Forensic Science International: Genetics 31 (2017) 135–148

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Research paper

# Massively parallel sequencing of forensic STRs and SNPs using the Illumina® ForenSeq™ DNA Signature Prep Kit on the MiSeq FGx™ Forensic Genomics System

Fei Guo[a,*], Jiao Yu[b], Lu Zhang[a], Jun Li[a]

[a] Department of Forensic Medicine, Criminal Investigation Police University of China, No. 83, Tawan Street, Huanggu District, Shenyang 110854, Liaoning, PR China
[b] Criminal Science and Technology Institute of Liaoning Province, No. 2, Qishan Middle Road, Huanggu District, Shenyang 110032, Liaoning, PR China

ARTICLE INFO

*Keyword:*
Massively parallel sequencing (MPS)
ForenSeq™ DNA Signature Prep Kit
MiSeq FGx™ Forensic Genomics System
Short tandem repeat (STR)
Single nucleotide polymorphism (SNP)
Evaluation

ABSTRACT

The ForenSeq™ DNA Signature Prep Kit (ForenSeq Kit) is designed to detect more than 200 forensically relevant markers in a single reaction on the MiSeq FGx™ Forensic Genomics System (MiSeq FGx System), including Amelogenin, 27 autosomal short tandem repeats (A-STRs), 7 X chromosomal STRs (X-STRs), 24 Y chromosomal STRs (Y-STRs) and 94 identity-informative single nucleotide polymorphisms (iSNPs) with the option to contain 22 phenotypic-informative SNPs (pSNPs) and 56 ancestry-informative SNPs (aSNPs). In this study, we evaluated the MiSeq FGx System on three major parts: methodological optimization (DNA extraction, sample quantification, library normalization, diluted libraries concentration, and sample-to-cell arrangement), massively parallel sequencing (MPS) performance (depth of coverage, sequence coverage ratio, and allele coverage ratio), and ForenSeq Kit characteristics (repeatability and concordance, sensitivity, mixture, stability and case-type samples). Results showed that quantitative polymerase chain reaction (qPCR)-based sample quantification and library normalization and the appropriate number of pooled libraries and concentration of diluted libraries provided a greater level of MPS performance and repeatability. Repeatable and concordant genotypes were obtained by the ForenSeq Kit. Full profiles were obtained from ≥100 pg input DNA for STRs and ≥200 pg for SNPs. A sample with ≥5% minor contributors was considered as a mixture by imbalanced allele coverage ratio distribution, and full profiles from minor contributors were easily detected between 9:1 and 1:9 mixtures with known reference profiles. The ForenSeq Kit tolerated considerable concentrations of inhibitors like ≤200 μM hematin and ≤50 μg/ml humic acid, and > 56% STR profiles and > 88% SNP profiles were obtained from ≥200-bp degraded samples. Also, it was adapted to case-type samples. As a whole, the ForenSeq Kit is a well-performed, robust, reliable, reproducible and highly informative assay, and it can fully meet requirements for human identification. Further, sensitive QC indicator and automated sample comparison function in the ForenSeq™ Universal Analysis Software are quite helpful, so that we can concentrate on questionable genotypes and avoid tedious and time-consuming labor to maximum the time spent in data analysis.

## 1. Introduction

DNA analysis has become the cornerstone of contemporary forensic science for over 30 years. At the current time, most forensic DNA typing utilizes polymerase chain reaction (PCR) and capillary electrophoresis (CE), which is based on detecting fragment length variation in short tandem repeat (STR) markers [1]. A small percentage of forensic investigations also use CE-based Sanger sequencing to analyze specific regions of mitochondrial DNA (mtDNA) [2,3]. With the development of technology and the growth of public expectation, forensic scientists are faced with much more difficult casework samples, such as heavily degraded samples, complicated kinship testing, missing person identifications, ancestry investigations and other human or non-human identifications as well. All of these demands are limited by the fixed capabilities of CE technology [4], and many scientists have begun to focus on some new analysis methods [5,6].

Massively parallel sequencing (MPS), also known as next generation sequencing (NGS), was introduced in 2005 [7] and is actively applied to solve various questions in areas such as evolutionary biology, oncology, microbial genomics, agrigenomics, and disease genomics worldwide [8]. MPS technology has the advantage of improving sample throughput, locus multiplexing, workflow efficiency and detection

* Corresponding author.
*E-mail address:* nevergfaye@gmail.com (F. Guo).

http://dx.doi.org/10.1016/j.fsigen.2017.09.003
Received 12 February 2017; Received in revised form 6 August 2017; Accepted 6 September 2017
Available online 08 September 2017

resolution that can not only address the limitations of CE but also give more informative clues in forensic science [9–20]. In early 2015, Illumina launched a complete sample-to-profile MPS workflow for forensic DNA analysis by the MiSeq FGx™ Forensic Genomics System (MiSeq FGx System). The workflow begins with the ForenSeq™ DNA Signature Prep Kit (ForenSeq Kit), which allows targeted PCR amplification of more than 200 forensically relevant markers in a single reaction, including Amelogenin, 27 autosomal STRs (A-STRs), 7 X chromosomal STRs (X-STRs), 24 Y chromosomal STRs (Y-STRs) and 94 identity-informative single nucleotide polymorphisms (iSNPs) with the option to contain 22 phenotypic-informative SNPs (pSNPs) and 56 ancestry-informative SNPs (aSNPs) depending on the DNA Primer Mix (DPM) chosen [21]. Additionally, the MiSeq FGx™ is a benchtop sequencing instrument that accomplishes bridge PCR amplification and sequencing-by-synthesis (SBS) of up to 96 samples in a single run by measuring fluorescence signals of labeled nucleotides through the use of instrument-specific reagents and flow cells, imaging hardware, and data analysis software [22]. Finally, the ForenSeq™ Universal Analysis Software (ForenSeq UAS) is on a standalone server with a user interface specifically designed for forensic genomics that enables run setup, sample management, genotype analysis, and report generation and provides population statistics and automated sample comparison, as well as an optional feature for estimating bio-geographical ancestry, hair color, and eye color [23].

In 2015, the Illumina® Beta Version ForenSeq™ DNA Signature Prep Kit (including Amelogenin, 28 A-STRs, 9 X-STRs, 25 Y-STRs, 95 iSNPs, 22 pSNPs and 56 aSNPs) was evaluated by J.D. Churchill et al. [24] on aspects of performance, reliability, sensitivity, mixture, concordance and challenged samples. There are now multiple other papers out that developmentally validate or independently evaluate the MiSeq FGx System [25–29]. In this study, we evaluated the MiSeq FGx System on three major parts: methodological optimization (DNA extraction, sample quantification, library normalization, diluted libraries concentration, and sample-to-cell arrangement), MPS performance (depth of coverage, sequence coverage ratio, and allele coverage ratio), and ForenSeq Kit characteristics (repeatability and concordance, sensitivity, mixture, stability and case-type samples) according to the guidelines 'Validation Guidelines for DNA Analysis Methods (2016)' issued by the Scientific Working Group on DNA Analysis Methods (SWGDAM) [30].

## 2. Materials and methods

### 2.1. Samples, DNA extraction and quantification

Details of samples, extraction and quantification methods used in this study are listed in Table S1.

For DNA extraction, sample quantification and sample-to-cell arrangement (i.e., the number of samples loaded on a flow cell) studies, a total of 63 buccal swabs were collected from unrelated healthy Han Chinese after informed consent, which were equally divided into four groups (I: AutoMate + qPCR; II: AutoMate + Qubit; III: Chelex + qPCR; IV: Chelex + Qubit) by two extraction methods (1) the PrepFiler *Express* BTA ™ Forensic DNA Extraction Kit on the AutoMate *Express* ™ Forensic DNA Extraction System (Thermo Fisher Scientific, MA, USA) according to the manufacturer's recommendations [31] and (2) the Chelex 100 protocol [32] and by two quantification methods (1) the Quantifiler ® Human DNA Quantification Kit on the Applied Biosystems ® 7500 Real-time PCR System (Thermo Fisher) according to the manufacturer's recommendations [33] and (2) the Qubit ® dsDNA HS Assay Kit on the Qubit ® 3.0 Fluorometer (Thermo Fisher) according to the manufacturer's recommendations [34].

For the repeatability study, the Male Control DNA 2800M (Promega, WI, USA) was prepared as follows (1) the repeatability of targeted PCR: three libraries of 2800M were independently prepared, but simultaneously sequenced in one flow cell; (2) the repeatability of bridge PCR: one library of 2800M was prepared, but independently sequenced in two different flow cells.

For concordance, library normalization, diluted libraries concentration and MPS performance studies, a total of 8 samples were employed including the Standard Reference Material ® (SRM) 2391c Components A–C and E–F (the National Institute of Standards and Technology (NIST), MD, USA), Male Control DNA 2800M and 9948 (Promega), and Female Control DNA 9947A (Thermo Fisher).

For the sensitivity study, a series of 2800M dilutions (5 ng, 2 ng, 1 ng, 500 pg, 200 pg, 100 pg, 50 pg, 20 pg and 10 pg) were quantified and added in library preparation.

For the mixture study, mixtures of 2800M and 9947A were prepared at various ratios (1:49, 1:19, 1:9, 1:1, 9:1, 19:1 and 49:1) while holding the total amount of input DNA mixed constantly at 1 ng in library preparation. The SRM 2391c Component D was also included in this experiment, ratio of which was documented as Component A:Component C = 3:1 in the Certificate of Analysis for Standard Reference Material ® 2391c [35].

For the stability study, stock solutions were made in 40 mM NaOH per 5000 μM hematin (Sigma–Aldrich, MO, USA) and in 1 mM NaOH per 500 ng/μl humic acid (Sigma–Aldrich) and then added to 1 ng 2800 M prior to targeted PCR amplification in concentrations varying from 100 to 800 μM and 10–200 μg/ml, respectively. Degraded samples (100 bp, 200 bp and 300 bp) were prepared according to F. Guo et al. [14].

For case-type samples, four archived bloodstains were collected in the years of 1995, 1995, 2005 and 2010, respectively.

### 2.2. Library preparation

Details of DPM in conjunction with the number of pooled libraries chosen in this study are listed in Table S1.

In general, library preparation was performed using the ForenSeq Kit (Illumina, CA, USA) according to the manufacturer's recommendations [21], mainly including six steps as follows: (1) Amplify and tag targets: except for dilutions in sensitivity study, degraded samples (200 pg) and 9948 (200 pg), amplification of 1 ng input DNA was performed in a volume of 15 μl containing 4.7 μl PCR1 Reaction Mix, 0.3 μl Enzyme Mix and 5 μl DPM on the GeneAmp® System 9700 (Thermo Fisher) using the thermal cycling conditions of enzyme activation for 3 min at 98 °C, amplification for 8 cycles of 45 s at 96 °C, 30 s at 60 °C, 2 min at 54 °C (ramp speed of 8%), 2 min at 68 °C (ramp speed of 8%) and 10 cycles of 30 s at 96 °C, 3 min at 68 °C (ramp speed of 8%), extension for 2 min at 68 °C, and hold at 10 °C forever. The ramping mode depends on selected thermal cycler. (2) Enrich targets: a total of 4 μl Index 1 (i7) adapters, 4 μl Index 2 (i5) adapters and 27 μl PCR2 Reaction Mix were consecutively added into each amplified target and then incubated for 30 s at 98 °C, 15 cycles of 20 s at 98 °C, 30 s at 66 °C and 90 s at 68 °C, 10 min at 68 °C, and hold at 10 °C forever. This step adds indexes for sample multiplexing and adapters for cluster generation. (3) Purify libraries: each enriched library was purified twice using 45 μl Sample Purification Beads (SPB) and then added 52.5 μl Resuspension Buffer (RSB). (4) Normalize libraries: each of 20 μl purified library was mixed with 45 μl master mix of Library Normalization Additives 1 (LNA1) and Library Normalization Beads 1 (LNB1), then washed twice using 45 μl Library Normalization Wash 1 (LNW1) and mixed with 32 μl 0.1 N NaOH (HP3). Finally, each of 30 μl Library Normalization Storage Buffer 2 (LNS2) was added into 30 μl normalization library. (5) Pool libraries: normalized libraries were pooled in equal volumes (5 μl) but the number of pooled libraries depended on DPM used and/or data quality required, although the manufacturer's protocol recommended at most 32 libraries for casework samples and at most 96 libraries for database samples. (6) Denature and dilute libraries: 7 μl pooled libraries was added into 591 μl Hybridization Buffer (HT1) and then mixed with 2 μl Human Sequencing Control (HSC) mixture.