JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, July 2012, Vol. 57, No. 4
doi: 10.1111/j.1556-4029.2012.02110.x
Available online at: onlinelibrary.wiley.com

TECHNICAL NOTE

PATHOLOGY AND BIOLOGY

Zhaoshu Zeng,[1] Ph.D.; Li Wang,[2] M.D.; Qingchuan Feng,[3] Ph.D.; Lirong Zhang,[1] M.D., Ph.D.; Li Lee,[4] M.D.; Lei Wang,[5] M.D.; Yangli Yue,[6] M.D.; Yu Fang,[7] M.D., Ph.D.; Weihong Yang,[1] M.D.; Hanmei Qiu,[1] M.D.; and Ziming Dong,[1] M.D., Ph.D.

Evaluation of 96 SNPs in 14 Populations for Worldwide Individual Identification*

ABSTRACT: Great advances have been made recently in searching for individual identification single-nucleotide polymorphisms (IISNPs or IDSNPs). Such SNPs as suggested by SNPforID scientists and by Pakstis et al., are promising, although they were selected from older or smaller databases rather than the most recent database. Here, we describe a new computational strategy for developing IDSNPs based on HapMap. We searched through HapMap r27 for SNPs having minor allele frequencies ≥0.30 in all its 11 populations and found more than 1881 qualified SNPs. We examined 96 of them with 183 DNA samples from three Chinese populations using Illumina arrays. The average allele frequency for these 96 SNPs among the three populations was 0.495/0.505, the average number of identical SNP genotypes shared by two individuals among the 14 populations (three Chinese and 11 HapMap) was 37.9, and the random matching probability for two unrelated Hans to match in all 96 genotypes was 9.793×10^{-39}. Thus, most of these 96 SNPs are universally applicable.

KEYWORDS: forensic sciences, individual identification, single-nucleotide polymorphism, HapMap, minor allele frequency, Illumina GoldenGate assay, random matching probability

In forensic science, polymerase chain reaction (PCR)-based short tandem repeat (STR) analyses have been successfully used for human identification and parentage testing for nearly 20 years (1,2). Currently, *c.* 20 STRs, mainly suggested by scientists in America and Europe (3–5), have been applied for identification purposes worldwide because of high rates of polymorphisms among various populations, straightforward interpretations into digital data records, and sophisticated typing technologies. However, with rapid advances in DNA sequencing that allow detection of single-base differences, we are approaching a time when all of the DNA sequence differences among individuals can be identified and utilized. A common view is that single-nucleotide polymorphisms (SNPs) offer a potentially cheaper, faster, and more automatable alternative to STR analyses for many forensic DNA applications, including human identification, phenotypic analyses, ancestry inferences, and lineage studies (6–8). However, there are still limitations with using SNPs, including technical problems with amplification, difficulties in plate reading, and low polymorphism information content (PIC) per locus, among others; additionally, SNPs can only act as an adjunct to STRs for solving special problems in forensic genetics.

Recently, there has been a dramatic increase in effort in searching and genotyping SNPs for human identification. SNPs that, when assayed collectively, give very low probabilities for two individuals having the same multi-locus genotypes are called individual identification SNPs (IISNPs or IDSNPs) (7). These collections of SNPs can be used to discriminate between individuals. In 2001, Gill (9) suggested that relatively small arrays of *c.* 50 SNPs could be comparable to STR multiplexes. To date, many researchers have suggested a variety of IISNP panels, each containing 50–100 SNPs, for human identification (10–12).

There are currently two major panels of SNPs, and both of these have attracted significant attention. One was developed as part of the high-throughput analysis of SNPs for the forensic Identification of Persons (SNPforID) project (10,11), and it includes 52 SNPs. The other panel was developed by Pakstis et al. (12), and it includes 92 SNPs. Both panels were validated in many populations, and they may be developed into commercial kits (10–13). However, close examination of these panels revealed that the source databases used for selecting these SNPs were older or smaller than the most recent databases available from the International HapMap Project (14). For the SNPforID project, their four source databases were all created in 2003, and only a few populations were investigated (15). As a consequence, some of their SNPs have uneven allele frequencies in some major populations. For example, rs1335873 has an allele frequency of 0.947/0.053 among Sub-Saharan Africans (n = 226, http://www.ncbi.nlm.nih.gov/snp, accessed on April 20, 2011). As for the 92 SNP panel, because of the small

[1]Department of Forensic Medicine, School of Basic Medical Sciences, Zhengzhou University, Zhengzhou, Henan, China.
[2]Department of Parasitology, School of Basic Medical Sciences, Zhengzhou University, Zhengzhou, Henan, China.
[3]Department of Cellular Biology and Medical Genetics, School of Basic Medical Sciences, Zhengzhou University, Zhengzhou, Henan, China.
[4]Department of Obstetrics and Gynecology, Central Hospital of Zhengzhou, Zhengzhou, Henan, China.
[5]Department of Forensic Sciences, Police Station of Zhengzhou, Zhengzhou, Henan, China.
[6]Department of Oral Pathology, The Affiliated Hospital of Stomatology, Zhengzhou University, Zhengzhou, Henan, China.
[7]Department of Emergency Medicine, the First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan, China.
*Funded by two grants from the National Natural Science Foundation of China (NNSFC) (Grant No. 30700966 and No. 31071100).

Received 11 Jan. 2011; and revised form 24 April 2011; accepted 26 June 2011.

number of unique SNPs per genome provided in their source databases, their IISNP pool was not large enough to avoid linkage disequilibrium (LD) for six of their SNPs (12,16,17).

Because HapMap release #27 (r27) exceeded many earlier and smaller databases in population inclusions and/or unique genomic SNPs, a genome-wide screen for all potential IISNPs based on this batch of data has become necessary. This is mainly because a number of factors need to be considered for a satisfactory IISNP panel. These factors include avoiding LD with other IISNPs or forensic STRs, avoiding unsatisfied flanking sequences that contain too many GC bases or consecutively repeated single bases (15), and avoiding anchoring at copy number variation (CNV) areas (18,19), among others. Thus, an IISNP candidate pool should contain as many as tens of thousands of potential IISNPs. Given that the transferability of HapMap data to more populations has been well verified (20,21), it is now possible to set up a genome-wide IISNP pool. Here, we report on our approaches for setting up such a genomic IISNP pool using HapMap data. We also report on our further assessment of 96 SNPs in three Chinese populations in comparison with genotyping data from the 11 HapMap populations.

Materials and Methods

Genome-Wide IISNP Screening

HapMap (r27) files including more than 2.3×10^9 genotypes from 1017 individuals from 11 populations were downloaded from the HapMap website (http://hapmap.ncbi.nlm.nih.gov/). The 11 HapMap populations include African Americans (ASW); Utah Caucasians (CEU); Han Chinese from Beijing (CHB); Han Chinese from Denver (CHD); Gujarati Indians from Houston (GIH); Japanese from Tokyo (JPT); Luhya from Webuye, Kenya (LWK); Mexican Americans (MEX); Maasai from Kinyawa, Kenya (MKK); Tuscans from Italy (TSI); and Yorubans from Ibadan, Nigeria (YRI). These population samples were sourced as described in reference paper (14). The original frequency files were in.gz format and were unzipped and read by Microsoft Excel 2007 (Microsoft Corp., Redmond, WA). Rs numbers corresponding to SNPs with reference allele frequencies between 0.30 and 0.70 (i.e., minor allele frequency or MAF ≥ 0.30) were listed in a new.xlsx file containing one column for each population. Rs numbers that were present in all 11 columns were extracted by a short macro written with the computer language—Visual Basic Applications (for macro codes, please see the electronic supplementary materials at http://blog.sina.com.cn/legalmed) (21). We identified 48,631 MAF ≥ 0.30 SNPs. Subsequently, a second round of selection based on genotype frequency data for SNPs in Hardy–Weinberg equilibrium (HWE) ($p > 0.05$) was performed with qualified frequencies set at 0.04–0.65 for homozygotes and 0.40–0.50 for heterozygotes. This second type of selection picked 11,509 SNPs. After both types of selection, 1881 SNPs remained.

Selection of 96 SNPs

The pool of 1881 SNPs was used for our final selection of 96 SNPs. To avoid linkage, we chose SNPs with at least 5 Mb between-marker distance between SNPs and at least 1 Mb distance from other promising markers, including the 20 frequently used forensic STRs (3–5), the 52 SNPforID's SNPs (11), and the 92 SNPs determined by Pakstis et al. (12). Additionally, SNPs were not located in CNV regions, which are defined in the Database of Genomic Variants (http://projects.tcag.ca/variation/) (19), and there was appropriate GC content (between 31% and 55%) in the flanking sequences (±120 bp) (22) and appropriate scores (>0.700) to allow interface with the Illumina GoldenGate Genotyping Assay (Illumina, San Diego, CA). Parts of these selection strategies were published in our previous work (21,23). Table S1 shows detailed information about these 96 SNPs, and Fig. S1 shows chromosome locations and relative distances to the three other marker panels (both available at http://blog.sina.com.cn/legalmed). These 96 SNPs were custom designed and synthesized into two blocks of Illumina GoldenGate arrays (Serial no.: 4831397001, 4911733011) by Illumina.

Sample Collection and DNA Preparation

Approximately 2 mL of peripheral venous blood was taken from 183 individuals from three Chinese ethnic groups after informed consent. Selected individuals included Han Chinese from Henan in central China ($n = 61$), Mongolians from Inner Mongolia ($n = 62$), and Tibetans from a Tibetan Autonomous State in Sichuan ($n = 60$). Donors were informed that the samples would be anonymized and analyzed for human identification purpose by SNP typing. The ethics committee of Zhengzhou University gave institutional approval for the entire study. All genomic DNA was prepared by standard phenol–chloroform procedures (24) and K562 (Promega, Madison, WI), and 9947A (Applied Biosystems, Warrington, UK) standard DNA samples were used as positive controls.

SNP Genotyping

A total of 192 DNA samples (including four duplicate samples and the two positive control samples) were genotyped with two Illumina GoldenGate array blocks. Quality control was performed by examining duplicate samples, use of standard positive control DNA samples, and direct Sanger sequencing of randomly selected samples. A list of the PCR and hybridization sequences is also shown in Table S1 (available at http://blog.sina.com.cn/legalmed).

Statistical Analysis and Genotype Comparison

Allele frequencies were calculated by direct counting. LD and departures from HWE were analyzed with SNPSTATS online at http://bioinfo.iconcologia.net/snpstats/start.htm (25). The matching probability (MP), power of discrimination (PD), PIC, and other forensic parameters for each SNP were calculated as described by Tereba using PowerStats (Promega) (26). Chi-square tests were used to compare genotype frequency differences between any two of the five examined populations (three Chinese populations and two HapMap populations—CHB and CHD) for all 96 loci. All genotypes, including genotypes from our experiments and genotypes downloaded from HapMap (r27), were integrated into one table to determine whether any individuals were identical for all 96 loci, with samples excluded when even one "NN" was present in the genotype results.

Results

Universal IISNP Selection

In the first round of selection that focused on SNPs with MAF ≥ 0.300 among each of the 11 HapMap populations, 48,631 MAF ≥ 0.300 SNPs were found to be shared by the 11 populations. Of these, 1881 SNPs were further identified to have

probabilities of HWE ≥ 0.05 in all of the 11 populations after a second round of selection. A full list of these 1881 universally applicable candidate IISNPs is publicly available at http://legalmed.blog.sohu.com/ or http://blog.sina.com.cn/legalmed. For example, rs6108756, which has MAF ≥ 0.30 and HWE ≥ 0.05, had C/T allele frequencies that ranged between 0.309 and 0.691 among the 11 populations. The average allele frequency was 0.532/0.468, and the lowest HWE probability was 0.17 (more data is shown in Table S2 at http://blog.sina.com.cn/legalmed or http://blog.sina.com.cn/s/blog_5ec450150100p33q.html).

Genotyping Results for the 96 SNPs

In general, allele frequencies of the tested 96 SNPs ranged from 0.500/0.500 to 0.190/0.810, with an average allele frequency of 0.495/0.505, an average heterozygosity (Het) of 0.473, and an average Fst (fixation index between subpopulations and total population) of 0.006269 among the three Chinese populations. Detailed allele frequencies observed for the 96 SNPs and their Het and Fst values in three Chinese populations are summarized in Table S3 and graphically displayed in Fig. S2 (both available at http://blog.sina.com.cn/legalmed). Detailed genotypes of K562 and 9947A are summarized in Table S4 (available at http://blog.sina.com.cn/legalmed). Although most of these SNPs had HWE ratios larger than 0.05 ($p \geq 0.050$), departures from the HWE ratios were seen in some SNPs, including the following SNPs: rs1015665 in all the three populations ($p < 0.019$); rs11139320 and rs122452 in Han Chinese ($p = 0.039$); rs224045, rs7046039, rs7448547, and rs10756626 in Mongolians ($0.014 \leq p \leq 0.049$); and rs11151684, rs10405035, and rs6446700 in Tibetans ($0.002 \leq p \leq 0.025$). LD tests of these SNPs showed that the average *D*' value of all randomly paired markers was 0.0802, and the largest value (0.3873) was obtained when rs1476909 and rs947098 were tested as a pair. The average r^2 value of all of paired markers was 0.0059, and the largest value (0.0810) was obtained between rs4131354 and rs947098.

Chi-square tests on both allele frequencies and genotype frequencies between our three Chinese populations and two HapMap populations (CHD and CHB) showed that $p \leq 0.05$ was obtained for six SNPs: rs1015665, rs12467794, rs12959775, rs2000546, rs221899, and rs27419.

Genotype Matching and Random Matching Probabilities

The average number of identical SNP genotypes found between two unrelated individuals among the three populations was 38.0, with a minimum of eight and a maximum of 60 matching loci (183 samples, 33,306 comparisons). Additionally, the average number of identical SNP genotypes between two unrelated individuals in the 14 total examined populations (three Chinese and 11 HapMap) was 37.9, with a minimum of four and a maximum of 64 matching loci (980 samples, 959,420 comparisons, with 797 samples downloaded from HapMap).

Frequencies for each SNP were calculated and applied to obtain each MP, which ranged from 0.335 to 0.536 among the three Chinese populations. Detailed MPs, PDs, PICs, and other forensic parameters for each SNP are summarized in Table S5 (available at http://blog.sina.com.cn/legalmed). When all 96 SNPs were used in combination, from the sample sets with different numbers analyzed in our arrays, the random matching probabilities of complete matching at all 96 loci between any two unrelated individuals was the product of 96 MPs, which was 9.793×10^{-39} for the Han population ($n = 61$ for each MP), 1.338×10^{-39} for the Mongolian population ($n = 62$), and 5.018×10^{-39} for the Tibetan population ($n = 60$). An overview of the differentiation between the 14 populations is shown in Fig. S3.

Discussion

Ideal SNPs for individual identification should have two primary characteristics: high Het values (*c.* 50% for a biallelic system) and low Fst values (<0.06 in multiple populations) among world populations (6,7,12,16,17). When meeting these criteria, as few as 50–100 autosomal loci are sufficient for obtaining the high levels of discrimination that are currently allowed by analyses of the 13–20 forensic STRs (3–5,9), and fewer reference population data sets will be required for IISNPs' statistical assessments for forensic casework. These two criteria can be simultaneously achieved by finding SNPs with allele frequency distributions of *c.* 0.50/0.50 across most or all of the world's populations (21). Fortunately, the currently suggested IISNP panels meet the above criteria and have good allele frequencies and similar frequency patterns among multiple populations, especially the 92 SNPs recommended by Pakstis et al., (12) which have Het values > 0.4 and Fst values < 0.06 in 44 studied populations. However, some drawbacks exist in these IISNP sets that will greatly hamper their universal application. For example, SNPforID (11) recommends using rs1335873, which has uneven allele distributions in Africans (0.947/0.053, $n = 226$) and may lead to low discriminative power in African populations. Additionally, the LD of six SNPs in the panel recommended by Pakstis et al. (12) may result in insufficient power for effective analyses. The proximity of some of these SNPs to some forensic STRs may cause linkage to those STRs, which would prevent joint applications. Additionally, anchoring in CNV areas may make technical analyses challenging. All these problems arise from a number of factors, but a major factor may be the small size of the original data sets used for selecting the SNPs.

With recent advances in the HapMap Project, SNP genotypes from across the genomes of *c.* 1000 individuals from 11 populations were generated (14). Additionally, more SNPs will be identified by the 1000 Genomes Project, which will investigate *c.* 2000 unidentified people from *c.* 20 populations around the world with the help of next-generation platforms and technologies (27). With these high-throughput SNP sequencing advances that have assayed many different populations, more SNPs that can be used for individual identification will be determined. Therefore, a thorough investigation of these recent data sets should be performed as soon as they are available.

In our previous work, we succeeded in finding SNPs with high Het and low Fst values by setting MAF ≥ 0.45 based on HapMap data r21a (21). Under these restrictions, all target SNPs have balanced allele frequencies from 0.450/0.550 to 0.500/0.500 (or 0.500/0.500–0.550/0.450) in each of the four populations (CEU, CHB, JPT, and YRI), with very low frequency variations between these populations and versatile genotypes within populations. However, HapMap r21a only had four populations, and we needed to update our recommendations when more populations became available in HapMap r27. Now, with this recent and abundant source of data, we updated our universal IISNP selections to include SNPs with MAF 0.30 and HWE ($p \geq 0.05$) across the 11 populations in this study, and we found 1881 generally applicable IISNPs. This candidate pool offers enough flexibility that it avoids the drawbacks described previously, including self-linkage among SNPs, linkage to any existing forensic STRs or IISNP panels, and linkage to CNV anchoring regions. Most importantly, this large pool provides sufficient opportunities to balance each SNP with an even allele

frequency distribution of *c.* 0.500/0.500 and a similar frequency pattern across various populations.

Our experimental testing of these unlinked 96 SNPs demonstrated that most of them had good allele frequencies of *c.* 0.500/0.500 and had similar frequency patterns among the three Chinese populations. For example, the C/T allele frequencies of rs1996872 were 0.510/0.490 in Han Chinese, 0.500/0.500 in Mongolians, and 0.540/0.460 in Tibetans. Our calculations clearly demonstrated that, when tested in combination, these 96 SNPs provide high discriminatory power and low MPs in a range of values that are comparable to the 13 or more currently used forensic STR markers (3–5). Given that HapMap data are reliable (20,21) and cover numerous divergent populations, our test results on the 96 unlinked SNPs are expected to be applicable to other large human populations because the essential characteristics should be the same. Our panel of SNPs also has additional benefits, including absolute independence from any coding SNPs or other existing forensic STRs/IISNPs systems, good flanking sequences, and no CNV area anchoring.

One minor problem is that allele frequencies of a few of our 96 SNPs fluctuated greatly when populations beyond Asia were examined. An MAF $\geq$ 0.30 criterion means that allele frequencies can only vary from 0.30 to 0.70. Although most of the 96 SNPs in our panel had frequency patterns within this zone in different populations, some of them (such as rs7046039) did not demonstrate this allele frequency and varied between 0.190 in Han Chinese to 0.637 in Yorubans. Such allele frequency fluctuations may be because of inaccuracies resulting from small sample sizes in different tests. However, these fluctuations suggested that we should limit MAF values to 0.35 or higher to ensure that the MAFs will not range too far from 0.500. Additionally, we should put more emphasis on allele frequencies than on certain other factors (such as HWE $\geq$ 0.05) and use more specialized criteria when available (such as Fst values for IISNP screen). These steps will allow us to refine our 96 SNPs in our future study and increase the scientific community's abilities to resolve mixed DNA analyses with the minimum number of independent SNPs (28).

Acknowledgments

We would like to thank Drs. Zhengxi Li, Guowei Che, Weihua Zhang, and Enwu Yuan for their help in obtaining some of the blood samples and two senior Office fans ''hshj'' and ''YGZ'' at http://www.officefans.net for their help in macro-editing and data comparison. The majority of the experiments were conducted in the Henan Key Laboratory of Molecular Medicine (Zhengzhou, Henan, China) and the United Gene Holdings in Shanghai. Additionally, the latter helped with the Illumina genotyping.

References

1. Brinkmann B, Rand S, Bajanowski T. Forensic identification of urine samples. Int J Legal Med 1992;105(1):59–61.
2. Urquhart A, Kimpton CP, Downes TJ, Gill P. Variation in short tandem repeat sequences—a survey of twelve microsatellite loci for use as forensic identification markers. Int J Legal Med 1994;107(1):13–20.
3. Budowle B, Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM. Population data on the thirteen CODIS core short tandem repeat loci in African Americans, U.S. Caucasians, Hispanics, Bahamians, Jamaicans, and Trinidadians. J Forensic Sci 1999;44(6):1277–86.
4. Zeng ZS, Zheng XD, Zhu YL, Wang ZQ, Xiang ZD, Meng XS, et al. Population genetic data of 15 STR loci in Han population of Henan province (central China). Leg Med (Tokyo) 2007;9(1):30–2.
5. Amigo J, Phillips C, Salas T, Formoso LF, Carracedo Á, Lareu M. pop.STR—an online population frequency browser for established and new forensic STRs. Forensic Sci Int Gene Suppl 2009;2(1):361–2.
6. Budowle B, van Daal A. Forensically relevant SNP classes. BioTechniques 2008;44(5):603–8.
7. Butler JM, Budowle B, Gill P, Kidd KK, Phillips C, Schneider PM, et al. Report on ISFG SNP panel discussion. Forensic Sci Int Gene Suppl 2008;1(1):471–2.
8. Kidd JR, Friedlaender FR, Speed WC, Pakstis AJ, De La Vega FM, Kidd KK. Analyses of a set of 128 ancestry informative single-nucleotide polymorphisms in a global set of 119 population samples. Investig Genet 2011;2(1):1.
9. 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.
10. Phillips C, Prieto L, Fondevila M, Salas A, Gomez-Tato A, Alvarez-Dios J, et al. Ancestry analysis in the 11-M Madrid bomb attack investigation. PLoS ONE 2009;4(8):e6583.
11. Tomas C, Axler-Diperte G, Budimlija ZM, Borsting C, Coble MD, Decker AE, et al. Autosomal SNP typing of forensic samples with the GenPlex HID System: results of a collaborative study. Forensic Sci Int Genet 2011;5(5):369–75.
12. 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.
13. Fang R, Pakstis AJ, Hyland F, Wang D, Shewale J, Kidd JR, et al. Multiplexed SNP detection panels for human identification. Forensic Sci Int Gene Suppl 2009;2(1):538–9.
14. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature 2007;449(7164):851–61.
15. Phillips C, Lareu M, Sanchez J, Brion M, Sobrino B, Morling N, et al. Selecting single nucleotide polymorphisms for forensic applications. Int Congress Series 2004;1261:18–20.
16. Pakstis AJ, Speed WC, Kidd JR, Kidd KK. Candidate SNPs for a universal individual identification panel. Hum Genet 2007;121(3–4):305–17.
17. 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.
18. Pang AW, MacDonald JR, Pinto D, Wei J, Rafiq MA, Conrad DF, et al. Towards a comprehensive structural variation map of an individual human genome. Genome Biol 2010;11(5):R52.
19. Zhang J, Feuk L, Duggan GE, Khaja R, Scherer SW. Development of bioinformatics resources for display and analysis of copy number and other structural variants in the human genome. Cytogenet Genome Res 2006;115(3–4):205–14.
20. Hu C, Jia W, Zhang W, Wang C, Zhang R, Wang J, et al. An evaluation of the performance of HapMap SNP data in a Shanghai Chinese population: analyses of allele frequency, linkage disequilibrium pattern and tagging SNPs transferability on chromosome 1q21-q25. BMC Genet 2008;9:19.
21. Zeng Z, Yan H, Wang L, Yuan E, Yang W, Liao Z, et al. Genome-wide screen for individual identification SNPs (IISNPs) and the confirmation of six of them in Han Chinese with pyrosequencing technology. J Forensic Sci 2010;55(4):901–7.
22. Zeng Z, Yan H, Zheng X, Hu G, Chen Y, Ding M. High GC amplification: a comparative study of betaine, DMSO, formamide and glycerol as additives. Life Sci J—Acta Zhengzhou Univ Oversea Version 2006;3(1):67–71.
23. Zeng Z, Wang L, Fang Y, Zhang S, Zhang G, Guo K. The screening and genotyping of highly informative and universally applicable SNPs for individual identification. J Zhengzhou Univ (Medical Sci) 2010;45(3):378–81.
24. Maniatis T, Fritsch EF, Sambrook J. Molecular cloning. A laboratory manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory, 1982.
25. Sole X, Guino E, Valls J, Iniesta R, Moreno V. SNPStats: a web tool for the analysis of association studies. Bioinformatics 2006;22(15):1928–9.
26. Tereba A. Tools for analysis of population statistics. Profiles in DNA 1999;2(3):14–6.

27. Via M, Gignoux C, Burchard EG. The 1000 genomes project: new opportunities for research and social challenges. Genome Med 2010;2(1):3.
28. Visscher PM, Hill WG. The limits of individual identification from sample allele frequencies: theory and statistical analysis. PLoS Genet 2009;5(10):e1000628.

Additional information and reprint requests:
Zhaoshu Zeng, Ph.D.
Department of Forensic Medicine
School of Basic Medical Sciences
Zhengzhou University
100 Science Road
Gaoxin District
Zhengzhou
Henan 450001
China
E-mail: zs_zeng@yahoo.com

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. The chromosome locations and the relative positions of our 96 SNPs to the other three marker panels.

Figure S2. The frequency variations of allele N1, and Het and Fst values of the 96 SNPs in the three Chinese populations.

Figure S3. Dendrogram obtained by pairwise analyses of the genetic distances between the 14 populations using Ward's method.

Table S1. rs#, chromosomal localization (NCBI build 36, dbSNP b126), 5′-3′ sequences of PCR primers and probes and other information for the 96 SNPs' analysis with the custom-designed Illumina Sentrix Array Matrix.

Table S2. The allele frequencies of rs6108756 in the HapMap (r27) database and the statistic results on several forensic parameters.

Table S3. The allele frequencies and Fst values of 96 SNPs among three Chinese ethnic groups (for Han, $n = 61$; for Mongols, $n = 62$; and for Tibetan, $n = 60$).

Table S4. K562 and 9947A genotyping results on the 96 SNPs.

Table S5. Forensic statistics about 96 SNPs among the three Chinese populations (for Han, $n = 61$; for Mongols, $n = 62$; and for Tibetan, $n = 60$).

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.