JOURNAL OF FORENSIC SCIENCES

*J Forensic Sci,* November 2012, Vol. 57, No. 6
doi: 10.1111/j.1556-4029.2012.02154.x
Available online at: onlinelibrary.wiley.com

**PAPER**

**CRIMINALISTICS**

*Yi-Liang Wei,*[1,2] *Ph.D.; Cai-Xia Li,*[3,4] *Ph.D.; Jing Jia,*[5] *Ph.D.; Lan Hu,*[3,4] *Ph.D.; and Yao Liu*[1,3] *Ph.D.*

# Forensic Identification Using a Multiplex Assay of 47 SNPs*

**ABSTRACT:** As a powerful alternative to short tandem repeat (STR) profiling, we have developed a novel panel of 47 single nucleotide polymorphisms (SNPs) for DNA profiling and ABO genotyping. We selected 42 of the 47 SNPs from a panel of 86 markers that were previously validated as universal individual identification markers and identified five additional SNPs including one gender marker and four ABO loci. Match probability of the 42 validated SNPs was found to be $9.5 \times 10^{-18}$ in Han Chinese. SNP analysis correctly assessed a panel of historical cases, including both paternity identifications in trios and individual identifications. In addition, while STR profiling of degraded DNA provided information for 11 loci of 16 potential markers with low peak intensities, SNPstream® genotyping was sufficient to identify all 47 SNPs. In summary, SNP analysis is equally effective as STR profiling, but appears more suited for individual identification than STR profiling in cases where DNA may be degraded.

**KEYWORDS:** forensic science, DNA profiling, SNP array, individual identification, ABO genotyping, population genetics, heterozygosity, $F_{ST}$

The standard practice in forensic identification, short tandem repeat (STR) technology, uses 13 loci to distinguish one DNA profile from another. The odds that two individuals have matching STR profiles is *c.* 1 in $10^{13}$ (1). Similarly, multiplexed single nucleotide polymorphism (SNP) analysis is a very effective method for DNA profiling. SNP analysis, however, has several advantages over the STR profiling (2–5), including low mutation rates usefulness in paternity testing (6), and the use of short amplicons allowing the amplification of degraded DNA samples (7). The ability to amplify, and thus analyze, the degraded DNA samples is particularly attractive for forensic identification (8).

Individual identifications by SNP analysis require generation of a panel of SNPs that together give an extremely remote probability that two individuals would have the same DNA profile. A multiplex assay panel of 52 SNPs was developed in several European countries (9). Owing to variations in population genetics, match probabilities of an SNP panel can vary greatly among populations (10). A universal panel of 92 SNPs was developed for individual identification (IISNPs), with average heterozygosity at each locus of >0.4 and $F_{ST}$ values <0.06 for 44 major populations across the world (11). Six of the 92 SNPs exhibit strong linkage disequilibrium (LD) values, while the remaining 86 have no significant pairwise LD. The match probability of 42 of the 86 IISNPs was found to be equivalent to that of the 15 loci STR sites found in the commonly used Identifiler® kit (12). As such, 45 of the unlinked IISNPs are recommended for individual identification (11).

The IISNP system was validated for 44 populations; however, mainland Han Chinese were not included in the analysis. Based on the study of Pakstis et al. (11), we describe in this report a multiplexed SNP assay, using Genomelab™ SNPstream® genotyping (13,14), for use in individual identifications of Han Chinese. This system represents an alternative, potentially more sensitive method for forensic applications compared to STR analysis. Forty-two loci were chosen from the pool of 86 unlinked IISNPs, including 37 from among the panel of 45 recommended IISNPs (re-IISNPs). As ABO genotyping is another useful tool in forensic practice for individual identification and paternity testing (15), an additional four ABO gene loci at nucleotide positions 261, 297, 467, and 803 used for classifying ABO genotypes were also chosen between A101, A102, B101, O01, O02, and cisAB01. In addition to the 42 IISNPs and four ABO loci, an amelogenin gender marker was chosen to complete a 47-SNP panel. This panel was arranged into four PCR pools for sample analysis using a 12-plex genotyping system.

## Materials and Methods

### *Sample Collection*

DNA samples from 233 unrelated Han Chinese volunteers (125 men and 108 women) from Beijing were obtained with written informed consent. Each participant provided either a buccal swab ($n$ = 103) or a blood sample ($n$ = 130), and completed a brief questionnaire of general information including sex, age, blood type, and personal and family medical history.

Historical samples were provided by the DNA Department of Forensic Science Institute, Ministry of Public Security (Beijing,

[1]Key Laboratory of Ministry of Public Health for Forensic Science, Department of Forensic Science, School of Medicine, Xi'an Jiaotong University, Xi'an 710061, China.
[2]Key Laboratory of Environment and Genes Related to Diseases, Ministry of Education, Xi'an Jiaotong University, Xi'an 710061, China.
[3]Ministry of Public Security, Institute of Forensic Science, Beijing 100038, China.
[4]Key Laboratory of Forensic Genetics, Ministry of Public Security, Beijing 100038, China.
[5]Faculty of Basic Medical Sciences, Chongqing Medical University, Chongqing 400016, China.
*Supported by Basic Research Project (Grant 2007JB001) awarded by the Institute of Forensic Science, Ministry of Public Security, China.

Received 11 May 2011; and in revised form 8 Aug. 2011; accepted 21 Aug. 2011.

China) for test validation. Samples obtained included 20 DNA samples for paternity identification of trios and 18 forensic DNA samples for individual identifications.

### *DNA Preparation*

Genomic DNA was isolated using the Mini M48 Kit from QIAGEN Corporation (Hilden, Germany), and samples from body fluids, bones, hairs, and epidermal cells were prepared using Chelex-100 or magnetic beads (M48). DNA quantification was performed on 1.5 μL of DNA sample in solution using a NanoDrop® ND-1000 Spectrophotometer (Thermo Fisher Scientific, Wilmington, DE). DNA degradation was performed using 4–5 units of DNase I (Sigma-Aldrich, St. Louis, MO) at 25°C for 20 min and 70°C for 10 min.

### *SNPstream® Genotyping*

SNPstream® genotyping was performed according to the manufacturer's instructions (Beckman Coulter, Brea, CA). All 47 SNPs were arranged into four PCR pools, with one pool of 11-plex primers for typing C/G substitution and three pools of 12-plex primers for A/G. Primer design was performed using the *Autoprimer.com* online service and modified through PRIMER PREMIER 5.0 (PREMIER Biosoft, Palo Alto, CA). The amplicon size ranged from 89 to 158 bp. Details of loci, PCR primers, extension primers, and product sizes are outlined in Table 1. The 45 bp extension primer is composed of a 5′ tag array tail (20 nt) and a 3′ SNP region (25 nt).

Multiplex PCR was carried out in 5 μL volume, containing 3-μL PCR mix with 24 mixed-primers (50 nM each primer, 75 μM dNTPs, 5 mM $MgCl_2$, 0.5 unit of HotStarTaq DNA Polymerase; QIAGEN), and 2-μL DNA template. The amplification was performed in a Mastercycler Pro (Eppendorf AG, Hamburg, Germany). PCR conditions were as follows: predenaturation for 15 min at 95°C; 30 sec at 94°C, 30 sec at 55°C, and 1 min at 72°C for 40 cycles; and 4°C soak. To rid the samples of nucleotide overhangs, a clean-up of the PCR products was achieved using Shrimp Alkaline Phosphatase and Exonuclease I (USB, Cleveland, OH).

The extension procedure was completed with the addition of 7 μL extension mix (3.75 μL extension dilution buffer, 0.2 μL 20× extension mix, 0.02 μL DNA polymerase, 0.03 μL extension primer mix, and 3 μL $ddH_2O$) to the clean-up product utilizing conditions as follows: 96°C for 3 min, 94°C for 20 sec, and 40°C for 11 sec for 46 cycles; hold at 4°C.

Array hybridization was performed on a washed SNPware 12-plex tag array plate. Eight microliter hybridization mix (7.56 μL hybridization solution and 0.44 μL hybridization additive) was added to each well of extension products, and 10 μL of the mixture was added to the array plate, which was later incubated at 42°C for 2 h. Finally, the hybridized plate was washed and centrifuged to dry three times. The dried plate was imaged on a data analysis (Beckman Coulter).

### *STR Kits*

The 15-loci Identifiler® Plus kit, used as the standard STR method for comparison purposes, was obtained from Applied Biosystems (Foster City, CA). PCR conditions were as follows: 95°C for 11 min, 30 cycles of 94°C for 1 min, 59°C for 1 min, 72°C for 1 min; followed by 60°C for 60 min; hold at 4°C. PCR products were transferred to 10 mL of formamide and applied to a 3130xl Genetic Analyzer (Run Module Name: HIDFragmentAnalysis36_POP4_1; GENEMAPPER™ ID3.2 software; Applied Biosystems).

### *DNA Sequencing*

SNP genotyping performed using the SNPstream® assay was compared with the results obtained using dideoxy dye-terminator sequencing chemistry. For each of the 47 loci, eight randomly selected samples were sequenced in both the forward and reverse directions on PCR products obtained using individual primer pairs. PCR was performed using the ABI BigDye® Terminator Cycle Sequencing Kit, performed on a 3130xl DNA Analyzer from Applied Biosystems.

### *Statistical Analysis*

Population indices were calculated, including allele and genotype frequency, heterozygosity, match probability, and success rate of the SNPstream® assay. Hardy–Weinberg equilibrium was computed with HAPLOVIEW genetics software (version 4.2). LD values ($r^2$) (16) for each pair of the 46 autosomal SNPs were computed to evaluate the statistical independence of each locus (loci independence is reflected in pairwise $r^2$ values closest to zero). Haplotype frequencies were analyzed within four ABO genetic markers. An $F_{ST}$ value (17) was calculated for the measurement of the variance in allele frequencies among populations.

## Results

### *System Performance*

DNA concentration was quantified using a NanoDrop® ND-1000 Spectrophotometer. DNA extractions yielded samples with concentrations in the range of 1–30 ng/μL. In our assay, using 2 μL of sample, all 47 markers among the four PCR pools yielded clear genotype clusters. A total of 10,951 possible genotypes from the 233 samples resulted in a call rate of 98.43% (10,779 genotypes receiving a call score).

Assay accuracy was estimated by reproducibility, concordance with sequencing, and consistency with Mendelian inheritance. Five samples were repeated in triplicate. Genotypes at each locus were identical for each of the five samples. Random selection of eight samples for DNA sequencing of all 47 SNPs generated results in complete concordance with those derived from SNPstream® analysis, and with outcomes identified by ABO and gender markers and individual information of blood type and sex.

To test the ability of SNP analysis to positively identify degraded DNA, DNA was degraded using 4–5 U of DNase I, and analysis was performed. While SNP analysis generated complete genotypes of the degraded DNA (as shown in Fig. 1), STR analysis using the Identifiler® Plus kit failed at five of 16 loci. Furthermore, in the 11 detected loci, five demonstrated low peak intensities (ranged from 50 to 200 RFU).

### *Population Indices*

The allele frequency distributions for the SNP panel in the Han Chinese population are outlined in Fig. 2. No meaningful deviations from Hardy-Weinberg equilibrium were observed for any of the 46 autosomal SNPs. Previous data established that a panel of 40–45 SNPs with heterozygosities >0.4 provides match probabilities $<10^{-15}$ (18). Heterozygosities of the 42 IISNPs (listed in Table 1) ranged from 0.3907 to 0.5011 in the training set. There was no significant pairwise LD among the 42 IISNPs. The majority of the pairwise $r^2$ values were <0.03, while three of the pairwise $r^2$ values were <0.04 and two were <0.05. Therefore, allele

TABLE 1—*A panel of 47 SNPs for forensic DNA profiling.*

| PCR Pools | Function | dbSNP rs# | Chr | Nucleotide Position Map Build 37.1 | Alleles | Heterozygosity (Mainland Chinese) | Amplicon Length (bp) | Oligo Type-Mod. | Sequence |
|---|---|---|---|---|---|---|---|---|---|
| CG | ABO | rs8176747 | 9 | 136131315 | 803G>C | – | 109 | PCRU | AGGCCTACATCCCCAAGG |
| | | | | | | | | PCRL | ATCATGGCCTGGTGGCAG |
| | | | | | | | | SNPU | ACGCACGTCCACGGTGATTTCTCTTGCACCGACCCCCCGAAGAAC |
| | re-IISNPs | rs10488710 | 11 | 115207176 | C/G | 0.494835941 | 140 | PCRU | AAAATTAAATAAGGGTACTCATTAACCA |
| | | | | | | | | PCRL | TAAAAGACTTTCAATTTATGTCAGCA |
| | | | | | | | | SNPU | GGATGGCGTTCCGTCCTATTTTGTTCTTTTTCTGCAAATGTGGCA |
| | re-IISNPs | rs445251 | 20 | 15124933 | C/G | 0.451889796 | 103 | PCRU | ATCACACTATCCTGACATGAACAA |
| | | | | | | | | PCRL | ACCTGGCCTACAATTCAAATTA |
| | | | | | | | | SNPL | GCGGTAGGTTCCCGACATATATTAATGTAAAAACTGCAAGTGGTT |
| | re-IISNPs | rs722290 | 14 | 53216723 | C/G | 0.500990756 | 121 | PCRU | ACCAGAAGTGTTTCAGATTTCAG |
| | | | | | | | | PCRL | AGTATATATGGTGTGTGTGAAACATAAAT |
| | | | | | | | | SNPU | GACCTGGGTGTCGATACCTATTTGAAATATTTGCATATACATACT |
| | re-IISNPs | rs1821380 | 15 | 39313402 | C/G | 0.458396788 | 151 | PCRU | TTTTCCTCTTTTCCAGTGCA |
| | | | | | | | | PCRL | TATGGAACCTCTGCTGTCTTTT |
| | | | | | | | | SNPU | AGAGCGAGTGACGCATACTAACTGCAGTGCAAAAATGCAGTAAGG |
| | re-IISNPs | rs2269355 | 12 | 6945914 | C/G | 0.501066039 | 90 | PCRU | TCTCAGAGGCCCCCTTAC |
| | | | | | | | | PCRL | CTCTCTGGTACGGGCTGG |
| | | | | | | | | SNPU | CGACTGTAGGTGCGTAACTCTGTAGGAAGCTCTCCCGAGTTCTCT |
| | IISNPs | rs2272998 | 6 | 148761456 | C/G | 0.483206424 | 93 | PCRU | TGCCTTTTTTTTTTAAGTGACAC |
| | | | | | | | | PCRL | CGACTCCTACGAGAGAAGATTC |
| | | | | | | | | SNPU | CGTGCCGCTCGTGATAGAATCTGTACAAATCAGATGAAGCCTGCT |
| | IISNPs | rs5746846 | 22 | 19920646 | C/G | 0.45146487 | 120 | PCRU | AAAGCTTGGGAAGCTGCA |
| | | | | | | | | PCRL | GCACACATCTCACGTGCA |
| | | | | | | | | SNPU | GGCTATGATTCGCAATGCTTTGCCCCACAACCAAACCCACCCACT |
| | IISNPs | rs7205345 | 16 | 7520254 | C/G | 0.462079468 | 158 | PCRU | TTAATCTGACTGTATACCCCAGG |
| | | | | | | | | PCRL | GGAAACAATGGTGAACAACTTAC |
| | | | | | | | | SNPU | AGGGTCTCTACGCTGACGATAAGAGCTGATTTCTGTGTCTGCCT |
| | IISNPs | rs3744163 | 17 | 80739859 | C/G | 0.492768471 | 96 | PCRU | TTAAACAGGAACTAAGCAGCC |
| | | | | | | | | PCRL | AGCCCCACCTGTTGCCAG |
| | | | | | | | | SNPL | GTGATTCTGTACGTGTCGCCCAGAGGGCAGTGGTGCCACGCAGCC |
| | IISNPs | rs521861 | 18 | 47371014 | C/G | 0.50062301 | 154 | PCRU | GCATCCTTCTAGTTAAATGATAACTGTA |
| | | | | | | | | PCRL | CCTACAAAATACATTTTCCTGGATA |
| | | | | | | | | SNPU | AGATAGAGTCGATGCCAGCTTCTTGAGTACATGGTTGACATTTGG |
| AG1 | re-IISNPs | rs7520386 | 1 | 14155402 | A/G | 0.429599889 | 114 | PCRU | ATTTAGCAACCTGCTGAGTGA |
| | | | | | | | | PCRL | AAGAAGACTCTGTCCCAGCC |
| | | | | | | | | SNPU | ACGCACGTCCACGGTGATTTCCAAGGGAACGTGAGGAGGCCACAC |
| | re-IISNPs | rs1736442 | 18 | 55225777 | A/G | 0.475289052 | 121 | PCRU | ACTTGAAGATATATATAAGACTTTCTAATAAGTGG |
| | | | | | | | | PCRL | AATGGAGACCCCCCCACT |
| | | | | | | | | SNPU | GGATGGCGTTCCGTCCTATTGAAGGCTGCTTTTGCCTGCCCTGTC |
| | re-IISNPs | rs10773760 | 12 | 130761696 | A/G | 0.470828659 | 154 | PCRU | TTCTCATTTTCGGTAGACTGC |
| | | | | | | | | PCRL | ATCATCACAGCCATCACTGTC |
| | | | | | | | | SNPU | CGTGCCGCTCGTGATAGAATCTGGAAGTTCGTCAAATTGCAGTTA |
| | re-IISNPs | rs13218440 | 6 | 12059954 | A/G | 0.497004938 | 112 | PCRU | TGAATCTCTTGAGTCTTTTTCTGC |
| | | | | | | | | PCRL | AAGCAGCCCCAACCCATC |
| | | | | | | | | SNPU | AGCGATCTGCGAGACCGTATCTGAGATTCACCTCTAGTCCCTCTG |
| | re-IISNPs | rs560681 | 1 | 160786670 | A/G | 0.447764087 | 142 | PCRU | CAACCCTGTCAGCAGGAG |
| | | | | | | | | PCRL | TTACTGCACTTCATAAAGAATCAGTC |
| | | | | | | | | SNPU | GCGGTAGGTTCCCGACATATTCTGTTCAGGTTTCTCTCCATCTCT |
| | re-IISNPs | rs221956 | 21 | 43606997 | C/T | 0.495758918 | 112 | PCRU | TCCATGCTAGAAAAAGCTGAG |

*Continued.*

TABLE 1—*Continued.*

| PCR Pools | Function | dbSNP rs# | Chr | Nucleotide Position Map Build 37.1 | Alleles | Heterozygosity (Mainland Chinese) | Amplicon Length (bp) | Oligo Type-Mod. | Sequence |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | PCRL | TCCTCTGAGATGATGAATGCT |
| | | | | | | | | SNPU | GGCTATGATTCGCAATGCTTAGCTGTGAAGGAAGCACTAAAGTCA |
| | re-IISNPs | rs4530059 | 14 | 104769149 | A/G | 0.444921316 | 89 | PCRU | AGTCGATGATGACTTCCCA |
| | | | | | | | | PCRL | GCTTCCTCAGTAGCAGCAC |
| | | | | | | | | SNPU | AGGGTCTCTACGCTGACGATAGCTCCAGAAGCAACTCCAGCACAC |
| | re-IISNPs | rs338882 | 5 | 178690725 | C/T | 0.465596013 | 122 | PCRU | AGCCAAAGGGACCAAGTC |
| | | | | | | | | PCRL | TTCATTTTTCTCTCCTTCTGTCTC |
| | | | | | | | | SNPU | GTGATTCTGTACGTGTCGCCTCCAGCCCTTGTCCCAAACGTGTGT |
| | re-IISNPs | rs430046 | 16 | 78017051 | C/T | 0.443472241 | 152 | PCRU | AGTCAAGGCATTTGACCATC |
| | | | | | | | | PCRL | AATCTTAAATGTTAAGGAACAAGGTC |
| | | | | | | | | SNPU | GACCTGGGTGTCGATACCTACCATTAAACCCAGCACCTACCCTCA |
| | re-IISNPs | rs13182883 | 5 | 136633338 | A/G | 0.501066039 | 90 | PCRU | TCTGGCCTAGTAGAGGGCC |
| | | | | | | | | PCRL | TTACTTTCTTCCTGCCTTTGTC |
| | | | | | | | | SNPU | AGATAGAGTCGATGCCAGCTGCCTGCAGTGAGCATTCAAATCCTC |
| | re-IISNPs | rs9951171 | 18 | 9749879 | A/G | 0.493839125 | 132 | PCRU | TCTGTCCTGTAGGGGAGAAA |
| | | | | | | | | PCRL | AATTTTCTTGTCCCTGCTTTC |
| | | | | | | | | SNPU | AGAGCGAGTGACGCATACTAAGAGCAGCACACTGAGGCTTTATGG |
| | re-IISNPs | rs1109037 | 2 | 10085722 | A/G | 0.496569464 | 135 | PCRU | TAAGAAACTTTGTCACAGAGCTG |
| | | | | | | | | PCRL | CAGTTCATCAGAACCCGG |
| | | | | | | | | SNPU | CGACTGTAGGTGCGTAACTCCCCTGCCTCCCACACCAGTTTCTCC |
| AG2 | ABO | rs8176719 | 9 | 136132908:136132909 | 261del(A)>G | – | 91 | PCRU | ACCGCACGCCTCTCTCCAT |
| | | | | | | | | PCRL | ATGTTGAATGTGCCCTCC |
| | | | | | | | | SNPU | AGAGCGAGTGACGCATACTAGCAGTAGGAAGGATGTCCTCGTGGT |
| | ABO | rs8176720 | 9 | 136132873 | 297A>G | – | 92 | PCRU | ATGTCCTCGTGGTACCCC |
| | | | | | | | | PCRL | TGTTCTGGAGCCTGAACTG |
| | | | | | | | | SNPU | GTGATTCTGTACGTGTCGCCTGGCTCCCATTGTCTGGGAGGGCAC |
| | SEX | Amelogenin | XY | chrX: 9198433 | C | – | 127 | PCRU | ATGATTTGGTGGTGCAGC |
| | | | | chrY: 4086980 | T | | | PCRL | AAAGAAAACACTGCTGCTTCTC |
| | | | | | | | | SNPU | AGGGTCTCTACGCTGACGATCCCATGGGCTCGTAACCATAGGAAG |
| | re-IISNPs | rs159606 | 5 | 17374898 | A/G | 0.480332118 | 141 | PCRU | TGTTATTTTTGGTTGTGAGCAC |
| | | | | | | | | PCRL | ATATGCCCCCATTATAACATTG |
| | | | | | | | | SNPU | ACGCACGTCCACGGTGATTTTCTCATCCTGTTATTATTTGTTTAC |
| | re-IISNPs | rs321198 | 7 | 137029838 | C/T | 0.499958466 | 102 | PCRU | TCCACACTTTATACAGGTGAAATCT |
| | | | | | | | | PCRL | TACACACAGGCTTCAGGTTACC |
| | | | | | | | | SNPU | GGATGGCGTTCCGTCCTATTAAAACAAGCACTGCTTGCTTCACAC |
| | re-IISNPs | rs4606077 | 8 | 144656754 | C/T | 0.499746181 | 89 | PCRU | TGTGCTGAAGCCCGAAGG |
| | | | | | | | | PCRL | TGTGGCTTTGTAGTTCTAGTGTG |
| | | | | | | | | SNPU | CGTGCCGCTCGTGATAGAATAGTGAGGAGTTCCCCGGCTTTGGGT |
| | re-IISNPs | rs1336071 | 6 | 94537255 | A/G | 0.496569464 | 113 | PCRU | CACCTATATATTATACCTGAAAGCATATTAAA |
| | | | | | | | | PCRL | TGTTTTGTCCATCTGAAATTCTAA |
| | | | | | | | | SNPU | AGCGATCTGCGAGACCGTATAAGAATGAATTTGAAAAAGCATCAG |
| | re-IISNPs | rs1294331 | 1 | 233448413 | A/G | 0.423314277 | 142 | PCRU | TTTGGAGTTTTATGTGTTGCTACA |
| | | | | | | | | PCRL | AAATTCTCTGTATGAAGGTAGCATAAA |
| | | | | | | | | SNPU | GCGGTAGGTTCCCGACATATTCATCATTCATAATTGTGATTGAGT |
| | re-IISNPs | rs3780962 | 10 | 17193346 | C/T | 0.484793945 | 154 | PCRU | ATAATCCTCAAAAACAAAGAAACATG |
| | | | | | | | | PCRL | ATTGATCCTTCCAAACCCTC |
| | | | | | | | | SNPU | GGCTATGATTCGCAATGCTTGGTCAAGATATCAGCTTTCACCCGT |
| | re-IISNPs | rs9905977 | 17 | 2919393 | A/G | 0.490937208 | 111 | PCRU | ATAAGGAAAATTCATGAGCTGGT |
| | | | | | | | | PCRL | AAAGCTGACAAGCCAGGG |
| | | | | | | | | SNPU | GACCTGGGTGTCGATACCTACTGGGTGACTCGTGGCTCAGTCAGC |
| | re-IISNPs | rs1058083 | 13 | 100038233 | A/G | 0.489768794 | 122 | PCRU | AATGGTGATAACCATGCACTG |

*Continued.*

TABLE 1—*Continued.*

| PCR Pools | Function | dbSNP rs# | Chr | Nucleotide Position Map Build 37.1 | Alleles | Heterozygosity (Mainland Chinese) | Amplicon Length (bp) | Oligo Type-Mod. | Sequence |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | PCRL | TTGTTCTTCTCCATCCCATTT |
| | | | | | | | | SNPU | AGATAGAGTCGATGCCAGCTTGGGTTAATTTTGCTCAGAGTATCC |
| | re-IISNPs | rs740598 | 10 | 118506899 | A/G | 0.483206424 | 139 | PCRU | TAAACTTTAGAAATGCCTTCTCAGG |
| | | | | | | | | PCRL | AGCTTTGCAGGGCCAGGG |
| | | | | | | | | SNPU | CGACTGTAGGTGCGTAACTCCAATGGCTCGTCTATGGTTAGTCTC |
| AG3 | ABO | rs1053878 | 9 | 136131651 | 467C>T | – | 150 | PCRU | GGAGACGGCGGAGAAGCACTTCA |
| | | | | | | | | PCRL | ACGTCCTGCCAGCGCTTGTAGGC |
| | | | | | | | | SNPU | GGATGGCGTTCCGTCCTATTCAGCGTCACGCGGGGCACCGCGGCC |
| | re-IISNPs | rs8078417 | 17 | 80461935 | C/T | 0.390718029 | 140 | PCRU | TTCTGAGCCACGTGCATC |
| | | | | | | | | PCRL | TAGAGAACTCTGTACGTGGTCACC |
| | | | | | | | | SNPU | ACGCACGTCCACGGTGATTTTGTGAGAAGAGCCTCAAGGACAGCC |
| | re-IISNPs | rs2342747 | 16 | 5868700 | A/G | 0.451889796 | 90 | PCRU | AAAAAGTTGTCAGCATGGGA |
| | | | | | | | | PCRL | CACTCCTAATCATACCTTGAAGCT |
| | | | | | | | | SNPU | CGTGCCGCTCGTGATAGAATGAAGAAAACAGAGAGTCTTGACCGT |
| | re-IISNPs | rs10092491 | 8 | 28411072 | C/T | 0.456368098 | 152 | PCRU | AACCAGTATCCCCGCAAA |
| | | | | | | | | PCRL | TGAGGAAAGGACTTGAATTAAACA |
| | | | | | | | | SNPU | AGCGATCTGCGAGACCGTATACTAGGATAAATCTCTATAAGGAAA |
| | re-IISNPs | rs6444724 | 3 | 193207380 | C/T | 0.479626149 | 113 | PCRU | AATCAGGAAATAGTCACTTCCTACTC |
| | | | | | | | | PCRL | ATTGGAATGGGAGGAAAGG |
| | | | | | | | | SNPU | GCGGTAGGTTCCCGACATATGAGCAAGTTTGTAAATACCTAGCAC |
| | re-IISNPs | rs1498553 | 11 | 5709028 | C/T | 0.496192718 | 157 | PCRU | AATGAACTCATTTATCCAGAGACC |
| | | | | | | | | PCRL | AATTAATACTTCATGAGATAAATCAGACC |
| | | | | | | | | SNPU | GGCTATGATTCGCAATGCTTATCTCATAGCCATCCTTTTATTCTC |
| | re-IISNPs | rs12997453 | 2 | 182413259 | A/G | 0.45053302 | 94 | PCRU | TTGTCAGAATGTTTTATGCTTTAA |
| | | | | | | | | PCRL | CATTTAACAGCTCTGATGATGTG |
| | | | | | | | | SNPU | AGGGTCTCTACGCTGACGATGATACAGGTTATCTGTATTACATTG |
| | re-IISNPs | rs7041158 | 9 | 27985938 | C/T | 0.481205415 | 132 | PCRU | AATTTCTGAGAATAACATTGCCTC |
| | | | | | | | | PCRL | TTGTTATTTTGTCATTCTGATGGA |
| | | | | | | | | SNPU | GTGATTCTGTACGTGTCGCCCCAGTGAGAAGTGTCTTGGGTTGGA |
| | re-IISNPs | rs214955 | 6 | 152697706 | A/G | 0.486307628 | 141 | PCRU | ACATTCTAAGAACTGGTGATTCTATCA |
| | | | | | | | | PCRL | TGTAAAATCTGTTCTTTATTCTGGC |
| | | | | | | | | SNPU | GACCTGGGTGTCGATACCTACTTTTTCCTGCCATTAAATTTTTGC |
| | re-IISNPs | rs6955448 | 7 | 4310365 | C/T | 0.429599889 | 93 | PCRU | AGTTCTTTTCTCCGGGCTA |
| | | | | | | | | PCRL | ATAGGTGATGCAAAGCCCT |
| | | | | | | | | SNPU | AGATAGAGTCGATGCCAGCTCTAGTTGCGTTTACAACTTTCTCCC |
| | re-IISNPs | rs993934 | 2 | 124109213 | C/T | 0.500152291 | 97 | PCRU | AATAGAGCAAAGTATTGTGATAACA |
| | | | | | | | | PCRL | ATTTTTCCCATGATGAAACAGT |
| | | | | | | | | SNPU | AGAGCGAGTGACGCATACTATCTCCAGAGTATATTAGCTTAGTTC |
| | re-IISNPs | rs1523537 | 20 | 51296162 | C/T | 0.481545064 | 90 | PCRU | GTGAGACAATGCACAGAACT |
| | | | | | | | | PCRL | CTGCATGGGTGGGGTTTC |
| | | | | | | | | SNPU | CGACTGTAGGTGCGTAACTCTGATCACCTAATAGCCAGCGATAGC |

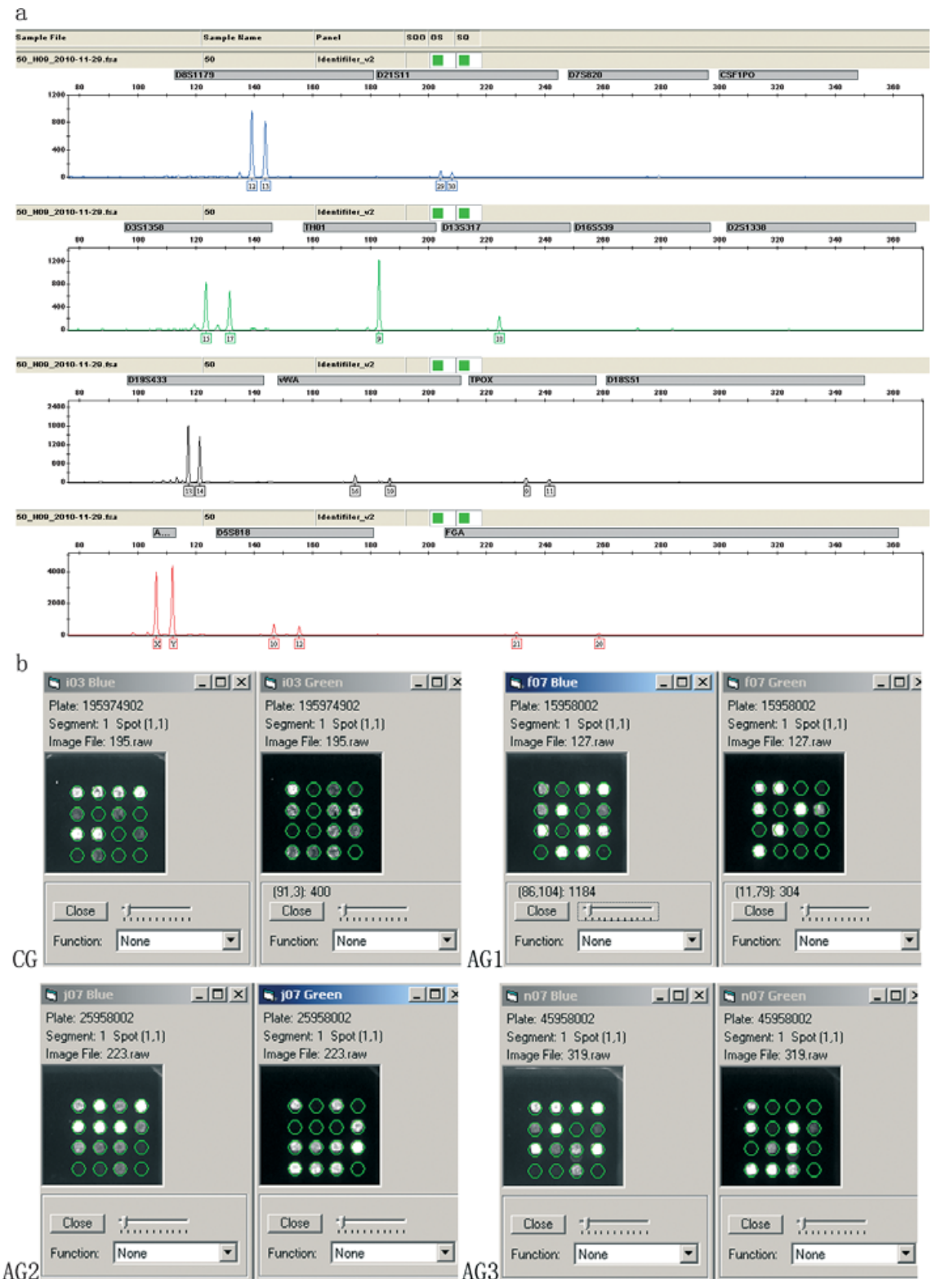


FIG. 1—*The STR profile and SNP analysis of degraded DNA. DNA was extracted from blood, and degraded with 4–5 units of DNase I at 25°C for 20 min and 70°C for 10 min. (a) STR analysis of degraded DNA resulted in five STR loci dropouts. (b) SNPstream® genotyping was performed on 233 samples. Blue and green fluorescence from each spot, respectively, represents each of two alleles of an SNP marker. Controls of hybridization are indicated in the four corner spots: three positive (top-right for genotype 1/1 (Blue-on, Green-off), top-left for 1/2 (Blue-on, Green-on), bottom-left for 2/2 (Blue-off, Green-on)) and one negative (bottom-right; Blue-off, Green-off). Eleven SNP markers were included in the CG panel, with one spot left blank.*

frequencies noted for each SNP can be multiplied generating a match probability for the 42 statistically independent IISNPs of $9.5 \times 10^{-18}$.

Each of the four ABO markers used in this panel is found within the chromosomal region 9q34.2. Allele and genotype frequencies are outlined in Table 2. Allele cisAB01 was not detected in this

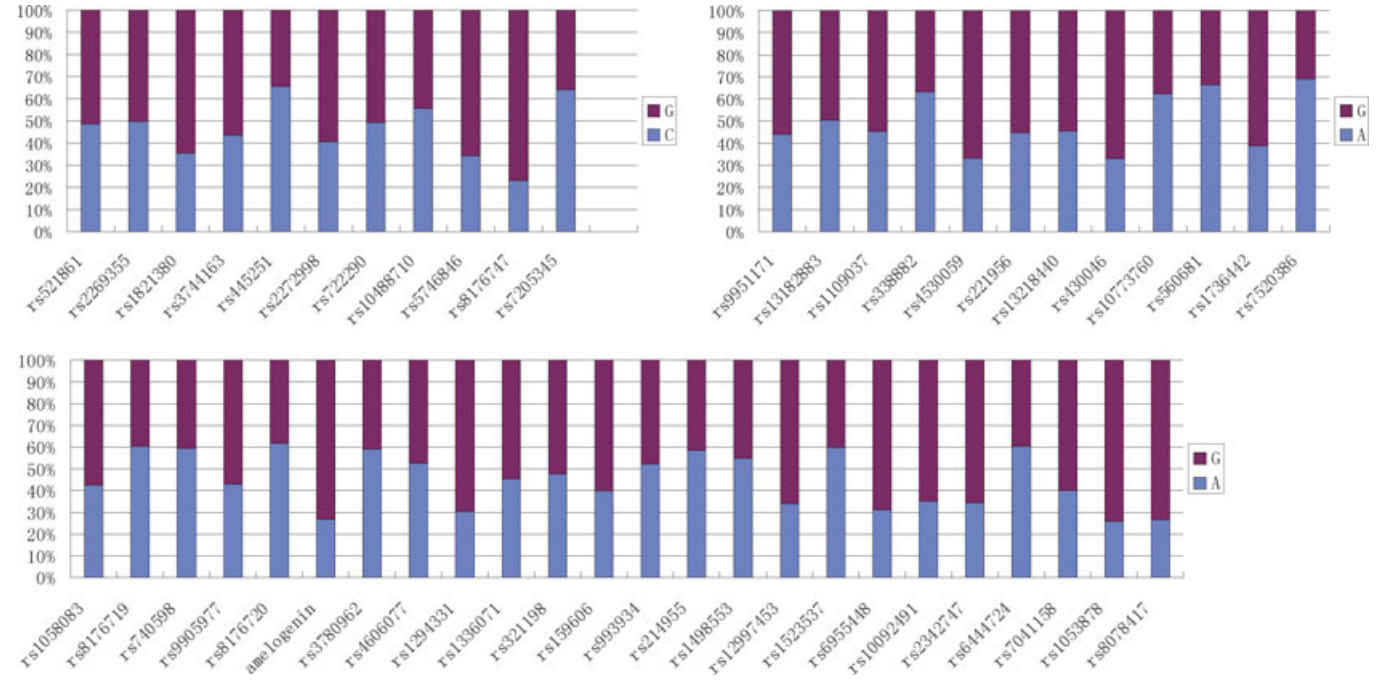


FIG. 2—*Allele frequencies of 47 SNPs in Han Chinese. Analysis of DNA by SNPstream® genotyping was performed according to the manufacturer's instructions. Data indicates population frequency of 47 alleles derived from 233 unrelated Han Chinese samples.*

TABLE 2—*Allele and genotype frequencies of ABO loci in Han Chinese.*

| | | | | Loci Position | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Genotypes | Count | Frequency (%) | Alleles | 261 | 297 | 467 | 803 | Count | Frequency (%) |
| A101/O01 | 4 | 1.78 | O01 | A | A | G | C | 160 | 35.56 |
| A101/O02 | 2 | 0.89 | O02 | A | G | G | C | 80 | 17.78 |
| A102/A101 | 2 | 0.89 | A102 | G | A | A | C | 100 | 22.22 |
| A102/A102 | 12 | 5.33 | A101 | G | A | G | C | 11 | 2.44 |
| A102/O01 | 26 | 11.56 | B101 | G | G | G | G | 99 | 22.00 |
| A102/O02 | 24 | 10.67 | cisAB01* | G | A | T | C | – | – |
| B101/A101 | 3 | 1.33 | | | | | | | |
| B101/A102 | 24 | 10.67 | | | | | | | |
| B101/B101 | 8 | 3.56 | | | | | | | |
| B101/O01 | 51 | 22.67 | | | | | | | |
| B101/O02 | 5 | 2.22 | | | | | | | |
| O01/O01 | 18 | 8.00 | | | | | | | |
| O01/O02 | 43 | 19.11 | | | | | | | |
| O02/O02 | 3 | 1.33 | | | | | | | |
| SUM | 225 | 100 | | | | | | 450 | 100 |

*Allele cisAB01 was not detected in the population examined.

experiment. In statistical analysis, no significant difference ($p = 0.353$, $\alpha = 0.05$; $\chi^2$ test) was found in ABO allele frequencies between Han Chinese and Korean populations (Korean data were obtained from a published report by Lee et al. [15]).

$F_{ST}$ values were calculated between our training sample set and the data from 11 populations (ASW: African ancestry in Southwest U.S.A.; CEU: Utah residents with Northern and Western European ancestry from the CEPH collection; CHB: Han Chinese in Beijing, China; CHD: Chinese in Metropolitan Denver, Colorado; GIH: Gujarati Indians in Houston, Texas; JPT: Japanese in Tokyo, Japan; LWK: Luhya in Webuye, Kenya; MEX: Mexican ancestry in Los Angeles, California; MKK: Maasai in Kinyawa, Kenya; TSI: Tuscan in Italy; YRI: Yoruban in Ibadan, Nigeria) available on the HapMap database (release #28) (19). $F_{ST}$ values for the 42 IISNPs over 12 populations ranged from 0.002 to 0.080.

### *Case Sample Analysis*

Multiplexed SNP analysis was used to test a panel of samples from archived forensic or paternity cases. Each case was provided with Identifiler® genotypes, allowing a comparison of results between STR and SNP DNA profiling (details in Table 3). Multiplexed SNP was equally effective compared with STR analysis for both forensic individual identifications and for paternity testing. Each method correctly identified all forensic cases. Similarly, SNP analysis correctly performed paternity identifications of 20 trio family paternity tests, including three cases in which paternity was

TABLE 3—*Identity results of nine cases analyzed by both STR and SNP methods.*

| Cases | Sample from Crime Scene | Blood Sample from Suspect/Victim | Identity (Yes/No) | |
|---|---|---|---|---|
| | | | Identifilier® Plus | SNPstream® |
| 1 | Blood | Victim | Y | Y |
| 2 | Blood | Victim | Y | Y |
| 3 | Semen | Suspect | Y | Y |
| 4 | Blood | Suspect | N | N |
| 5 | Semen | Suspect | Y | Y |
| 6 | Blood | Victim | Y | Y |
| 7 | Semen | Suspect | Y | Y |
| 8 | Saliva | Suspect | Y | Y |
| 9 | Semen | Suspect | N | N |

excluded (genotypes of loci by STR and SNP analysis revealed that the parent and child did not follow Mendel's laws; Table 4).

## Discussion

We have established a multiplex SNP detection system that is very useful for human identification and ABO blood type analysis. Based on a 12-plex SNPstream® system, 42 IISNPs, four ABO, and one gender markers were arranged into four SNP panels, including one for detecting C/G substitutions and three for detecting A/G substitutions. In these 42 IISNPs, 37 re-IISNPs were chosen from the recommended set of 45 unlinked IISNPs, and five were selected from the remaining list of 41 markers. These 42 IISNPs had been tested with no significant LD between any pair in 44 populations. These findings were validated in our training samples as all pairwise $r^2$ were determined to be <0.05. As a result, each IISNP could be statistically independent for calculating a match probability using the panel. Previous data established that a panel of 40 to 45 SNPs with heterozygosities >0.4 could provide match probabilities $<10^{-15}$ (18). Using our SNP panel, the match probability reached $9.5 \times 10^{-18}$ in the training set, which is comparable with that obtained with the 15-STR Identifiler® kit. Furthermore, in light of the 52-SNP panel reported by Sanchez et al. (9), in which a mean match probability was calculated to be $10^{-18}$ in the populations studied, our panel with less markers was efficient for human identification in Han Chinese.

A significant problem with SNPs in forensic identification is that the frequency of an allele can vary dramatically among different populations, causing very large variations in match probability (20). Using a low $F_{ST}$ strategy for identifying IISNPs can minimize differences in allele frequency among populations, as the match probability would be nearly constant irrespective of the population (10). Low $F_{ST}$ values tested with our training samples, and 11 worldwide populations further validate this SNP panel for universal use in forensic human identification.

Different typing technologies can be used to establish the multiplexed SNP detection systems, that is, SNaPshot® (21) and GenPlex™ (12). SNPstream® is a high-throughput and automated system, which has been described as one of the SNP typing strategies allowing possible identification on some highly degraded samples because of the small amplicon size (13). In our design, all amplicons were between 89 and 158 bp, and the benefit of the system for detecting highly degraded samples was shown by making a comparison with Identifiler® Plus. Using degraded DNA samples, when the STR kit failed to amplify the alleles of larger sizes, the SNP panel, on the other hand, provided a complete genotype of the 47 SNPs. As such, compared to the Identifiler® Plus kit, our SNP

TABLE 4—*Genotypes of loci by STR and SNP analysis that the parent and child do not follow Mendel's laws.*

| Case | STR Loci | | | | | | | | | | | SNP Markers | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Case 1 | D8S1179 | CSF1PO | D3S1358 | TH01 | D13S317 | D16S539 | D2S1338 | D19S433 | vWA | TPOX | D18S51 | rs 521861 | rs 1821380 | rs 13182883 | rs 1109037 | rs 13218440 | rs 7520386 | rs 9905977 | rs 8176720 | rs 4606077 | rs 2342747 | | | | | | | | | | | |
| Father | 12,15 | 11,11 | 16,16 | 6,6 | 10,12 | 11,11 | 19,19 | 14,14 | 17,17 | 11,11 | 21,21 | CC | GG | GG | GG | GG | AA | AA | GG | GG | AA | | | | | | | | | | | |
| Mother | 11,13 | 10,12 | 15,17 | 6,7 | 10,10 | 11,12 | 19,23 | 13,14.2 | 16,17 | 11,11 | 17,18 | CC | GG | GA | GA | GA | AA | AA | GA | GG | GA | | | | | | | | | | | |
| Child | 11,14 | 10,12 | 15,15 | 7,9 | 9,10 | 8,12 | 23,24 | 13,16.2 | 14,16 | 8,11 | 13,18 | CG | CG | AA | AA | AA | GA | GA | AA | GA | GG | | | | | | | | | | | |
| Case 2 | D21S11 | D7S820 | D3S1358 | D16S539 | vWA | FGA | | | | | | rs 10488710 | rs 338882 | rs 430046 | rs 7520386 | rs 740598 | rs 9905977 | rs 4606077 | rs 159606 | rs 1498553 | rs 10092491 | rs 2342747 | rs 8078417 | | | | | | | | | |
| Father | 29,30 | 11,12 | 15,17 | 10,10 | 14,16 | 21,25 | | | | | | GG | AA | AA | GG | AA | AA | AA | GG | GG | GG | GG | GG | | | | | | | | | |
| Mother | 29,32 | 11,12 | 17,17 | 13,13 | 15,18 | 18,18 | | | | | | CG | AA | GA | GG | AA | GA | AA | GG | GG | GA | GA | GA | | | | | | | | | |
| Child | 32,35 | 8,12 | 16,17 | 12,13 | 18,20 | 18,22 | | | | | | CC | GA | GG | GA | GA | GG | GA | GA | GA | AA | AA | AA | | | | | | | | | |
| Case 3 | D21S11 | D7S820 | CSF1PO | TH01 | D16S539 | D2S1338 | TPOX | D18S51 | D5S818 | FGA | | rs 521861 | rs 445251 | rs 722290 | rs 9951171 | rs 1109037 | rs 1058083 | rs 740598 | rs 9905977 | rs 3780962 | rs 321198 | rs 159606 | rs 993934 | rs 214955 | rs 1523537 | rs 6955448 | | | | | | |
| Father | 31,32.2 | 11,11 | 12,12 | 9,9 | 11,13 | 20,20 | 8,9 | 14,19 | 11,11 | 23,25 | | GG | CC | GG | AA | AA | GG | AA | GG | AA | AA | GG | GG | GG | GG | GG | | | | | | |
| Mother | 30,30.3 | 8,11 | 10,11 | 6,9 | 10,12 | 18,24 | 9,11 | 16,16 | 12,12 | 23,23 | | GG | CG | CC | AA | GG | GA | GA | AA | GA | AA | GG | AA | AA | AA | AA | | | | | | |
| Child | 29,30.3 | 11,12 | 9,10 | 7,9 | 12,12 | 19,24 | 11,11 | 16,18 | 10,12 | 23,24 | | CG | GG | CC | GA | GG | AA | GG | AA | GA | GA | GA | AA | AA | AA | AA | | | | | | |

SNP, single nucleotide polymorphism; STR, short tandem repeat.

panel was effective in analyzing degraded DNA templates. Because all PCR primers are automatically generated by the http://www.Autoprimer.com website, in the future, additional work to narrow down the amplicon size may be considered in primer design. This may result in significant increase in the performance efficiency of our system for degraded samples (22).

In addition to individual identification, the 47-SNP panel can also be useful for the determination of ABO genotypes. The panel includes four markers that are commonly used in cases of ABO discrepancy between cell typing and serum typing, as well as in forensic practice for personal identification and paternity test (15). Interestingly, no significant difference was observed in ABO allele frequencies between Han Chinese and Korean. We attribute this result to the close geographic distance between these two populations. The ABO data obtained in this study could also be used as a fundamental reference for Han Chinese.

Case sample analysis confirmed the validation of the multiplexed SNP assay for both forensic individual identifications and for paternity testing. While SNP analysis performed equally well as STR typing in these cases, we believe that SNP analysis may actually perform better than STR typing in more difficult cases where DNA integrity is low. We hope to validate this hypothesis with cases where STR typing fails in the future.

## Conclusions

We established a multiplexed SNP assay system, which equaled the performance of the Identifiler® kit in match probability. Further, SNP analysis was more effective than STR analysis to profile DNA from a degraded sample. In summary, this SNP DNA profiling method represents an ideal alternative strategy to standard STR analysis for forensic individual identifications, particularly for the analysis of degraded DNA samples.

*Acknowledgment*

The authors gratefully acknowledge B.-W. Jiang for his help in DNA preparation.

## References

1. Einum DD, Scarpetta MA. Genetic analysis of large data sets of North American Black, Caucasian, and Hispanic populations at 13 CODIS STR loci. J Forensic Sci 2004;49(6):1381–5.
2. Amorim A, Pereira L. Pros and cons in the use of SNPs in forensic kinship investigation: a comparative analysis with STRs. Forensic Sci Int 2005;150(1):17–21.
3. Gill P, Werrett DJ, Budowle B, Guerrieri R. An assessment of whether SNPs will replace STRs in national DNA databases—joint considerations of the DNA working group of the European Network of Forensic Science Institutes (ENFSI) and the Scientific Working Group on DNA Analysis Methods (SWGDAM). Sci Justice 2004;44(1):51–3.
4. Kim JJ, Han BG, Lee HI, Yoo HW, Lee JK. Development of SNP-based human identification system. Int J Legal Med 2010;124(2):125–31.
5. Gill P. An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. Int J Legal Med 2001;114(4–5):204–10.
6. Reich DE, Schaffner SF, Daly MJ, McVean G, Mullikin JC, Higgins JM, et al. Human genome sequence variation and the influence of gene history, mutation and recombination. Nat Genet 2002;32(1):135–42.
7. Sanchez JJ, Endicott P. Developing multiplexed SNP assays with special reference to degraded DNA templates. Nat Protoc 2006;1(3):1370–8.
8. Vallone PM, Just RS, Coble MD, Butler JM, Parsons TJ. A multiplex allele-specific primer extension assay for forensically informative SNPs distributed throughout the mitochondrial genome. Int J Legal Med 2004;118(3):147–57.
9. Sanchez JJ, Phillips C, Borsting C, Balogh K, Bogus M, Fondevila M, et al. A multiplex assay with 52 single nucleotide polymorphisms for human identification. Electrophoresis 2006;27(9):1713–24.
10. Kidd KK, Pakstis AJ, Speed WC, Grigorenko EL, Kajuna SL, Karoma NJ, et al. Developing a SNP panel for forensic identification of individuals. Forensic Sci Int 2006;164(1):20–32.
11. Pakstis AJ, Speed WC, Fang R, Hyland FC, Furtado MR, Kidd JR, et al. SNPs for a universal individual identification panel. Hum Genet 2010;127(3):315–24.
12. Fang R, Pakstis AJ, Hyland FC, Wang D, Shewale JG, Kidd JR, et al. Multiplexed SNP detection panels for human identification. Forensic Sci Int Genet Suppl Ser 2009;2(1):538–9.
13. Budowle B. SNP typing strategies. Forensic Sci Int 2004;146 (Suppl.):S139–42.
14. Bell PA, Chaturvedi S, Gelfand CA, Huang CY, Kochersperger M, Kopla R, et al. SNPstream UHT: ultra-high throughput SNP genotyping for pharmacogenomics and drug discovery. Biotechniques 2002;2:74.
15. Lee SH, Park G, Yang YG, Lee SG, Kim SW. Rapid ABO genotyping using whole blood without DNA purification. Korean J Lab Med 2009;29(3):231–7.
16. Devlin B, Risch N. A comparison of linkage disequilibrium measures for fine-scale mapping. Genomics 1995;29(2):311–22.
17. Hartl DL, Clark AG. Principles of population genetics, 4th edn. Sunderland, MA: Sinauer Associates, 2007.
18. Pakstis AJ, Speed WC, Kidd JR, Kidd KK. Candidate SNPs for a universal individual identification panel. Hum Genet 2007;121 (3–4):305–17.
19. Thorisson GA, Smith AV, Krishnan L, Stein LD. The International HapMap Project web site. Genome Res 2005;15(11):1592–3.
20. Lewontin RC, Hartl DL. Population genetics in forensic DNA typing. Science 1991;254(5039):1745–50.
21. Phillips C, Salas A, Sanchez JJ, Fondevila M, Gomez-Tato A, Alvarez-Dios J, et al. Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. Forensic Sci Int Genet 2007;1 (3–4):273–80.
22. Wiegand P, Kleiber M. Less is more—length reduction of STR amplicons using redesigned primers. Int J Legal Med 2001;114(4–5):285–7.

Additional information and reprint requests:
Dr. Yao Liu, Ph.D.
Ministry of Public Security
Institute of Forensic Science
Muxidi Nanli
Xicheng District
Beijing 100038
China
E-mail: liuyao1123@yahoo.cn

Dr. Lan Hu, Ph.D.
Ministry of Public Security
Institute of Forensic Science
Muxidi Nanli
Xicheng District
Beijing 100038
China
E-mail: hulan328@yahoo.com

<Correction added after online publication 26 April 2012: Two corresponding authors worked on this paper, but Dr. Lan Hu's address was mistakenly omitted from the original version of this paper. It has been added above.>