1 **Title**

2 Approaches for identifying multiple-SNP haplotype blocks for use in human
3 identification

4
5
6
7
8
9
10
11

12 Corresponding author: Nakahara Hiroaki
13 6-3-1 Kashiwanoha, Kashiwa, Chiba 277-0882, Japan
14 Tel.: 81-4-7135-8001
15 Fax: 81-4-7133-9159
16 E-mail: senju@nrips.go.jp
17

1 Original Article

2

3 # Approaches for identifying multiple-SNP haplotype blocks for use in human identification

4

5

6 Nakahara Hiroaki[ab*], Fujii Koji[a], Kitayama Tetsushi[a], Sekiguchi Kazumasa[a],

7 Nakanishi Hiroaki[b], Saito Kazuyuki[b]

8

9 [a] *National Research Institute of Police Science, Chiba 277-0882, Japan*

10 [b] *Juntendo University School of Medicine, Tokyo 113-8421, Japan*

11

Abstract

Single nucleotide polymorphism (SNP) discrimination effectiveness is low due to the bi-allelic nature of SNPs, and large numbers of loci must be analyzed for human identification in forensic casework. To resolve these issues, the authors support the use of multiple SNP haplotypes that will generate many haplotypes based on the combination of SNP alleles. First, 27 regions were selected from the JSNP database (http://snp.ims.u-tokyo.ac.jp) according to the following criteria: (1) 3 or more SNP loci within 100 bp; (2) on-intron or out-of-gene location; and (3) frequency of more than 40% for each SNP allele. PCR amplification and high-resolution melting curve (HRM) analysis were then carried out for all selected regions to determine variation in the haplotypes of each. HRM analysis indicated that 7 regions (1q25, 1q42.2, 3p24, 10p13, 11p15.1, 14q12–q13, and 20q12) containing 3 SNP loci had more than 2 haplotypes. The frequencies of the haplotypes for each region were observed via direct sequencing of more than 100 individuals. Not only haplotyping increases the effectiveness of individual identification but also the analysis region is shorter than in common short tandem repeat analysis, representing a further advantage for fragmented DNA samples in SNP typing.

1 *Keywords:* forensic; DNA typing; SNP; haplotype; individual identification

2

1. Introduction

The main technique currently used for human DNA identification in forensic casework is short tandem repeat (STR) typing. However, several methods involving the use of single nucleotide polymorphism (SNP) typing for forensic application have been reported based on their related advantages [1-5]. For example, millions of SNP loci are found throughout the human genome [1, 6], and SNP typing can be used with fragmented DNA from degraded forensic human tissues thanks to its shorter detection region in comparison with those of commercially available multiplex STR typing systems that produce amplicons ranging in size from 100 to 450 base pairs (bp). However, forensic sample SNP typing techniques for individual identification have specific requirements, such as the need to analyze large numbers of loci for more effective discrimination, due to the bi-allelic nature of SNPs [7-9].

To resolve this issue, the authors support the use of multiple SNP haplotypes. Some researchers have previously investigated and validated the use of polymorphic haplotypes (also known as mini haplotypes, haplotype blocks, or microhaplotypes) for forensic, genetic, and ancestral study [1, 10-12]. Typing numerous haplotypes generated by a

combination of SNP alleles is expected to increase the effectiveness of individual identification. Additionally, if all SNP loci are sited in short regions, the analysis region will also be short and the advantage of SNP typing for fragmented DNA samples will be maintained. However, SNPs in short regions are likely to be linked to each other and end up making only 2 haplotypes in the same way as a single SNP loci. Accordingly, the authors examined haplotype regions with a length of less than 100 bp made from a combination of multiple SNPs to determine the possible presence of 3 or more haplotypes. High-resolution melting curve (HRM) analysis for amplified sequence regions was used to determine haplotype varieties.

2. Materials and methods

2.1. DNA samples

Blood samples were collected from unrelated Japanese individuals who had given written consent, and DNA was isolated from the samples using a MagNA Pure LC instrument (Roche Diagnostics, Germany) for quantification with a Human DNA Quantification Kit (Takara Bio, Japan) and a Real-Time PCR Smart Cycler II System

(Cepheid, USA) in line with the manufacturers' instructions.

2.2. Selection of haplotype loci

Haplotype regions in the human genome were searched for in the online JSNP database (http://snp.ims.u-tokyo.ac.jp/index.html) with the following criteria:

1. Three or more SNP loci sited in a domain of less than 100 bp;
2. All SNPs sited within an intron or other non-coding region;
3. Allele frequency of more than 0.4 for each SNP.

The search produced 27 candidate regions for the analysis outlined below.

2.3. PCR amplification and HRM analysis

Primer pair candidates for the amplification of the 27 loci were designed using the online Primer3Plus resource (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) (Table 1).

PCR amplification and HRM analysis were carried out using a Rotor-Gene 6000 system (Corbett Life Science, Australia). The PCR mixture contained 0.2 μM of each

1 primer for the region, 12.5 μL SYBR® Premix Ex Taq™ (Takara Bio), 0.5 μL
2 LightCycler® 480 ResoLight Dye (Roche Diagnostics), and 0.5–1 ng template DNA in a
3 total volume of 25 μL. ResoLight Dye (an intercalation type) was added to the reaction
4 mixture to enhance fluorescence. The PCR conditions involved heating at 95°C for 1 min
5 and 40 cycles at 95°C for 10 s and 60°C for 30 s. Following PCR, HRM analysis was
6 carried out with the reaction mixture heated from 60°C to 90°C at a rate of 0.1°C every 2
7 s and continuous measurement of fluorescence at 510 nm. The melting curve data were
8 plotted using Rotor-Gene 6000 Series Software v1.7 (Corbett Life Science). Curve
9 pattern variations were observed as main peaks indicating different melting temperatures,
10 bimodal peaks indicating the coexistence of different sequences, or small subsidiary
11 peaks indicating the presence of heteroduplexes (Fig. 1). Regions judged visually to have
12 more than 3 melting curve patterns and more than 2 haplotypes were selected for
13 subsequent sequencing analysis.

Fig. 1

14

15 *2.4. Direct sequencing*

16

17 Direct sequencing analysis of the amplified haplotype regions was carried out using an
18 ABI PRISM BigDye® Terminator v1.1 Cycle Sequencing Kit (Life Technologies, USA)

and an Applied Biosystems 3130xl Genetic Analyzer (Life Technologies), and the results were analyzed using Sequencing Analysis Software v5.4 (Life Technologies) in line with the manufacturer's instructions. The primers used for the sequencing reaction were the same as those used for amplification. Analysis, sequence alignment, and SNP typing were carried out using Sequencher^{TM} v5.0 (Gene Codes Co., USA).

2.5. Data analysis

The haplotypes of all loci were defined from the sequencing data in which all SNPs were homozygous. In addition, PHASE version 2.1 [13, 14] was used to estimate the unknown haplotypes in each region. The frequencies of haplotypes were also estimated using PHASE version 2.1. The allele frequencies and power of discrimination (PD) of each haplotype region were calculated from the obtained haplotype frequencies.

3. Results

The selection of regions from the JSNP database produced 27 candidates containing 3 or 4 SNPs (Table 1).

Table 1

PCR and HRM analysis produced melting curve patterns for 7 of the 27 candidate

1 regions (regions 3, 4, 7, 14, 17, 21, and 26) that were distributed visually into more than 4
2 groups (Fig. 1) and expected to have 3 or more haplotypes, and these were labeled #a to
3 #g, respectively. Each region was dispersed across different chromosomes, except for #a
4 and #b on the long arm of chromosome 1. The inter-region distance was approximately 46
5 Mbp (JSNP database). In order to examine their genotypes and determine haplotype
6 frequencies, direct sequencing of each region was carried out using more than 100
7 individuals. The observed genotypes are shown in Table 2. Three regions, #a (at 1q25
8 containing rs338564, rs338565, and rs543912), #c (at 3p24 containing rs2293140,
9 rs2293141, and rs2293142), and #g (at 20q12 containing rs2070640, rs2070639, and
10 rs2070638), were predicted to have 3 or more haplotypes because 6 genotypes were
11 observed from these regions. Actually, 3 haplotypes were defined in each region because
12 all 3 SNPs in each region were observed as homozygous in the sequencing data from a
13 number of samples. However, 4 regions, #b (at 1q42.2 containing rs2296796, rs2296797,
14 and rs2296798), #d (at 10p13 containing rs2277214, rs2277215, and rs2277216), #e (at
15 11p15.1 containing rs2074312, rs2074311, and rs2074310), and #f (at 14q12–q13
16 containing rs2281629, rs2281630, and rs2281631), appeared to have 4 or more
17 haplotypes, based on the observation of more than 6 genotypes. PHASE predicted the
18 presence of 4 or more haplotypes for all regions, although some of their calculated

1 frequencies were less than 0.1%. One predicted haplotype of region #a (-G-G-A-), 3 of
2 region #b (-G-T-A-, -A-G-G-, and -A-T-A-), 1 of region # c (-C-A-G-), 1 of region # d
3 (-G-G-C-), 3 of region #e (-C-A-A-, -C-A-G- and -T-A-A-), 1 of region #f (-C-C-A-), and
4 1 of region #g (-C-C-G-) were not defined as homozygous. In consideration of the human
5 identification capacity of individual regions, PD values were calculated from the
6 observed haplotype frequencies (Table 2).
7
8 **4. Discussion**
9
10 In this study, the online JSNP database [7, 15] was used to search for and select regions
11 containing multiple SNPs. Information on the database includes allele frequencies,
12 chromosome mapping, nucleotide positions of SNP loci, and the surrounding sequences.
13 This nucleotide position information also helps to clarify the mutual proximity of SNP
14 loci and the distance between them. Here, a search was performed to find regions
15 containing 3 or more SNP loci within 100 bp. Longer regions are more likely to contain
16 more SNP loci and can make more polymorphisms, but have less advantage in relation to
17 degraded forensic casework DNA samples. From the database search results, 27 regions
18 with 3 or 4 SNP loci within 100 bp or less were selected (Table 1). More regions with high

1 numbers of polymorphisms may have been found with less stringent selection criteria,

2 including those regarding allele frequency or the number of SNP loci.

3 HRM analysis was used to estimate tentatively the number of haplotypes in each

4 region. Melting curve analysis can be applied to determine DNA fragment sequence

5 differences, which are observed as differences in melting temperature [16, 17]. With

6 HRM analysis in particular, melting temperatures are measured to a tenth of a degree, and

7 even single base differences in DNA fragment sequences can be distinguished [16, 17]. It

8 can also be used to detect the presence of heteroduplexes from heterozygous loci, which

9 are observed by their characteristic melting curve pattern. In this study, PCR

10 amplification and subsequent melting were carried out continuously in a reaction mixture

11 containing SYBR Green I intercalation dye and ResoLight Dye (Roche Diagnostics) to

12 increase detection sensitivity. Real-time HRM analysis using an intercalator dye is a

13 simple and inexpensive way of detecting DNA polymorphisms without the need for

14 additional reagents such as labeled hybridization probes or steps such as electrophoresis.

15 Typical melting curve patterns for each fragment sequence are plotted on negative first

16 derivative graphs using analysis software. HRM is a convenient way of searching for

17 polymorphisms in haplotypes because the number of curve patterns is theoretically

18 equivalent to the number of sequence patterns. Regions that appeared to have only 2

1 haplotypes from which fewer than 4 melting curve patterns could be observed were
2 rejected. Although such regions may have had more than 2 haplotypes because different
3 sequences may result in similar curve patterns by chance, no additional analysis was
4 performed to determine this here.

5 In sequencing electrophoresis, heterozygotes from different haplotypes were observed
6 as mixtures of different nucleotides at SNP positions. However, it is difficult to determine
7 existing haplotype sequences in this way because some conceivable combination patterns
8 of 3 SNP alleles result in the same sequencing electrophoresis pattern. Nonetheless, many
9 haplotypes were identified clearly based on homozygote results obtained from the same
10 haplotypes. Ambiguous haplotypes and their frequencies were estimated using PHASE
11 for each region (Table 2). PHASE distributes the unobserved haplotypes and haplotype
12 frequencies given the observed genotype data in Bayesian approaches[14]. The common
13 SNP typing method for individual loci and traditional sequencing analysis do not support
14 the determination of haplotype sequences for individual DNA molecules as mentioned
15 above. However, approaches currently under development such as next-generation
16 sequencing, which can be used to analyze individual sequences of DNA molecules, can
17 be adopted for this purpose.

18 Theoretically, a locus with 3 haplotypes will have 6 genotype patterns (3 homozygotes

1 and 3 heterozygotes), and a locus with 4 haplotypes will have 10 genotypes, while a locus
2 with 2 haplotypes will only have 3 genotypes. Estimated genotypes that can be made
3 from estimated haplotypes and their frequencies are not shown in the table because of
4 space limitations. All of the conceivable genotypes calculated to have a frequency of
5 0.1% or more, except for 3 (-G-H-C-: estimated frequency of 0.4% in region #d; -C-C-A-:
6 estimated frequency of 0.4% in region #f; and -C-H-H-: estimated frequency of 0.7% in
7 region #f), were observed in our experiment.

8 PD values for each region were calculated from the frequencies of genotypes.
9 Theoretically, the PD value of a SNP locus that has 2 alleles will be 0.625 or less.
10 Meanwhile, all currently available common STR loci have many alleles and high PD
11 values of more than 0.8 or 0.9 [18]. The PD values of all haplotype loci validated in this
12 study were higher than those of SNP loci, but lower than those of STR loci. However, one
13 of them (region #b) was higher than the PD value of TPOX observed in a previous study
14 [18]. The combined random match probability of 7 regions is 2.3×10^{-5}. (The one of the
15 15 STR loci is 1.8×10^{-17}[18].)

16 It was initially thought that haplotypes within small regions of the genome may have a
17 low polymorphism rate because neighboring SNPs are closely linked and rarely seem to
18 be separated by the crossing-over of genes. However, in this study, 7 of the 27 regions

1 within 100 bp (approximately 26% of all candidates) had 3 or more haplotypes. It is

2 highly likely that there are more potential regions in the human genome that have a high

3 polymorphism rate and can be used for forensic identification.

4

Figure labels

Fig. 1. (A) HRM analysis for a region containing 3 SNP loci (rs2295561, rs2295562, and rs2295563). Only 3 melting curve patterns resulted from 2 homozygotes and 1 heterozygote with 2 haplotypes. (B) HRM analysis for a region containing 3 SNP loci (rs2070640, rs2070639, and rs2070638). More than 3 melting curve patterns were observed, and the region was identified as having 3 or more haplotypes.

1

2 **References**

3

4 [1] Giardina E, Pietrangeli I, Martone C, Asili P, Predazzi I, Marsala P, et al. In silico and in
5 vitro comparative analysis to select, validate and test SNPs for human identification. BMC
6 Genomics. 2007;8:457.
7 [2] Budowle B. SNP typing strategies. Forensic Sci Int. 2004;146 Suppl:S139-42.
8 [3] Hiratsuka M, Tsukamoto N, Konno Y, Nata M, Hashiyada M, Funayama M, et al.
9 Forensic assessment of 16 single nucleotide polymorphisms analyzed by hybridization probe
10 assay. Tohoku J Exp Med. 2005;207:255-61.
11 [4] Nakamura Y. [SNP Idenshi Takei no Senryaku (SNP genetic polymorphism strategy)]:
12 Nakayama Shoten; 2000.
13 [5] Romanini C, Catelli ML, Borosky A, Pereira R, Romero M, Salado Puerto M, et al. Typing
14 short amplicon binary polymorphisms: supplementary SNP and Indel genetic information in
15 the analysis of highly degraded skeletal remains. Forensic science international Genetics.
16 2012;6:469-76.
17 [6] Pakstis AJ, Speed WC, Kidd JR, Kidd KK. Candidate SNPs for a universal individual
18 identification panel. Hum Genet. 2007;121:305-17.
19 [7] Nakahara H, Hosono N, Kitayama T, Sekiguchi K, Kubo M, Takahashi A, et al.
20 Automated SNPs typing system based on the Invader assay. Leg Med (Tokyo). 2009;11 Suppl
21 1:S111-4.
22 [8] Amorim A, Pereira L. Pros and cons in the use of SNPs in forensic kinship investigation: a
23 comparative analysis with STRs. Forensic Sci Int. 2005;150:17-21.
24 [9] Chakraborty R, Stivers DN, Su B, Zhong Y, Budowle B. The utility of short tandem repeat
25 loci beyond human identification: implications for development of new DNA typing systems.
26 Electrophoresis. 1999;20:1682-96.
27 [10] Pakstis AJ, Fang R, Furtado MR, Kidd JR, Kidd KK. Mini-haplotypes as lineage
28 informative SNPs and ancestry inference SNPs. Eur J Hum Genet. 2012;20:1148-54.
29 [11] Ge J, Budowle B, Planz JV, Chakraborty R. Haplotype block: a new type of forensic DNA
30 markers. Int J Legal Med. 2010;124:353-61.
31 [12] Kidd KK, Pakstis AJ, Speed WC, Lagace R, Chang J, Wootton S, et al. Current
32 sequencing technology makes microhaplotypes a powerful new type of genetic marker for
33 forensics. Forensic science international Genetics. 2014;12:215-24.

[13] Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. American journal of human genetics. 2001;68:978-89.
[14] Stephens M, Donnelly P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. American journal of human genetics. 2003;73:1162-9.
[15] Nakahara H, Sekiguchi K, Hosono N, Kubo M, Takahashi A, Nakamura Y, et al. Criterion values for multiplex SNP genotyping by the invader assay. Forensic science international Genetics. 2010;4:130-6.
[16] Zhou L, Wang L, Palais R, Pryor R, Wittwer CT. High-resolution DNA melting analysis for simultaneous mutation scanning and genotyping in solution. Clin Chem. 2005;51:1770-7.
[17] Reed GH, Kent JO, Wittwer CT. High-resolution DNA melting analysis for simple and efficient molecular diagnostics. Pharmacogenomics. 2007;8:597-608.
[18] Yoshida K, Takahashi K, Kasai K. Allele frequencies of 15 loci using AmpFlSTR Identifiler Kit in Japanese population. J Forensic Sci. 2005;50:718-9.

Figure 1

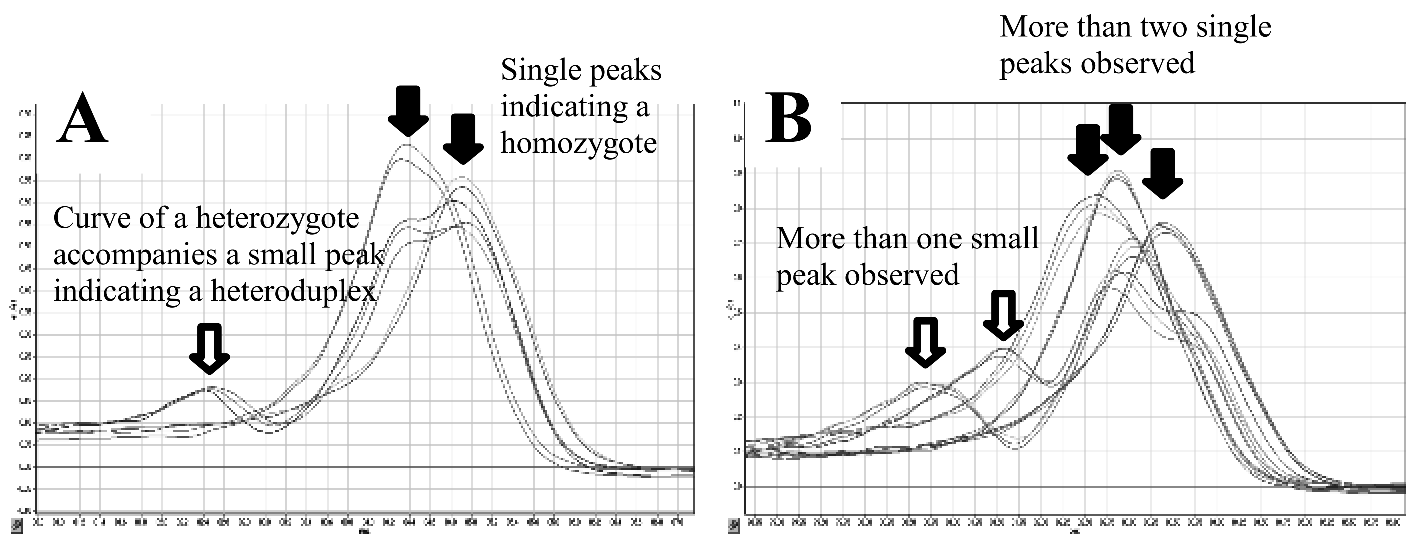

Table1 Candidate regions selected from SNPs database.

region number	chromosome	rs ID	JSNP ID	gene	range*	primers for PCR amplification and HRM analysis
1	1p36.13	rs910230 rs2311483 rs3818042	IMS-JST178572 IMS-JST178571 IMS-JST178570	PADI2	97	GGCCTCATCTGGGAAGAGA ACAGACTTGCCCTCCAGAAA
2	1p36.13	rs2235919 rs2076612 rs2076611	IMS-JST016063 IMS-JST004861 IMS-JST004860	PADI2	89	GGAGTGGGGAATGCAAGATA GGAGACAGACCCTCAGGACA
3	1q25.3	rs338564 rs338565 rs543912	IMS-JST178853 IMS-JST030330 IMS-JST030331	RGL	99	AGTGGAGGGACAGAACGTGT TGGTGGTTCAGAGGACACAG
4	1q42.2	rs2296796 rs2296797 rs2296798	IMS-JST051105 IMS-JST051106 IMS-JST051107	LDLC	53	CCCATGTTTGTCGATTTGGT AGTGGCCACAATCCAAGTTC
5	2p23.1	rs2303324 rs2303323 rs2303322 rs2303321	IMS-JST059945 IMS-JST059944 IMS-JST059943 IMS-JST059942	GALNT14	98	CATCCCTATGCCAGATGCTT GGGTTGTGTTCATGCAGCTA
6	2q35-q37	rs3769644 rs3769643 rs3769642	IMS-JST124464 IMS-JST124463 IMS-JST124462	COL4A4	84	TGCAAGAACAAAATGAACAGG CGACCAGTTTGGTTGAATTTG
7	3p24	rs2293140 rs2293141 rs2293142	IMS-JST046140 IMS-JST046141 IMS-JST046142	KAT2B	73	TAGCCACCAGCCCATTTTTA CATTCTCCAGGTGGGAAACA
8	4q13	rs1026338 rs2278918 rs2278917	IMS-JST026415 IMS-JST026414 IMS-JST026413	UGT2B15	91	CCTCCAGAGCAGTGGCTAGT AATTCTGCCCTGTCATTTGC
9	5q15	rs2255633 rs2255634 rs2255637	IMS-JST059720 IMS-JST059719 IMS-JST059718	ERAP2	56	TCTCAGTTGCCTCAAACCAA GCTGAGGAGAGAGCCAAGAA
10	5q31	rs2237070 rs2237071 rs2237072	IMS-JST017573 IMS-JST017574 IMS-JST017575	TGFBI	71	TGGCAGGCTTTAAAACAGGT CTCAGCCTCAAGTGATCTGC
11	6q27	rs2294500 rs2294499 rs2294498	IMS-JST048012 IMS-JST048011 IMS-JST048010	WDR27	92	CACAAAACCCAGAGCCATCT CCGCAGACCATGAAGGTAGA
12	7p14.1	rs2073543 rs2073542 rs2073541	IMS-JST009932 IMS-JST009931 IMS-JST009930	POU6F2	81	TGTAGATGCAAATACTTCCCTTGA TCAATTGGCTCCCATTCATT
13	9p22	rs3818705 rs2282162 rs2282161	IMS-JST179271 IMS-JST031056 IMS-JST031055	GLDC	73	GGCACATTTTTAGCCTTCTCC GGGGCACCTTGTGAAATAAG
14	10p13	rs2277214 rs2277215 rs2277216	IMS-JST074071 IMS-JST074072 IMS-JST074073	CELF2	52	CACCAATGCAAACCCTCTCT ATAGTATCCCGGCTCCATCC
15	10p12	rs2284947 rs2284946 rs2284945	IMS-JST034951 IMS-JST034950 IMS-JST034949	NRP1	97	TCCTGGAAATTTGGAGAGAGA TTCAAAAGCTCTCCTCTGAAGG
16	10q25	rs911703 rs911701 rs911700	IMS-JST012575 IMS-JST012574 IMS-JST012573	HABP2	50	TCGCCCTGAAGTGTTACTCC GGATGCGATGTCACCTCTTT
17	11p15.3	rs2074312 rs2074311 rs2074310	IMS-JST001081 IMS-JST001080 IMS-JST001079	SOX6	97	TCTGAGGTGGGAGAGAGCAT CTGAGACCCACATACGCTGA
18	11q12.2	rs3794042 rs3794041 rs3794040	IMS-JST151626 IMS-JST151625 IMS-JST151624	TMEM132A	86	CCCCAATCACGTTCACCTAC TGGACTCCAGAAGCAAGTCA
19	11q12.2	rs2023855 rs2023854 rs2023853	IMS-JST151623 IMS-JST151622 IMS-JST151621	TMEM132A	50	GGGAGAGGCATCTCTTAGGC GTGAGCATAGGAGGGCGTAG
20	12p13.3	rs2239020 rs2239021 rs2239022 rs2239023	IMS-JST010022 IMS-JST010023 IMS-JST010024 IMS-JST010025	CACNA1C	100	AACTTAGCCGCAGTTGGAGA TCAGTCTGGGAGAACACAGAGA
21	14q12-q13	rs2281629 rs2281630 rs2281631	IMS-JST030256 IMS-JST030257 IMS-JST030258	NPAS3	94	CCCTTGTTGGATTTTTAGAGGTC GCACCTTGAAAACTGTAGTCTGA
22	14q24	rs872810 rs3742893 rs872809	IMS-JST094083 IMS-JST094084 IMS-JST094085	ACTN1	73	GGGGCTGTCTTTCTCTGACA AGGGCTCCTCCTTCAGTAGC
23	19p13.2	rs387315 rs408312 rs423926	IMS-JST104281 IMS-JST104280 IMS-JST104279	ICAM5	95	GTTGGAACCTGTGGGAAGTC CGCAGGGAAGAGACAGGAC
24	20p11.21	rs2295561 rs2295562 rs2295563	IMS-JST049426 IMS-JST049427 IMS-JST049428	CST9L	55	GACCAGAGAGGGGACCACTT GCTCCTGGCCAATCTGAGT
25	20p11.21	rs2208284 rs2295565 rs2295566	IMS-JST049430 IMS-JST049431 IMS-JST049432	CST9L	71	AGTCCCAACATTCTCCTCCA AGAGTGGGAAGTGCATCTGG
26	20q12	rs2070640 rs2070639 rs2070638	IMS-JST005566 IMS-JST005565 IMS-JST005564	SDC4	87	GGCAGAAGAATGACTCAATGC CCCCTTCCTTTTGCCTATTC
27	22q11.2	rs3788339 rs3788340 rs3788341	IMS-JST145278 IMS-JST145279 IMS-JST145280	RTDR1	82	ACCTACATGGCCACAGCTTC GCACTTTTTGATGGATTCTCC

* Sequence length between all including SNP loci.

Table 2 The observed haplotypes and their frequencies of the regions expected to have three or more haplotypes

Region	rs ID	Allele	Sample number	Observed genotype[a]	Observed frequency	Estimated haplotype[b]	Frequency[c]	PD[d] of each haplotype
#a	rs338564	G/A	n=116	-G-G-C-	0.276	-G-G-C-	0.517	0.736
	rs338565	G/A		-H-H-H[e]-	0.353	-G-G-A-	0.000	
	rs543912	C/A		-H-H-C-	0.129	-A-A-C-	0.104	
				-A-A-A-	0.172	-A-A-A-	0.379	
				-A-A-H-	0.060			
				-A-A-C-	0.009			
#b	rs2296796	G/A	n=127	-A-T-G-	0.134	-G-G-A-	0.319	0.863
	rs2296797	T/G		-H-H-H-	0.228	-G-G-G-	0.154	
	rs2296798	G/A		-H-H-G-	0.126	-G-T-A-	0.000	
				-H-T-G-	0.134	-G-T-G-	0.150	
				-G-G-A-	0.094	-A-G-G-	0.000	
				-G-G-H-	0.118	-A-T-A-	0.000	
				-G-H-H-	0.102	-A-T-G-	0.378	
				-G-G-G-	0.008			
				-G-H-G-	0.047			
				-G-T-G-	0.008			
#c	rs2293140	T/C	n=132	-C-A-T-	0.106	-T-T-G-	0.443	0.798
	rs2293141	T/A		-H-H-H-	0.295	-T-T-T-	0.223	
	rs2293142	G/T		-H-H-T-	0.159	-C-A-G-	0.000	
				-T-T-G-	0.189	-C-A-T-	0.333	
				-T-T-H-	0.212			
				-T-T-T-	0.038			
#d	rs2277214	A/G	n=125	-G-A-T-	0.272	-A-G-C-	0.372	0.749
	rs2277215	A/G		-H-H-H-	0.376	-G-A-C-	0.088	
	rs2277216	C/T		-G-A-H-	0.096	-G-A-T-	0.516	
				-A-G-C-	0.144	-G-G-C-	0.024	
				-H-H-C-	0.048			
				-G-A-C-	0.016			
				-G-H-H-	0.016			
				-H-G-C-	0.032			
#e	rs2074312	A/G	n=134	-C-G-A-	0.209	-C-G-A-	0.466	0.782
	rs2074311	A/G		-H-H-H-	0.358	-C-G-G-	0.153	
	rs2074310	C/T		-C-G-H-	0.149	-C-A-A-	0.000	
				-T-A-G-	0.119	-C-A-G-	0.011	
				-H-H-G-	0.134	-T-A-A-	0.000	
				-C-G-G-	0.007	-T-A-G-	0.369	
				-C-H-H-	0.007			
				-H-A-G-	0.007			
				-C-H-G-	0.007			
#f	rs2281629	C/T	n=127	-C-C-C-	0.094	-C-C-A-	0.067	0.758
	rs2281630	T/C		-H-H-H-	0.402	-C-C-C-	0.346	
	rs2281631	C/A		-C-C-H-	0.047	-C-T-C-	0.055	
				-T-T-A-	0.268	-T-T-A-	0.531	
				-H-H-A-	0.087			
				-C-H-C-	0.055			
				-H-T-H-	0.039			
				-C-T-C-	0.008			
#g	rs2070640	A/C	n=121	-A-C-G-	0.215	-A-A-A-	0.141	0.764
	rs2070639	A/C		-C-A-A-	0.132	-A-C-G-	0.471	
	rs2070638	A/G		-H-H-H-	0.380	-C-A-A-	0.388	
				-H-A-A-	0.132	-C-C-G-	0.000	
				-A-H-H-	0.132			
				-A-A-A-	0.008			

[a] The observed genotypes by sunger sequencing.

[b] The estimated haplotypes by PHASE version 2.1.

[c] Frequencies of each haplotype calculated by PHASE version 2.1.

[d] Power of Discrimination value.

[e] The "-H-" means a SNP locus dedected as heterozygous.