Kinship Testing Based on SNPs Using Microarray System

Sohee Cho\textsuperscript{a}  Hee Jin Seo\textsuperscript{a}  Jihyun Lee\textsuperscript{a}  Hyung Jin Yu\textsuperscript{b}  Soong Deok Lee\textsuperscript{a,c}

\textsuperscript{a}Department of Forensic Medicine, Seoul National University, College of Medicine, Seoul, South Korea; \\
\textsuperscript{b}DNA Link, Seoul, South Korea; \\
\textsuperscript{c}Institute of Forensic Science, Seoul National University, College of Medicine, Seoul, South Korea

Introduction

Since single nucleotide polymorphism (SNP) has been introduced as a new promising marker in forensic genetics, its usefulness for personal identification and kinship analysis has been evaluated with the development of several SNP panels [1–4]. The advantage of SNP analysis for kinship testing has been demonstrated in several challenging cases with ambiguous STR genotyping or highly degraded DNA samples commonly found in mass disasters [5–7]. Because SNPs have lower mutation rate of approximately $10^{-8}$ compared to $10^{-4}$ for STRs and smaller amplicon sizes due to single-base polymorphism feature, employing SNPs can be beneficial in certain cases involving poor-quality DNA [8]. However, due to the biallelic nature of SNPs, generating the shortage of information, a higher number of SNPs is required in an analysis with a suitable genotyping platform to allow the discrimination power to be equivalent to currently used 13–15 STR loci.

Considering the potential for massive typing, microarray is a powerful tool with a high density of markers on an array, providing a fast and automated detection in a high-throughput system simultaneously [9]. In addition, it could be designed to include not only an informative marker set but also combined sets, such as lineage markers, phenotypic markers, and ancestral informative markers (AIMs) depending on its specific applications [4]. For forensic application, the AccuID\textsuperscript{®} chip ver 1.0, a SNP-based microarray system, created by combining the Affymetrix Resequencing array technology and multiplex PCR technology, has been introduced. This array contains 169 autosomal SNPs developed for forensic application was applied to our practical case for kinship analysis that had remained inconclusive due to partial STR profiles of degraded DNA and possibility of inbreeding within the population. Case Report: 169 autosomal SNPs were typed on array with severely degraded DNA of two bone samples, and the kinship compared to genotypes in a reference database of their putative family members. Results: Two bone samples remained unidentified through traditional STR typing with partial profiles of 10 or 14 of 16 alleles. Because these samples originated from a geographically isolated population, a cautious approach was required when analyzing and declaring true paternity only based on PI values. In a supplementary SNP typing, 106 and 78 SNPs were obtained, and the match candidates were found in each case with improved PI values than using only STRs and with no discrepant SNPs in comparison. Conclusion: Our case showed that the utility of multiple SNPs on array is expected in practical forensic caseworks with an establishment of reference database.
Case Report

Background of Samples

A number of bone remains were in the identification process through comparison to their putative family members. The bones had been kept underground for several decades and thereafter stored at room temperature until analysis followed by the cleaning procedure [11]. Conventional autosomal STR and Y-STR analysis using commercial kits along with mitochondrial DNA typing were performed but failed to generate full STR profiles for the bone DNA samples. Furthermore, the utility of lineage markers was limited due to the samples’ geographic and population background. The population has been isolated on a small island with limited interaction with other regional groups outside of the island for a long period of time. This may result in an increased risk for inbreeding within this population. Among the unidentified bones, thereafter, two bones with partial STR profiles but relatively higher number of alleles were chosen and applied to kinship analysis using the SNP-based microarray.

Sample Preparation and DNA Extraction

To obtain reference profiles from family members, blood was collected from 585 people with informed consents. DNA was extracted using the QuickGene DNA Whole Blood Kit S with the QuickGene-Mini80 instrument (Kurabo, Osaka, Japan). Bone DNA was isolated from the chopped fragment of two femur bones through a phenol/chloroform extraction method [11]. The DNA concentration was measured using NanoDrop TM 2000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA).

Autosomal STR and Y-STR Typing

The initial conventional autosomal STR and Y-STR typings were performed using AmpFISTR® Identifier® PCR Amplification Kit (Thermo Fisher Scientific) and AmpFISTR® Yfiler® PCR Amplification Kit (Thermo Fisher Scientific), respectively, following the manufacturer’s instructions. Moreover, for bone DNA samples, AmpFISTR® MiniFiler® PCR Amplification Kit (Thermo Fisher Scientific) was used for additional typing. Thermal cycling was performed using the GeneAmp® PCR System 9700 (Thermo Fisher Scientific), and analysis of amplified PCR products was carried out on ABI 3130 Genetic Analyzer (Thermo Fisher Scientific). The raw data were analyzed by using GeneMapper® ID v3.2 (Thermo Fisher Scientific).

SNP Genotyping

Two different platforms for SNP typing were applied to bone and blood DNA samples. The fluidigm-based genotyping platform was affordable to use for the massive amount of blood DNA samples, with sufficient quality and quantity of DNA, in terms of cost and time. This procedure was carried out on the 96.96 Dynamic Array® IFCs genotyping system (Fluidigm, San Francisco, CA, USA) according to the manufacturer’s instructions. The typing for bone DNA samples was performed using AccuID® Ver 1.0 chip (Affymetrix, Santa Clara, CA, USA) following the protocol described in our previous report [10]. Briefly, multiplex amplification of DNA was performed in a reaction containing DL Premix (DNA Link, Inc., Seoul, South Korea) and pooled 169 primer sets. Purified PCR products after amplification step were quantified using Agilent DNA 1000 Kit & High sensitivity kit (Agilent Technologies, Inc., Santa Clara, CA, USA), and applied to GeneChip CustomSeq Resequencing Array according to the Affymetrix manufacturers’ protocols [12]. Subsequently, the PCR product was labeled in GeneChip DNA Labeling Reagent (Affymetrix), and then hybridized to probes on array. Staining in GeneChip Fluidics Station (Affymetrix) and scanning using GeneChip Scanner 3000 (Affymetrix) were done by following the AccuID Standard Protocol. The raw data was exported using Gseq software (Affymetrix).

Data Analysis for Calculation of Paternity Probability

The raw data from two typing platforms were imported into AccuID console software for analysis, which calculated likelihood ratio (LR) and matching rate by comparing two pairs of SNP data. This uses two quality control parameters to generate reliable calls for assessing array data, especially when concerning the call of low-quality DNA. Here, the quality score was set to 6.0, a score based on the resequencing calling algorithm [13, 14], and typing call rate cutoff of 0.5, denoting the reliability with at least 50% calling of the surrounding sites [14]. The genetic relatedness of parent-child pairs was assessed by calculated LR values based on the identical-by-descent (IBD) algorithm with minor allele frequency (MAF), genotype data and the allele frequency of 1,197 Koreans acquired using Genome-Wide Human SNP Array 6.0 (Affymetrix).

Results

SNP Genotyping Result of Putative Family Samples

The genotyping results obtained from blood DNA samples were compared. The fluidigm assay was more applicable to blood samples than the AccuID array, providing the equivalent performance for sufficient quantity of DNA at a more affordable cost. The average number of SNPs for all 585 samples showed a success call rate of 99.69%, except for only 1–2 SNPs per sample producing no calls. This was common in almost all individual samples, suggesting that it might be originated from technical issues in arrays.

SNP Genotyping Result of Bone Samples Based on AccuID Microarray

Bone DNA samples were applied to the AccuID microarray platform. According to the standard protocol for AccuID. It is mentioned that only 1 ng of DNA in SNP multiplexing was sufficient to acquire equivalent result to full STR profiles, but a greater amount is recommended for degraded sample. In this experiment, the maximum volume of DNA extracts from bones was added in reactions due to undetermined quantity of DNA. For each sample, 106 and 78 SNPs were typed in array, with estimated call rates of 62.7% and 46.2%, respectively. In the parallel reaction of STRs performed simultaneously with microarray in this experiment, no allele was amplified (data not shown). This result demonstrates that SNPs were more reliable than STRs when obtaining results from degraded samples.

The Comparison of SNPs for Kinship Testing

A database containing SNP genotypes obtained from blood samples was constructed and referred to data of tested bone samples for random comparison using the AccuID console software for analysis. Of the commonly typed SNPs in both blood and bone samples, markers with no excluding genotypes were chosen for comparison. For each bone sample only one candidate was found to be in parent-child relationship with maximum number of SNP calls matching among comparable markers. Each pair was designated as case (1); bone (1) – putative family (1) or case (2); bone (2) – putative family (2) (table 1). In case (1), alleles at 106 SNP loci were shared with one another, and LR (or paternity index (PI) in these cases) was calculated to be 4,647,773,835. This value gave probability of 99.9997848% for the parent-child relationship at 50% prior, which was increased compared to 99.985559% by initial STR typing with partial profiles. In case (2), shared alleles at 76 SNP loci yielding a PI of 754,670.2037 showed slight improvement of the probability from 99.999841 to 99.999874%.

Y-STR typing
result, only available in case (2), indicated no discrepancy with partial profile. In summary, both cases were validated to have stronger paternity indexes with higher probabilities through SNP typing.

**Discussion**

It has been reported that SNPs can offer extra discrimination power to relationship tests that routinely are done using STR typing, which especially improves the test for close relatives [15]. Statistically, a larger number of SNPs can provide stronger evidential support either to exclude or to confirm the familial relationship with higher LR [16], which would make the use of SNP markers in kinship analysis relevant. Furthermore, the small sizes of PCR amplicons [10] expand their usability to challenging samples such as highly degraded DNA commonly encountered in real casework. Here, we demonstrated the successful use of an SNP array for kinship analysis in highly degraded bone samples. The array used for these samples contained 169 SNP probes designed in triplicate for each SNP marker, which could maximize its sensitivity and specificity. For bone samples of this case, each 78 and 106 SNPs were typed successfully in a single reaction despite the concerns resulting not only from the use of severely degraded DNA but also from the presence of PCR inhibitors in bone samples. The number of identical loci in each case was found to be over 50 SNPs that were required to declare paternity [16], and showed enhanced probabilities than only using incomplete STR results in the previous test. These results suggest that this platform has the potential to assist standard STR typing, especially in challenging samples.

Although no discrepancy was found in partial STR profiles of the initial experiment, even the approximate paternity probability over 99.9% could not assure the paternity due to a lack of co-ancestry coefficient established within this population [17, 18]. Previously, we had found cases with sufficient PI of 99.9% by standard STR typing in which paternity was excluded by more than 10 mismatched SNPs in supplementary SNP testing [19]. This concern with regard to inbreeding is frequently found in several reports, especially within undetermined subpopulations or relatively isolated populations with limited regional interactions [20, 21]. For instance, in a small regional population in Northern Germany, a group of unrelated people showed sufficient probabilities with no mismatch STRs in 26 pairs (0.022%) and probabilities over 99.99% even with 1–2 mutations [22]. Such a situation would result in even more serious effects when evaluating the discrimination power of partial STR profiles of degraded DNA.

Concerning the kinship analysis for related groups in the specific incident of e.g. a mass disaster limited to certain area, it would be valuable to establish a specialized genetic profile database for comparison [23, 24]. In our study, the reference profiles of putative family members related to this casework were established in a database and imported into the software, AccuID® console. The in-house developed software is capable of random matching between a reference data set and tested data by comparing genotypes in all possible pairs. The SNP data obtained from additional bone samples in the future assay could also be compared with the current reference software. This application is expected to be valuable for practical work such as massive typing, along with the advantage of automated array providing a less laborious and time-saving process.

**Conclusion**

In the present report an additional SNP-based array system was successfully used for resolving the kinship analysis in a case with highly degraded DNA, where initial data from established STRs could not establish paternity with convincing probabilities. The use of SNPs on array along with the establishment of a reference database for sample analysis are expected to improve the reliability of practical forensic caseworks.
Acknowledgements

This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2014M3A9E1069989) in Korea.

Disclosure Statement

The authors declare no conflict of interest.

References


