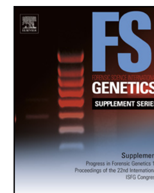
Forensic Science International: Genetics Supplement Series 5 (2015) e83–e85

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

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



Evaluation of the reliability of the data generated by Next Generation Sequencing from artificially degraded DNA samples

I. Carboni[a], P. Fattorini[b], C. Previderè[c], S. Sorçaburu Ciglieri[b], S. Iozzi[a], A.L. Nutini[a], E. Contini[a], C. Pescucci[a], F. Torricelli[a], U. Ricci[a,*]

[a] Diagnostic Genetics Unit, Careggi University Hospital, Florence, Italy
[b] Department of Sciences of Public Health, University of Trieste, Italy
[c] Department of Public Health, Experimental and Forensic Medicine, University of Pavia, Italy

ARTICLE INFO

Article history:
Received 31 August 2015
Accepted 13 September 2015
Available online 18 September 2015

Keywords:
Next Generation Sequencing
NGS
ForenSeq™
Forensic samples
Illumina
Degraded DNA

ABSTRACT

NGS has the potential to be a promising technology for recovering genetic information from challenging specimens in forensic genetics. In order to understand the role of DNA damage on the outcome of NGS, we investigated the performance of ForenSeq™ DNA Signature kit, Illumina (in its pre-commercial version) on a set of *in vitro* degraded trial DNA samples. After DNA quantification by qPCR, duplicate analyses of the samples were carried out. The resulting molecular products were then sequenced by using MiSeq® system (Illumina) and analyzed using ForenSeq™ Universal Analysis Software (Illumina). The coverage and error rate of the NGS data obtained from the degraded samples were compared to the ones gathered from the unmodified DNA. The NGS data showed that the ability of recovering genotypes and the frequency of analytical artifacts are strongly influenced by the degree of damage of the template. NGS was able to call 46–17% of the STR loci and 68–26% of the SNPs in the degraded samples. In addition, when the genotypes from the degraded samples were compared to the ones recovered from the unmodified control DNA, correct typing was achieved from 39 to 4% of the STRs and from 55 to 13% of the SNPs.

These data show that NGS is a powerful method for gathering genetic data from samples which failed the conventional approaches, even if in this experiment the risk of mistyping seems not to be negligible (up to 2%).

© 2015 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

In the last years, Next Generation Sequencing (NGS) has become a very promising technology for massive genetic sequencing [1,2]. In forensic genetics, although first kits are starting to be released [3], few data are available concerning samples with limited quantity and poor quality [4] due to factors such as heat, bad environmental conditions or contaminants. In this work we investigated the performance of the ForenSeq™ DNA Signature kit, Illumina (in its pre-commercial version) on a set of *in vitro* degraded trial DNA samples derived from aqueous hydrolysis. This kit offers the possibility to sequence up to millions of individual DNA strands at the same time combining 63 STRs and 95 identity informative SNPs amplification and supplying a theoretical discriminating power dramatically higher than traditional CE analysis.

* Corresponding author.
E-mail address: ricciu@aou-careggi.toscana.it (U. Ricci).

http://dx.doi.org/10.1016/j.fsigss.2015.09.034
1875-1768/© 2015 Elsevier Ireland Ltd. All rights reserved.

2. Materials and methods

2.1. DNA extraction

Whole blood from a male donor (FM) was collected after informed consent. DNA was extracted using a standard organic extraction protocol based on phenol/chloroform purification and ethanol precipitation [5].

2.2. DNA depurination protocol

15 μg of DNA, as assessed by Nanodrop, were artificially degraded by aqueous hydrolysis at 37 °C for 6 h, 8.5 h and 10 h using a protocol already described [5] with minor changes.

2.3. DNA quantification

About 40 ng (as assessed by Nanodrop) of the resulting depurinated samples, named FM-6, FM-8.5 and FM-10, were quantified together with the unmodified DNA sample FM with the

Quantifiler® Trio kit (Thermofisher) on the QuantStudio 6 Flex (Applied Biosystems) in triplicate experiments.

2.4. Library preparation and sequencing

One nanogram of the unmodified control DNA sample was used in this step while the highest input DNA volume (5 ul) was employed for the three degraded samples. ForenSeq™ DNA Signature protocol consists of different steps:

(1) Amplification: Amplification and DNA tagging using a primer mix with regions specific to DNA sequences upstream and downstream of STRs and SNPs. The primer mix A of the pre-release kit version provides a simultaneous amplification of 63 STRs (62 STRs and amelogenin) and 95 identity-informative SNPs.
(2) Enrichment: Amplification of tagged DNA using primers that add index sequences for sample multiplexing (i5 and i7), as well as common adapters required for cluster generation (P5 and P7). Indexed adapter primers tag DNA templates with a unique combination of molecular index sequences on their ends, allowing a sequencer to separate and isolate the data generated from each sample.
(3) Libraries purification: Purification of PCR products from the other reaction components
(4) Libraries normalization: Preparation of DNA libraries for cluster generation to make sure that samples of varying yields are equally represented within the sequencing run.
(5) Libraries pooling: Combination of equal volumes normalized library to create a pool of samples to be sequenced together.
(6) Libraries preparation for run: Dilution of libraries, adding of internal control (HSC) and libraries denaturation for sequencing. The sequencing run was performed on MiSeq® system (Illumina) converted in MiSeq® FGx system version. All these steps were performed in duplicate experiments.

2.5. Data analysis

Results interpretation was performed by ForenSeq™ Universal Analysis Software, by using default analytical settings for the beta version of the kit here reported: Analytical threshold and Interpretation threshold: 10% and 20%, respectively, of the total number of reads for each given locus. Low coverage: below 650 reads/locus, intra-locus balance threshold: 60%. Stutter: 0%.

Table 1
Results of quantification of the depurinated samples by Quantifiler® Trio DNA Quantification Kit (Thermofisher).

Samples	Y (75 bp)	SA (80 bp)	LA (214 bp)	D.I.
FM (ctrl)	28.6 ± 5.1	32.1 ± 5.5	43.8 ± .1	1.36
FM 6	0.274 ± 0.021	0.336 ± 0.030	<LOQ (u)	u
FM 8.5	0.124 ± 0.011	0.168 ± 0.007	<LOD (u)	u
FM 10	0.008 ± 0.001	0.011 ± 0.001	<LOD (u)	u

The results are given as mean ± CI at the 95% probability level and expressed in ng/μL; *u*: undetermined. Y: Y-specific probe; SA: small autosomal probe; LA: large autosomal probe; D.I.: degradation index.

2.6. Genetic typing

The genotype of the degraded samples was achieved by using the consensus method described in [5].

3. Results

qPCR analysis of the damaged trial samples showed that highly degraded DNA samples with undetermined degradation index were obtained (see Table 1).

The NGS data showed that the possibility of recovering genotypes and the frequency of analytical artifacts are strongly influenced by the degree of template damage. In fact, the ability of calling the STR loci decreased from 46 to 17% proportionally to the degradation of the trial samples and similarly from 68 to 26% for the SNP loci. In addition, correct typing according to the interpretation guidelines described in [5] was achieved from 39 to 4% of the STRs and from 55 to 13% of the SNPs when the genotypes from the trial samples were compared to the ones recovered from the unmodified control DNA (see Fig. 1).

4. Conclusions

Our data show that NGS is a powerful method for gathering genetic data from severely degraded samples, especially in cases where conventional approaches failed. Even if this new analytical approach seems to be very promising for the forensic application, the next crucial step will be the definition of proper thresholds for allele calling; in fact we observed that standard thresholds of ForenSeq™ Universal Analysis Software *beta version* gave a not negligible risk of mistyping in complex samples (up to 2%). For this reason, more data have to be generated by the forensic community in order to identify standard operating procedures for the definition and interpretation of DNA typing results.

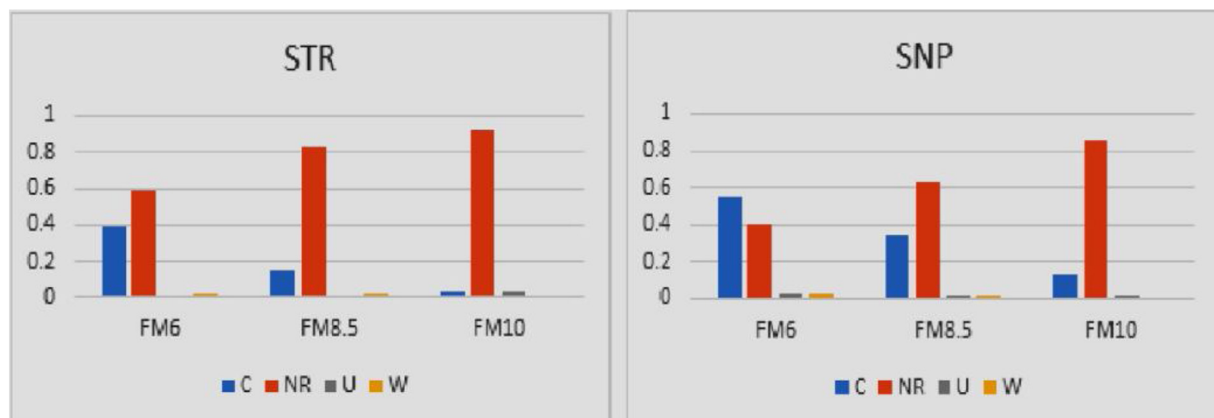


Fig. 1. Evaluation of the genotypes according to [5]. *Y* axis: frequency of the loci correctly typed (C), of the loci which gave no results (NR), ambiguous (U) or wrong typing (W).

Conflict of interest

None.

References

[1] B. Merriman, Ion Torrent R&D Team, J.M. Rothberg, Progress in Ion Torrent semiconductor chip based sequencing, Electrophoresis 33 (2012) 3397–3417.

[2] M.A. Quail, et al., A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers, BMC Genomics 24 (2012) 341–353.

[3] M. Eduardoff, et al., Inter-laboratory evaluation of SNP- based forensic identification by massively parallel sequencing using the Ion PGM^{TM}, Forensic Sci. Int. Genet. 17 (2015) 110–121.

[4] K.B. Gettings, K.M. Kiesler, P.M. Vallone, Performance of a next generation sequencing SNP assay on degraded DNA, Forensic Sci. Int. Genet. 19 (2015) 1–9.

[5] P. Fattorini, et al., The molecular characterization of a depurinated trial DNA sample can be a model to understand the reliability of the results in forensic genetics, Electrophoresis 35 (2014) 3134–3144.