energy producers of the cells and regulators of the cellular metabolism. In humans mitochondrial DNA (mtDNA) is a 16,569 bp circular sequence and it is inherited only through the maternal lineage. Mutations in the mtDNA are associated with a range of human diseases and have also been implicated as a driving force behind the aging process. The main goal of our study was to design and describe reliable protocol for sequencing mtDNA with next-generation sequencing (NGS) which is now widely used in biological and medical studies.

Methods

We describe a method for rapid sequencing of the entire human mitochondrial genome by Personal Genome Machine (Ion Torrent PGM). With the aim of finding new disease-causing mutations, we sequenced 6 patient and 2 control mtDNA genomes on one Ion Torrent 314 chip.

Isolated mtDNA was first amplified using long-range PCR. Two amplicons with sizes of 9389 and 8288 bp were designed and amplified using the GeneAmp XL PCR Kit (Life Technologies). Purified PCR products were mixed in equimolar ratio. To produce the sequencing library, PCR products were fragmented enzymatically, using the IonPlus Shear Reagents from Life Technologies. Ligation of sequencing adapters with the barcodes, nick repair and purification was carried out with the Ion Xpress Plus Fragment Library Kit. After size selection and library quantification, the Ion OneTouch instrument was used to prepare the template and then sequencing was performed on the PGM.

Figure 1:
Our Ion Torrent PGM laboratory setup

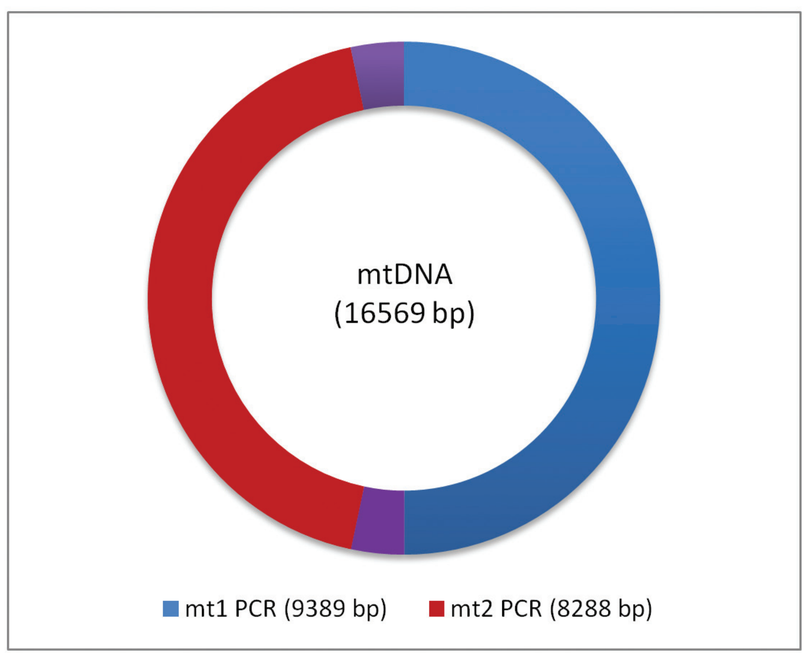

Figure 2:
Schematic representation of the long-range PCR amplicons

Results

Based on the alignment to the reference sequence, we achieved an average read depth of 238x per individual sample (AQ20: 188x). Good coverage of the entire genome was also obtained: on average, 100x coverage was achieved for 87,3% of the bases and 20x coverage for 99,0% of the bases. The combination of high read depth and broad coverage of the genome enabled us to search for genetic variants with high confidence. The number of identified variants ranged from 8 to 38, where majority of them overlapped between the different samples, giving them even more reliable pathogenic value. To better assess their significance, all the variants were studied further using the MITOMAP human mitochondrial genome database. In the case of several identified variants, these could be linked to the observed patient's clinical profile.

	Count	Percentage
Total Addressable Wells	1,262,521	
Wells with ISPs	591,562	47%
Live ISPs	514,933	87%
Test Fragment ISPs	7,005	1%
Library ISPs	507,928	99%

	Count	Percentage
Library ISPs / Percent Enrichment	507,928	87%
Filtered: Polyclonal	153,473	30%
Filtered: Primer dimer	159	<1%
Filtered: Low quality	18,789	4%
Final Library Reads	**335,507**	**66%**

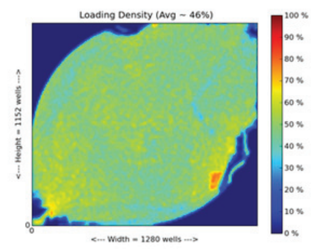

Figure 3:
Loading statistics for the pilot mtDNA sequencing project. The achieved loading densitiy was 47 % and 335.507 (66 %) library reads yielded useful data. With subsequent improvements in loading protocols, we now routinely reach > 80% loading, so we will be able to include a larger number of samples on one chip in the future.

Based on Full Library Alignment to Provided Reference

	AQ20	Perfect
Total Number of Bases [Mbp]	24.87	20.98
Mean Length [bp]	90	79
Longest Alignment [bp]	149	147
Mean Coverage Depth	1,501.10×	1,266.20×
Percentage of Library Covered	99%	99%

Figure 4:
Statistics for the alignment of the reads to the reference genome (rCRS human mtDNA sequence)

AQ20 Reads

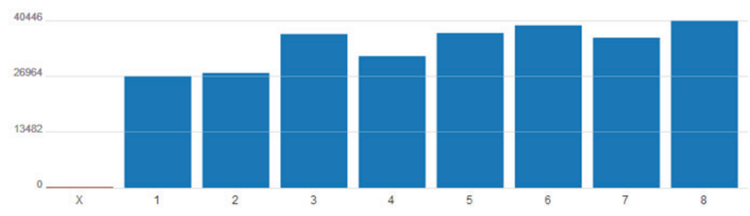

Figure 5:
Number of AQ20 reads for each of the barcoded samples. The results show good uniformity of coverage obtained for each of the individuals.

Table 1:
Detailed statistics for mapped reads, read depth, coverage and germline variants detected for each of the 8 individuals. For variant detection, the VariantCaller plugin for TorrentSuite 2.2 (Life Techonologies) was used.

Variant Caller Reports	Mapped Reads	Reads On-Target	Bases On-Target	Read Depth	1x Coverage	20x Coverage	100x Coverage	Variants Detected
mtDNA_sample_001	30,678	100.0%	100.0%	177.06	99.270%	98.558%	82.636%	37
mtDNA_sample_002	31,786	100.0%	100.0%	194.35	99.294%	98.552%	83.107%	34
mtDNA_sample_003	42,412	100.0%	100.0%	259.50	99.191%	99.179%	88.979%	31
mtDNA_sample_004	36,528	100.0%	100.0%	221.27	99.191%	99.040%	86.813%	38
mtDNA_sample_005	41,649	100.0%	100.0%	250.66	99.197%	99.028%	88.569%	8
mtDNA_sample_006	44,063	100.0%	100.0%	280.67	99.191%	99.125%	90.506%	27
mtDNA_sample_007	40,589	100.0%	100.0%	253.79	99.789%	99.173%	88.545%	18
mtDNA_sample_008	44,997	100.0%	100.0%	271.10	99.197%	99.119%	89.577%	8

Discussion

The described protocol enables a simple and reliable identification of known and novel mitochondrial variants. It offers a fast and cost effective alternative to the classical Sanger sequencing protocols, and the ability to multiplex even larger numbers of samples by barcoding can further decrease the cost per sample. Another advantage is a straightforward data analysis, which does not require any advanced bioinformatics knowledge. Based on the readily achievable deep coverage, the method can also be used to identify heteroplasmic variants with higher resolution, which were much more difficult to study in the past.