



Differentiating between monozygotic twins in forensics through next generation mtGenome sequencing



Zheng Wang, Suhua Zhang, Yingnan Bian, Chengtao Li*

Shanghai Key Laboratory of Forensic Medicine, Institute of Forensic Science, Ministry of Justice, Shanghai, PR China

ARTICLE INFO

Article history:

Received 5 August 2015

Accepted 7 September 2015

Available online 10 September 2015

Keywords:

Forensic genetics

Mitochondrial genome (mtGenome)

Monozygotic twins

Massively parallel sequencing (MPS)

ABSTRACT

Monozygotic (MZ) twins, considered being genetically identical, cannot be distinguished one from another by forensic short tandem repeat (STR) profiling. The high mutation rate of the mitochondrial DNA (mtDNA) has the potential to become a promising biomarker for the differentiation between MZ twins. With the advent of Next-Generation Sequencing (NGS) approaches, it is now possible to characterize minor differences of mtDNA genomes (mtGenomes) between MZ twins. In this study, we mapped nucleotide differences and heteroplasmies of MZ twins' mtGenomes by NGS technology. Blood samples were taken from 6 pairs of adult MZ twins, and the mtGenomes were sequenced using the Illumina HiSeq 2000 Sequencing System. Point heteroplasmies were observed in five sets of MZ twins and a single nucleotide variant (nt15301) was detected in four sets of MZ twins. Our results give experimental evidence for the hypothesis that variants of mtGenomes could be a perspective biomarker to distinguish MZ twins from each other.

© 2015 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

In 2014, Weber-Lehmann et al. [1] employed whole genome sequencing to search potential somatic mutations for the differentiation between MZ twins, and reported five SNPs present in the twin A, but not in the twin B. Compared with nuclear DNA, mitochondrial DNA (mtDNA), an extra-nuclear genome, exhibits higher mutation rates due to the presence of fewer DNA repair mechanisms. The 10-fold higher mutation rate, relative to nuclear DNA, helps introduce more variability in mitochondrial genome (mtGenome). In this study we explored the use of Next-Generation Sequencing (NGS) technology to identify minor differences in mtGenomes between MZ twins. For mtDNA sequencing with the Illumina's HiSeq 2000 Sequencing System, mtGenome are amplified, fragmented, modified with adaptors and dual indexes, pooled (for multiplexing), generated DNA cluster by bridge PCR, and sequenced.

2. Materials and methods

2.1. Samples preparation

Whole blood samples from six sets of MZ twins were collected by venipuncture without anticoagulation treatment. DNA was extracted using the QIAamp DNA blood mini kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions. The quantity of extracted DNA was estimated using the Quantifiler[®] Human DNA quantification kit (Thermo Fisher, Foster City, CA, USA) on an Applied Biosystems 7500 Real-Time PCR System following to the manufacturer's recommendations. Samples were normalized to 1 ng/ μ L and stored at -20°C until mtDNA enrichment.

2.2. Library preparation

The entire mtGenome was amplified by long PCR in two separate reactions according to the protocol described by Fendt et al. [2]. Negative controls (negative amplification control and reagent blank control) were used as controls for potential contamination. PCR products were purified using the QIAquick PCR Purification Kit (QIAGEN), and then imaged by agarose gel electrophoresis to confirm successful amplification. The quantity of amplicons were determined using the Qubit dsDNA BR Quantification Kit with the Qubit 2.0 Fluorometer (Thermo Fisher),

* Corresponding author at: West Guangfu Road 1347, Shanghai 200063, PR China. Fax: +86 21 52352959.

E-mail address: lichengtaohla@163.com (C. Li).

Table 1

The detailed sequence variant and PHPs of MZ twins' mtGenomes.

| | MZ 1 | MZ 2 | MZ 3 | MZ 4 | MZ 5 | MZ 6 |
|---------|--|---|---|--|---|---|
| Variant | G15301 1A: A (88.09%) 1B: G | | G15301 4A: G 4B: A (84.37%) | | G15301 9A: G (82.11%) 9B: A (88.08%) | G15301 10A: G (94.55%) 10B: A (88.59%) |
| PHP | A16183 1A: C 1B: C (82.72%) T16189 1A: C 1B: C (85.90%) | A10397 2A: G (92.02%) 2B: G C10400 2A: T (92.59%) 2B: T C12705 2A: T 2B: T (93.68%) G15301 2A: G 2B: G (89.3%) | C207 4A: A 4B: A (94.93%) T16189 4A: C (93.47%) 4B: C (92.81%) | G15301 6A: A (84.73%) 6B: A (68.27%) G16129 6A: G (93.37%) 6B: G (94.29%) | | T9540 10A: C (89.43%) 10B: C A10398 10A: G (90.02%) 10B: G G11719 10A: A 10B: A (94.42%) |

and then equal quantities of two amplicons were pooled to produce 1.0 ng DNA for library preparation. NGS libraries were prepared according to the common guidelines for shotgun library preparation. All libraries were normalized to the same concentration, pooled and denatured according to the manufacturer's instructions.

2.3. Sequencing and data analysis

Sequencing was performed on the Illumina HiSeq 2000 Sequencing System with chemistry v3.0 and using the 2 × 100 bp paired-end read mode according to the manufacturer's recommendations. The initial data analysis was carried directly on the HiSeq 2000 System during the run. VarScan2 software [3] was used to identify variants and Variant Call Format (VCF) files were generated. The analysis in this study utilized the manual settings for minimum base call quality (Q30), min coverage ≥ 500, min reads2 (minor component) ≥ 200, and min var freq (minor component frequency) ≥ 0.05. The variant nucleotides from the reference were annotated by base difference, and verified by manually viewing BAM files in Integrative Genomics Viewer (IGV) [4].

3. Results and discussion

Two overlapping PCR fragments were confirmed by agarose gel electrophoresis, and no amplification band was observed in the negative controls samples. The Illumina HiSeq generates approximately 8.20 Gigabases (Gb) Q30 data, each indexed sample would be expected to have over 40,000× coverage at each base position of the mtGenome. In practice, coverage was not dispersed evenly among individuals (mtGenome coverage ranged from 33,078 to 56,130). Point heteroplasmy (PHP) refers to the presence of more than one nucleotide call at a specific position. In this study, minor component detection threshold was set 5.0%, meaning at least 200× corresponding coverage required. With this threshold, we identified a total of 11 bases presenting varying degrees of heterogeneity in five sets of MZ twins, expect for MZ 5 (Table 1). Among these nucleotide positions, nt207, nt16183 and nt16189 are located in hypervariable regions (HV 1 and HV 2), and other 8 positions come from the coding region. One particular set of twins, MZ 2, showed a remarkably large difference; Point heteroplasmy (PHP) was observed at four positions (nt10397, nt10400, nt12705 and nt15301). However, MZ 5 was not detected PHP based on the application of a 5.0% minor variant threshold for detection. Interestingly, one nucleotide position, nt15301, was observed different bases (or major component) in four sets of MZ twins (Table 1). This polymorphic site is located in the coding

region of the cytochrome b, but G15301A variant will not cause amino change. In Human Mitochondrial Genome Database (<http://www.mtodb.igp.uu.se>) [5], 867 A nucleotide (32.1%) was observed in a database of 2704 sequences, and in this study four MZ samples were profiled A nucleotide in this polymorphic site.

4. Conclusions

In this study, we used the Illumina HiSeq 2000 Sequencing System to sequence whole mtGenome of six sets MZ twins. We identified 11 PHPs and one SNP in mtGenomes between MZ twins. Although there is considerably more work to be performed before mtGenome sequencing could be used in casework (e.g. methodical approach for interpretation heteroplasmy), it has been demonstrated that mtGenome sequencing could be used to differentiate between MZ twins.

Conflict of interest

The authors declare that they have no conflict of interest.

Role of funding

This study was supported by grants from the National Key Technology Research & Development Program of the Ministry of Science and Technology of People's Republic of China (2012BAK16B01).

Acknowledgements

We thank Di Zhou for bioinformatics assistance and Ruxin Zhu for technical assistance.

References

- [1] J. Weber-Lehmann, E. Schilling, G. Gradl, et al., Finding the needle in the haystack: differentiating identical twins in paternity testing and forensics by ultra-deep next generation sequencing, *Forensic Sci. Int. Genet.* 9 (2014) 42–46.
- [2] L. Fendt, B. Zimmermann, M. Daniaux, et al., Sequencing strategy for the whole mitochondrial genome resulting in high quality sequences, *BMC Genomics* 10 (2009) 139.
- [3] D.C. Koboldt, Q. Zhang, D.E. Larson, et al., VarScan2. Somatic mutation and copy number alteration discovery in cancer by exome sequencing, *Genome Res.* 22 (2012) 568–576.
- [4] H. Thorvaldsdóttir, J.T. Robinson, J.P. Mesirov, Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration, *Brief Bioinform* 14 (2013) 178–192.
- [5] M. Ingman, U. Gyllensten, mtDB: human mitochondrial genome database, a resource for population genetics and medical sciences, *Nucleic Acids Res.* 34 (2006) D749–D751.