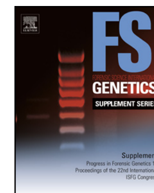
Forensic Science International: Genetics Supplement Series 5 (2015) e13–e15

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

journal homepage: www.elsevier.com/locate/FSIGSS



Evaluation of heteroplasmy detection in the Ion Torrent PGM

S. Magalhães[a,b,c], S.L. Marques[a,b], C. Alves[a,b], A. Amorim[a,b,c], L. Alvarez[a,b], A. Goios[a,b,*]

[a] Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal
[b] Institute of Molecular Pathology and Immunology of the University of Porto (Ipatimup), Porto, Portugal
[c] Faculdade de Ciências da Universidade do Porto, Porto, Portugal

ARTICLE INFO

Article history:
Received 26 August 2015
Accepted 7 September 2015
Available online 10 September 2015

Keywords:
mtDNA
Heteroplasmy
MPS
PGM
Long-range PCR

ABSTRACT

The use of mitochondrial DNA (mtDNA) heteroplasmy in forensic genetics casework and databasing has been limited due to the technical limitations of the PCR/Sanger sequencing technology. Massively Parallel Sequencing (MPS) has been suggested to be more sensitive and accurate in detecting and evaluating heteroplasmy levels than Sanger sequencing, particularly when contamination is controlled and the amplification strategy is carefully designed.

In this work, we present the results of a sensitivity study of heteroplasmy detection and quantification with Ion Torrent PGM, using control samples of predetermined haplotypes belonging to different haplogroups. The amplification strategy used for the PGM sequencing is based on a single long-range PCR, in order to even the proportions of the mixed bases throughout the molecule. To simulate different levels of heteroplasmy we used mixtures of the amplified control samples (amplicons) at variable proportions.

The PGM sequencing strategy accurately detected all artificial heteroplasmies ≥5%, with the exception of variants at repetitive regions. At the 1% level, some of the mixed bases were detected although the false positive background noise does not allow for a trustworthy variant call.

© 2015 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

The use of mitochondrial DNA (mtDNA) heteroplasmies in forensic genetics has been limited, however recent guidelines illustrate the importance of thoroughly detecting heteroplasmies both in casework and databasing [1].

Sanger sequencing technology is not sensitive enough to detect low-level heteroplasmy (≤10–20%) and it does not allow for accurate allele quantification. It has been demonstrated that generally Massively Parallel Sequencing (MPS) methods have higher sensitivity to detect low-level heteroplasmies, and depending on the amplification strategy it has the additional potential to quantify the heteroplasmy level.

Presently, several MPS platforms are available, with different chemistries that should be evaluated and compared when it comes to heteroplasmy detection. The Illumina GA was demonstrated to detect all ≥5% heteroplasmies with virtually no false positives [2]. We here present the results of a sensitivity study of heteroplasmy detection and quantification with Ion Torrent PGM, LifeTechnologies (LT), using control samples of predetermined haplotypes belonging to different haplogroups.

2. Material and methods

DNA samples of two healthy donors, with previously determined complete mtDNA haplotypes (belonging to haplogroups H13a1a (T4) and L1b1a12a (T5)), were extracted from 1 mL saliva samples through a user-developed protocol using DNeasy® Blood & Tissue Kit (Qiagen). In both samples, the complete mtDNA molecule was amplified in a single long-range PCR using the home-designed primers MT_16555L (5′-TAAGACATCACGATGGATCACAGGT-3′) and MT_16554H (5′-TTTAAGGGGAACGTGTGGGCTATTT-3′), with the enzyme mix Ranger DNA Polymerase (Bioline) following the manufacturer's instructions. Amplicons were quantified with Qubit 2.0 Fluorometer (Invitrogen Corporation) and normalized to the same concentration. Mixtures of the two amplicons were performed at different proportions: 1:99, 5:95, 10:90, 50:50, 90:10, 95:5, 99:1. Each mixture was further processed as an independent sample.

The 200 bp sequencing chemistry was used. Enzymatic shearing and Ion Xpress (LT) adapter ligation were carried out in the AB Library Builder (LT) according to the manufacturer's instructions. After Qubit quantification, the 7 different barcoded samples, corresponding to different mixtures, were pooled

* Corresponding author at: Institute of Molecular Pathology and Immunology of the University of Porto (Ipatimup), Porto, Portugal. Tel.: +351 225570700; fax: +351 225570799.
E-mail address: aalmeida@ipatimup.pt (A. Goios).

http://dx.doi.org/10.1016/j.fsigss.2015.09.006
1875-1768/© 2015 Elsevier Ireland Ltd. All rights reserved.

together at equimolar concentrations, and the library was size-selected (fragments ∼260 bp) using the E-Gel SizeSelect Agarose Gel (Invitrogen Corporation), following manufacturer's recommendations. The library was further subjected to 8 amplification cycles using the Platinum PCR SuperMix High Fidelity.

Template was prepared using the Ion One Touch (LT) and enriched in positive Ion Sphere Particles with the Ion One Touch ES system (LT), all according to manufacturer's recommendations. MPS was performed with the Ion Torrent PGM (LT) after the cleaning protocol, as recommended by the manufacturer.

All sequences were analysed with the Ion Torrent Software Suite using the plug-in Variant Caller V.4.4.3.3 (VC), based on the Generic – PGM – Somatic – Low Stringency parameters, producing a table with the list of differences to the rCRS (revised Cambridge Reference Sequence).

3. Results and discussion

The long-range amplification strategy proved to be efficient and specific when applied to reference samples. The primers were designed to anneal in a very conserved segment between HV1 and HV2, so that it is unlikely that a SNP would occur. Should there be the need to type these positions, these would need to be targeted in an independent PCR. When comparing the coverage map obtained from the long-range amplicons with those obtained from an alternative amplification of the mtDNA molecule in two ∼8 kb fragments [3] there were no significant differences observed, presenting generally an uneven coverage profile typical of the PGM sequencing of mtDNA genomes. However, when there is interest in quantifying heteroplasmy/mixture levels throughout the molecule, it is useful to work with a single amplicon so that there are no differences in PCR efficiency in the different regions.

A 316 chip with 7 samples corresponding to 7 different mixtures was run, resulting in a total of 1950575 reads with 96.7% aligned to the reference genome, and an average coverage depth of 15835.5*x* (predicted error rate of 1%). When analyzing the coverage variant sites reported by VC, there are positions that recurrently show lower coverage, which mostly (but not exclusively) are due to strand bias. This is the case of the region near the 310 heteroplasmic stretch and around positions 14560–14770, both showing exclusively sequences in the minus sense.

The VC SNP detection threshold was set to 0.1% to test the detections in the 1:99 mixtures. Erroneous variant calls are found in all mixture proportions (7–12 per sample) and recurrently in the same positions, most of which are near true variant sites or homopolymers. Typical positions/substitutions presenting these results are 71, 310, 520, 525, 8860C, 13046, 14768, 14770 and 16506. These positions are mostly found at very low frequencies (<5%), however when analyzing low-level heteroplasmies (<5%), they result in false positives.

For an easier visualization, a graphical display of the results is presented (Fig. 1), where all the variant positions are represented as differences to one of the samples (T5). Variants with known mixtures not called by the software were assigned a value of 0% or 100%, depending on whether they are equal or different from the rCRS. All artificial heteroplasmies were detected down to the 5 (T4):95(T5) mixture ratio, with the exception of indels at homopolymeric or dinucleotide repetitive regions. The method was not sensitive enough to detect the minor component of the 1 (T4):99(T5) or 99(T4):1(T5) mixtures. These two symmetric mixtures do not show exactly the same behavior. In fact, we observe that, the variants present in sample T4 (haplogroup H13a1a) are not detected in the 1(T4):99(T5) mixture and are under-represented in the 99(T4):1(T5) mixture, and the variants characteristic of sample T5 (L1b1a12a) present an inverse behavior in the same mixtures. This effect is visible also to a lesser extent in some control region variant positions that are not called in the 95 (T4):5(T5) mixture ratio. This suggests that the algorithm employed by the VC plugin appears to be biased towards the reference sequence used (rCRS). This is probably related with one or more default settings defined in the Generic – PGM – Somatic – Low Stringency parameters, since the plugin does not seem to detect variants at frequencies lower than 1% although the threshold was set to 0.1%.

Although some regions present an overestimation of the heteroplasmic frequencies, visible in particular in the region of HVR1 (Fig. 1), generally the observed frequencies were consistent

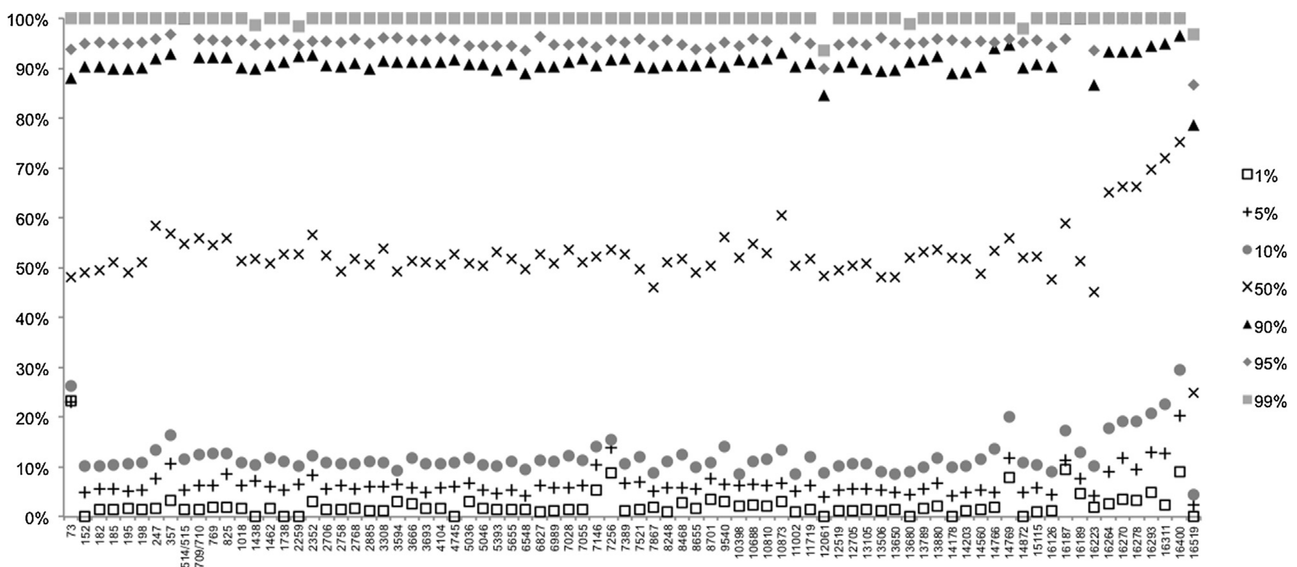


Fig. 1. Detected heteroplasmy levels for the expected substitutions present in the artificial heteroplasmic samples, reported as differences to sample T5. The different symbols represent the relative proportion of sample T4 in the mixture.

with the expected values for each mixture ratio ($y = 0.9897x + 0.0207$; $R^2 = 0.9998$) revealing a high accuracy at detecting heteroplasmy levels.

4. Conclusions

The Ion Torrent PGM MPS technology presents limitations in that (1) the coverage obtained throughout the mtDNA molecule is variable with some regions presenting extreme strand bias, and (2) false positives are mostly generated by alignment problems in the analysis algorithms. However, it proved to be sensitive and accurate at detecting and quantifying mixture/heteroplasmy fractions >5% in artificial heteroplasmic samples, when compared to Sanger sequencing. Note that in true mixtures/heteroplasmies the PCR reactions will alter the initial proportions of each component. By using a single long-range PCR this effect is minimized, and equalized throughout the mtDNA molecule.

Conflict of interest

None.

Acknowledgements

IPATIMUP integrates the i3S Research Unit, which is partially supported by FCT, (Portuguese Foundation for Science and Technology). This work is funded by FEDER funds through COMPETE (Operational Programme for Competitiveness Factors) and National Funds through the FCT, under projects PEst-C/SAU/LA0003/2013, PTDC/ATP-DEM/4545/2012, NORTE-07-0162-FEDER-00018 and NORTE-07-0162-FEDER-000067. LA and AG are supported by FCT fellowships SFRH/BPD/65000/2009, SFRH/BPD/43646/2008, respectively.

References

[1] W. Parson, L. Gusmão, D.R. Harese, et al., DNA commission of the international society for forensic genetics: revised and extended guidelines for mitochondrial DNA typing, Forensic Sci. Int. Genet. 13 (2014) 134–142.
[2] S. Tang, TH, uang, Characterization of mitochondrial DNA heteroplasmy using a parallel sequencing system, Biotechniques 48 (2010) 287–296.
[3] LB, M.D. Zimmermann, W.P. aniaux, Sequencing strategy for the whole mitochondrial genome resulting in high quality sequences, BMC Genomics 10 (2009) 193.