

# Direct molecular haplotyping of long-range genomic DNA with M1-PCR

Chunming Ding\* and Charles R. Cantor\*<sup>†‡</sup>

\*Bioinformatics Program and Center for Advanced Biotechnology, Boston University, 36 Cummington Street, Boston, MA 02215; and <sup>†</sup>SEQUENOM, Inc., 3595 John Hopkins Court, San Diego, CA 92121-1331

Contributed by Charles R. Cantor, April 25, 2003

**Haplotypes, combinations of several phase-determined polymorphic markers, are extremely valuable for studies of disease association and chromosome evolution. Here we describe a technique called M1-PCR (M for “multiplex” and 1 for “single-copy DNA molecules”) that enables direct molecular haplotyping of several polymorphic markers separated by as many as 24 kb. A genomic DNA sample first is diluted to approximately single-copy. The haplotype is directly determined by simultaneously genotyping several polymorphic markers in the same reaction with a multiplex PCR and base extension reaction. This approach does not rely on pedigree data and does not require previous amplification of the entire genomic region containing the selected markers.**

**W**ith the rapid discovery and validation of several million single-nucleotide polymorphisms (SNPs; refs. 1–3), it is now increasingly practical to use genome-wide scanning to find genes associated with common diseases (4, 5). Individual SNPs have statistical power for locating disease susceptibility genes. However, haplotypes, combinations of several phase-determined polymorphic markers, can provide additional statistical power in the mapping of disease genes (6–9), particularly for mapping human complex trait loci (10).

Haplotype determination of several markers for a diploid cell is complicated, because conventional genotyping techniques cannot determine the phases of several different markers. For example, a genomic region with three heterozygous markers can yield eight possible haplotypes. This ambiguity can, in some cases, be solved if pedigree genotypes are available. However, even for a haplotype of only three markers, genotypes of father–mother–offspring trios can fail to yield offspring haplotypes up to 24% of the time (11). Computational algorithms such as expectation-maximization (EM), subtraction, and PHASE are used for statistical estimation of haplotypes (6, 12, 13). However, these computational methods have limitations in accuracy, number of markers, and genomic DNA length. Alternatively, direct molecular haplotyping based on the physical separation of two homologous genomic DNAs before genotyping can be used. DNA cloning, somatic cell hybrid construction, and allele-specific, single-molecule, and long-range PCR (14–17) have been used, and these approaches are largely independent of pedigree information. However, some of these methods (allele-specific and single-molecule PCR) typically are limited to short genomic regions (<3 kb), and all are labor-intensive. Significant efforts have been spent on improving the aforementioned techniques for long-range haplotyping (17, 18), yet it is unclear whether these improvements are robust and simple enough for high-throughput haplotype analysis. Allele-specific PCR coupled with matrix-assisted laser desorption ionization/time-of-flight (MALDI-TOF) MS has been suggested for high-throughput molecular haplotyping (19). For haplotyping of some males, single-sperm typing (20) also can be used for direct molecular haplotyping.

We report here a direct molecule haplotyping approach called M1-PCR (M for “multiplex” and 1 for “single-copy DNA molecules”) that includes single-molecule dilution of genomic DNA for the separation of two homologous genomic DNAs,

followed by direct multiplex genotyping of several markers with the MALDI-TOF MS-based MassARRAY system (SEQUENOM; Fig. 1). In contrast to other PCR-based haplotyping techniques, the M1-PCR method does not require the amplification of the entire genomic region. Instead, only ≈100 bp of the genomic region surrounding each SNP is amplified by PCR. As a result, the distance between two SNPs in the genome does not affect the haplotyping efficiency. Thus, haplotyping can be achieved for virtually any genomic distance, as long as the genomic DNA is largely intact (no physical breaks). A genotyping success rate of ≈100% for single-copy DNA molecules has been achieved. Additionally, the multiplex genotyping assay approach enables direct haplotype determination without pedigree genotype information. High-throughput haplotyping (>1,000 haplotypes per day) can be achieved by incorporating the M1-PCR technique into the highly automated MassARRAY system.

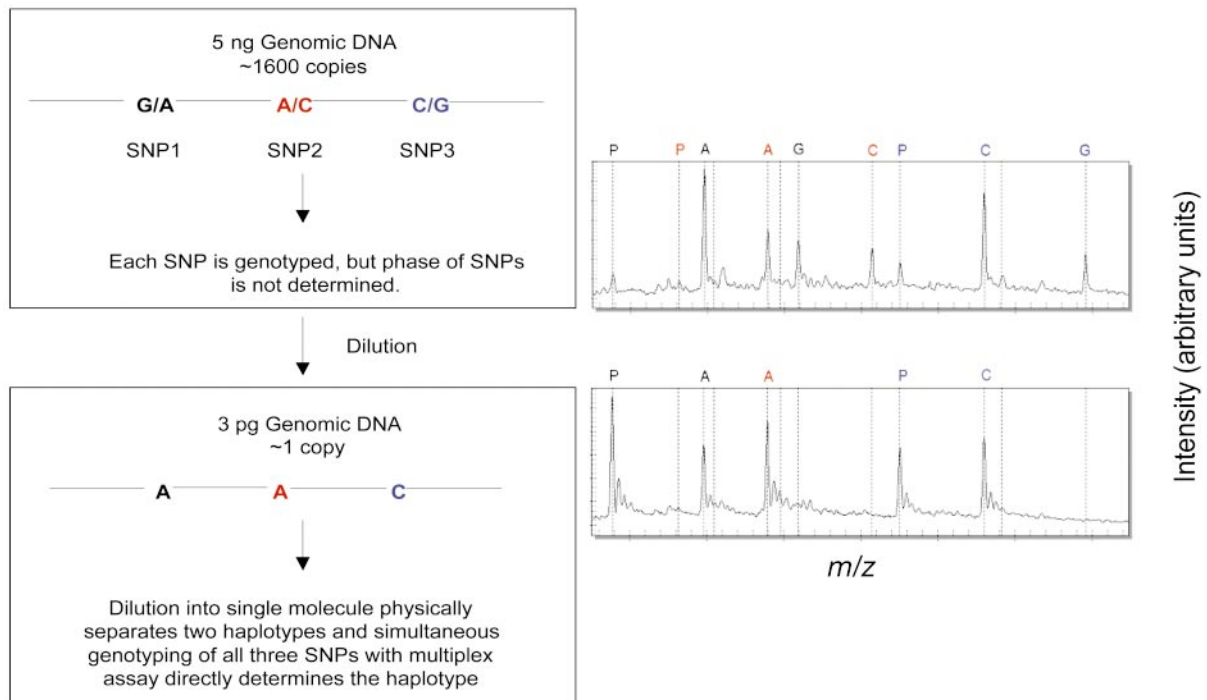
## Materials and Methods

**Genomic DNAs and Oligonucleotides.** Human genomic DNA samples used for haplotyping of the cholesteryl ester transfer protein locus were provided by SEQUENOM. These DNAs were isolated, by using the Puregene DNA isolation kit (Gentra Systems), from blood samples purchased from the Blood Bank of San Bernardino and Riverside Counties (San Bernardino, CA). The personal background of the blood donors is not accessible for these samples. Human genomic DNA samples for haplotyping of a 24-kb segment on chromosome 5q31 were Centre d'Etude du Polymorphisme Humain (CEPH)/French Pedigree 66 (repository nos. GM12547–GM12554 and GM12556–GM12559) DNAs purchased from Coriell Cell Repositories (Camden, NJ). Information on SNPs and oligonucleotides for genotyping is provided in Table 1.

**Genotyping and Haplotyping Analysis.** Genotyping analyses were carried out by using the MassARRAY system, following the protocol provided by SEQUENOM except for the genomic DNA concentrations. Briefly, genomic DNAs were amplified by PCR using HotStarTaq DNA Polymerase (Qiagen, Valencia, CA). PCR primers (Table 1) were used at 200 nM final concentrations for a PCR volume of 5  $\mu$ l. The PCR condition was 95°C for 15 min for hot start, followed by denaturing at 94°C for 20 sec, annealing at 56°C for 30 sec, extension at 72°C for 1 min for 45 cycles, and finally incubation at 72°C for 3 min. PCR products first were treated with shrimp alkaline phosphatase (SEQUENOM) for 20 min at 37°C to remove excess dNTPs. ThermoSequenase (SEQUENOM) was used for the base extension reactions. In contrast to SEQUENOM's standard Homogenous MassEXTEND protocol, extension primers were at a final concentration of 1.2  $\mu$ M in 9- $\mu$ l reactions. The base extension reaction condition was 94°C for 2 min, followed by 94°C for 5 sec, 52°C for 5 sec, and 72°C for 5 sec for 40 cycles. All reactions (reverse transcription, PCR amplification, and base extension)

Abbreviation: SNP, single-nucleotide polymorphism.

<sup>†‡</sup>To whom correspondence should be addressed. E-mail: ccantor@sequenom.com.



**Fig. 1.** Multiplex genotyping of single-copy DNA molecules for haplotype analysis. Traditional genotyping methods using 5 ng of genomic DNA ( $\approx 1,600$  copies of genomic templates) yield the genotypes of each individual SNP marker, but the phases of these SNPs are not determined (*Upper Right*). Simultaneous genotyping of several markers using multiplex assays with single-copy DNA molecules allows haplotyping analysis, because the two alleles can be physically separated at extremely dilute DNA concentrations (*Lower Right*; please refer to *Materials and Methods* for details of haplotyping and genotyping). In contrast to other molecular haplotyping methods, the entire haplotype block is not amplified in this approach. Instead, only  $\approx 100$  bp around each individual SNP is amplified for genotyping, resulting in very high efficiency of PCR amplification from single-copy DNA molecules. The SNP markers can be as far apart as desired, as long as physical breaks between the SNP markers exist only in a small proportion of the genomic DNAs. Peaks identified as P are from unextended primers.

were carried out in a GeneAmp PCR System 9700 thermal cycler (Applied Biosystems). The final base extension products were treated with SpectroCLEAN resin (SEQUENOM) to remove salts in the reaction buffer. This step was carried out with a

Multimek 96-channel autopipette (Beckman Coulter), and 16  $\mu$ l of resin/water suspension was added into each base extension reaction, making the total volume 25  $\mu$ l. After a quick centrifugation (2,000 rpm, 3 min) in an Eppendorf Centrifuge 5810,

**Table 1. SNPs and primers for the haplotyping assays**

SNP ID	Chromosome position	PCR primer 1	PCR primer 2	Extension primer	Terminator mix*
rs289741	47282625	TCTACCAGCTTGCTCCCTC	AAGTCCATCAGCAGCAGCAG	GGGAGTCAGCCCAGCTC	ACT
rs289742	47282337	ACTGGTGAGACAATCCCTTC	CCACTGGCATTAAAGTGCTG	AGCCACAGAAGAAGGACTCC	ACT
rs289744	47281997	TACCAGAAACCAGACCTCTG	AGTGTGGACAGAAAGTGAG	TGAGGATGGTGGGAGGG	ACT
rs289744 <sup>†</sup>	47281997	TCTACCAGAAACCAGACCTC	AGTGCTGGACAGAAAGTGAG	ACCTCTGAGGGCCCTTAC	CG
rs2228667 <sup>‡</sup>	47282820	CTCGAGTGATAATCTCAGGG	AGGTAGTGTTCACGCCCTC	TGATGATGTCGAAGAGGCTCATG	CG
rs5882 <sup>‡</sup>	47284007	TTACGAGACATGACCTCAGG	GCATTTGATTGGCAGAGCAG	CTGCAGGAAGCTCTGGATG	CG
rs5882 <sup>‡</sup>	47284007	GCATTTGATTGGCAGAGCAG	TTACGAGACATGACCTCAGG	AGAGCAGCTCCGAGTCC	ACT
rs5880 <sup>‡</sup>	47285008	GCAGCACATACTGGAAATCC	TTTCTCTCCCCAGGATATCG	GCTTTTTCTTAGAATAGGAGG	ACT
rs5881 <sup>‡</sup>	47288087	AGATCTTGGGCATCTTGAGG	ACCCTGTCTTCCACAGGTT	TGGGCCTGGCTGGGGAAGC	CG
rs5881 <sup>‡</sup>	47288087	ACCCTGTCTTCCACAGGTT	AGATCTTGGGCATCTTGAGG	TGTCTCCACAGGTTGTCCGC	ACT
rs291044 <sup>‡</sup>	47288647	GTAAAACCTGCAGCTGAGGAG	TGACTAGGTCAGGTCCCCTC	GGAGTATTTAAAGGAGAGACACA CTAG	CG
rs291044 <sup>‡</sup>	47288647	TGACTAGGTCAGGTCCCCTC	GTAAAACCTGCAGCTGAGGAG	CCCTCGTGCCACAGCCT	ACT
rs2033254 <sup>‡</sup>	47290114	GGACATCAAAGGAACAGGAC	ACTACAATATTGGGCAGGC	CAAGGGGCTAAGGGAGAAG	ACT
IGR2198a_1 <sup>§</sup>	506266 <sup>¶</sup>	GGGTTGCATGAGCATTAAAGT	CACATCAAGGATAAGACTGC	ATCTCTCAGTAGACGAAC	AC
IGR2175a_2 <sup>§</sup>	495082 <sup>¶</sup>	TGGCCTTGATTCAAACCTCG	AGATGAAGGAAATCCCAAGG	TGCCACTAACATACATAGTAAC	AC
IGR2150a_1 <sup>§</sup>	482171 <sup>¶</sup>	CCTTGGCTTGATAGTCAAAC	ATTTGGAGGAGTGCAGAGAG	AGTCAAACCTCACCAC	AC

\*ACT mix is ddATP/ddCTP/ddTTP/dGTP in which dd indicates the 2',3'-dideoxynucleotide. Similarly, CG mix is ddCTP/ddGTP/dATP/dTTP.

<sup>†</sup>Multiplex assay 1.

<sup>‡</sup>Multiplex assay 2.

<sup>§</sup>SNP ID from Daly et al. (23); see also [www-genome.wi.mit.edu/humgen/IBD5/haplodata.html](http://www-genome.wi.mit.edu/humgen/IBD5/haplodata.html).

<sup>¶</sup>Position of SNP from Daly et al. (23).

**Table 2. Sample haplotype analysis with a triplex genotyping assay**

Repeat	Genotype calls*
1	GGC
2	GGC
3	—†
4	—GC‡
5	—
6	GGC
7	—
8	ACA
9	—GC‡
10	A/GC/GA/C§
11	ACA
12	ACA

Genotypes of three SNP markers were determined with a triplex assay from 3 pg of genomic DNA.

\*The three SNP genotypes G, G, and C are represented as GGC, and A, C, and A are represented as ACA.

†Failed to genotype any of the three SNPs.

‡Failed to genotype the first SNP; the other two SNPs are G and C, respectively.

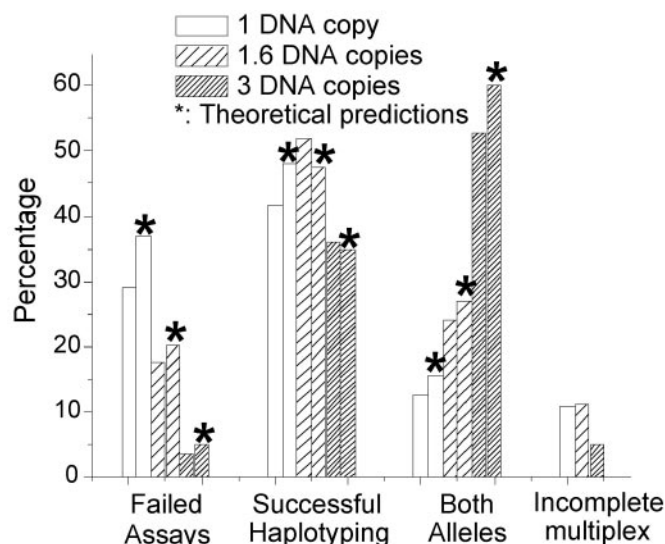
§Failed to separate the two alleles; thus, the genotypes are A/G, C/A, and A/C for the three SNPs.

≈10 nl of reaction solution was dispensed onto a 384 format SpectroCHIP (SEQUENOM) prespotted with a matrix of 3-hydroxypicolinic acid (3-HPA) by using a SpectroPoint nanodispenser (SEQUENOM). A modified Bruker Biflex matrix-assisted laser desorption ionization/time-of-flight mass spectrometer (Billerica, MA) was used for data acquisitions from the SpectroCHIP. Genotyping calls were made in real time with MASSARRAY RT software (SEQUENOM). For haplotyping analysis, multiplex genotyping assays were carried out by using 3 pg (or approximately one copy of genomic template, unless otherwise specified) of genomic DNA. Each SNP from every individual was genotyped individually by using 5 ng of genomic DNA to confirm that M1-PCR does not yield wrong genotypes.

**Analysis of Effects of Genomic DNA Concentration on Haplotyping.** To calculate the percentage of failed assays, we simply counted all failed assays (no call for any SNP) divided by the total number of assays. Twelve replicates for 12 individuals and 18 replicates for 6 individuals were carried out. The percentage of incomplete assays was calculated in the same way. To calculate percentage of successful haplotyping and both alleles, we excluded the data from those individuals with homozygous haplotypes. Theoretical predictions were based on the Poisson distribution of extremely dilute DNA solutions, according to a published method (21).

## Results and Discussion

We first investigated the effect of genomic DNA concentration on haplotyping efficiency for the M1-PCR technique. We used 3, 5, and 9 pg (equivalent to 1, 1.6, and 3 genomic template copies) of genomic DNA for PCR amplification and genotyping of three SNPs (GenBank SNP ID: rs289741, rs289742, and rs289744; Table 1) in the cholesteryl ester transfer protein region from 12 individuals. Each triplex assay was repeated 12–18 times to evaluate the PCR and haplotyping efficiency. A typical assay result is summarized in Table 2. The copy number of the genomic DNA region of interest in extremely dilute DNA solutions is estimated by the Poisson distribution (21). If a genomic DNA is diluted to approximately one copy of the genomic region of interest, zero, one, two, or more copies of this region will be obtained because of stochastic fluctuations. These fluctuations will lead to different outcomes for molecular haplotyping by the M1-PCR method. Haplotyping results are categorized into four



**Fig. 2.** Effect of genomic DNA concentration on haplotyping efficiency. Approximately 3, 5, and 9 pg (or 1, 1.6, and 3 copies of genomic templates) were used for haplotyping of three SNP markers (GenBank SNP ID: rs289741, rs289742, and rs289744; Table 1) in the cholesteryl ester transfer protein region. The DNA copy number in a specific reaction is estimated by the Poisson distribution. The haplotyping result can be a failed assay, successful haplotyping, the presence of both alleles (no phase determination for the markers), or an incomplete multiplex. Except for incomplete multiplexes, values are percentages from 54 to 144 individual multiplex assays (see *Materials and Methods* for details of the calculation), followed by predicted values (\*) determined by using the Poisson distribution.

groups (Table 2). Failed assays can result either from failed PCR amplifications from single-copy DNAs or from simply the absence of a template caused by the stochastic fluctuations of extremely dilute DNA solutions. Partially failed genotyping calls (or incomplete multiplexes) are those that have only one or two SNPs successfully genotyped. These incomplete multiplexes are most likely attributable to unsuccessful PCR amplifications for the genomic DNA regions (as opposed to the absence of the genomic regions) of the failed SNP assays, because in most cases the three closely positioned SNP markers (separated by <628 bp) are either all present or all absent in a PCR. The Poisson distribution also may result in the presence of both alleles in the solution and hence the inability to resolve the phase of the SNPs. Successful haplotyping analysis can be achieved when a single copy of the allele or multiple copies of the same allele are present, and the genotyping is successful.

Incomplete multiplex genotyping can be used to estimate the efficiency of genotyping from single-copy DNA molecules. A partial genotyping call suggests the presence of the SNP DNA but a failure to genotype some of the SNPs. We typically observe 5–10% incomplete multiplex genotyping calls (Fig. 2), suggesting a PCR efficiency of 90–95% with single-copy DNA molecules. This approach might overestimate the PCR efficiency, because we did not take the completely failed assays into account. We also carried out a detailed comparison between observed and theoretical values of failed assays, successful haplotyping, and the presence of both alleles (Fig. 2; also see *Materials and Methods* for details of calculation). Theoretical values are based on the Poisson distribution of extremely dilute DNA solutions and the assumption of 100% PCR amplification efficiency. The close agreement between theoretical estimates and experimental observations substantiates the earlier estimate of extremely high PCR efficiency with single-copy DNA molecules. This high PCR efficiency is likely attributable to the use of very short amplicons (typically only 100 bp) and the high sensitivity of matrix-assisted

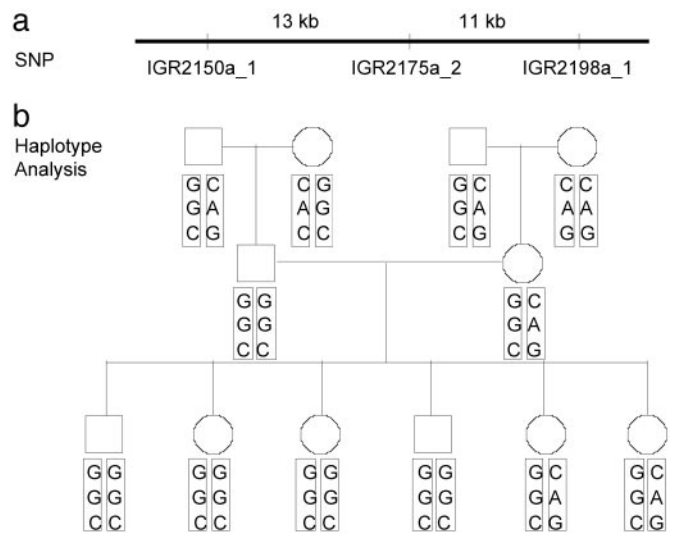
SNP	A	B	C	D	E	F	G
Multiplex Assay 1	+	+	+			+	+
Multiplex Assay 2				+	+		+
Individual 6, Assay 1	A	G	A			G	T
	C	G	G			G	C
Individual 6, Assay 2				G	G		T
				A	G		T
Individual 6, Haplotypes	A	G	A	G		G	T
	C	G	G	G		G	C

**Fig. 3.** Overlapping multiplex genotyping assays with single-copy DNA molecules. Seven SNP markers (A, rs289744; B, rs2228667; C, rs5882; D, rs5880; E, rs5881; F, rs291044; and G, rs2033254) from an 8-kb genomic region of the cholesteryl ester transfer protein locus were chosen (Table 1). Two five-plex genotyping assays were designed for these seven markers, and the overlapping heterozygous SNPs were used to obtain the entire haplotype of the markers. Assays on individual 6 were used to demonstrate how this approach was carried out. Multiplex assay 1 determined the haplotype of five SNPs as AGAGT and CCGGC. Multiplex assay 2 determined the other haplotype of five SNPs as GGGCT and AGGTT. Then, the genotypes of the overlapping SNPs (SNPs C, E, and F) were used to combine the two haplotypes of five SNPs into a haplotype of seven SNPs covering the entire region under investigation.

laser desorption ionization/time-of-flight MS detection of DNA oligonucleotides, consistent with our previous finding in transcriptional profiling, where as few as five cDNA copies can be quantified (22). High PCR efficiency is an absolute requirement for multiplex genotyping assays with single-copy DNA molecules. Low PCR efficiency (even at 50%) will cause allele dropout in genotyping assays when only a few DNA templates are present. The problems become more severe in multiplex genotyping assays, resulting in not only allele dropouts but also wrong haplotyping calls. This high PCR efficiency is also absolutely crucial for high-throughput haplotyping analysis. With our current PCR efficiency, we can achieve 40–45% haplotyping efficiency with a single reaction by using 3–4.5 pg of genomic DNA (close to the theoretical upper limit determined by the Poisson distribution of extremely dilute DNA samples). A replicate of four independent multiplex genotyping assays will enable  $\approx 90\%$  direct haplotyping success rate.

We next demonstrate an approach for determining haplotypes where there are too many markers to be determined in one multiplex genotyping assay. Overlapping informative SNPs can be used to combine haplotypes from several multiplex assays. Seven SNP markers in an 8-kb cholesteryl ester transfer protein genomic region were chosen, and two overlapping five-plex genotyping assays were used for haplotyping analysis (Fig. 3). We were able to determine the haplotypes of all 12 individuals for that genomic region, with absolutely no optimization of the assay system. A larger-scale study is needed to further validate the practicability of this approach for high-throughput haplotyping analysis.

It is inherent in the M1-PCR design that distance between SNPs is not a factor for successful haplotyping, as long as not many copies of the genomic DNA have physical breaks between the markers. This distance-independence is significant because allele-specific PCR and previous single-molecule PCR are limited to a short (at most a few kilobases) genomic region, attributable to lower PCR efficiency in amplifying the entire genomic DNA region (typically a few kilobases) containing the selected SNPs at single DNA molecule level. We chose three SNPs spanning  $>24$  kb in the human chromosome 5q31 region (23) and determined the haplotypes of two Centre d'Etude du Polymorphisme Humain (CEPH) families with a total of 24 individuals. The haplotypes of these individuals were first determined by our direct molecule haplotyping approach, without



**Fig. 4.** Haplotyping of three SNP markers spanning 24 kb in the chromosome 5q31 region. The SNPs chosen were IGR2150a.1, IGR2175a.2, and IGR2198a.1 (Table 1). The results from Centre d'Etude du Polymorphisme Humain (CEPH)/French Pedigree 66 are shown. (a) Relative positions and distances of the three SNP markers are shown. (b) Pedigree genotypes and haplotypes. Genotypes of the paternal grandfather (individual in upper left corner) in the figure are G/C for SNP IGR2150a.1, G/A for SNP IGR2175a.2, and C/G for SNP IGR2198a.1. The two haplotypes for the same individual are GGC and CAG. The son of that individual is homozygous for all three SNPs. Thus, his haplotype is GGC for both alleles.

the pedigree information. The haplotypes also were determined by genotyping individual SNPs and using the pedigree information. These two approaches produced consistent results (Fig. 4). Compared with the haplotyping of shorter genomic regions, there was a significantly higher percentage of wrong haplotyping calls. Thus, we used more replicates (10–14) of single-molecule PCRs for each individual. The false haplotype calls were identified clearly because more correct calls existed in the replicate experiments. The wrong haplotype calls are very likely caused by the breaks between the SNP markers in the original genomic DNA preparations. The vendor (Coriell Cell Repositories) of the genomic DNA states that the average genomic DNA size is 100–150 kb. Assuming that the DNA breaks are equally likely in all genomic regions, there would be an  $\approx 20\%$  chance of breakage between two markers 24 kb apart. The breaks between the selected SNP markers will, in effect, eliminate the linkage between the markers in the same haplotype and cause wrong haplotype calls. Recent publications on haplotype block structures (23, 24) indicate that the typical haplotype size is  $\approx 10$ –50 kb. Thus, our approach can be applied directly for long-range genomic DNA haplotyping.

We have shown that single-molecule haplotyping (15), a method tried and apparently abandoned a decade ago because of technical problems, can be resurrected and made to work robustly and at much longer genomic ranges, because of highly efficient M1-PCR and automated, sensitive matrix-assisted laser desorption ionization/time-of-flight MS detection. M1-PCR does not require the amplification of the entire genomic regions containing the SNPs of interest. Instead, only  $\approx 100$  bp of the genomic region surrounding each SNP is amplified by multiplex PCR, resulting in  $\approx 100\%$  PCR efficiency from single-copy DNA molecules. In addition, the M1-PCR has been incorporated into the highly automated MassARRAY system so that high-throughput (a few thousand haplotyping assays per day), direct molecular haplotyping can be achieved. The M1-PCR approach is independent of pedigree genotype information. Thus, it is not

necessary to collect families, set up cell lines, or subclone genomic DNAs to infer haplotypes, saving cost and time and eliminating various artifacts, which enables direct and rapid haplotyping in clinical settings or conventional case control studies. In addition, this approach produces haplotypes definitively and unambiguously where statistical and computational haplotyping falls short. The amounts of DNA needed are

≈100–1,000 times less than those currently used for genotyping; hence, whole genome scans will now be possible on normally accessible blood samples.

We thank Dr. Zhiping Weng for critical comments on the manuscript. This work was supported by a grant from SEQUENOM to Boston University.

1. Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., Sherry, S., Mullikin, J. C., Mortimore, B. J., Willey, D. L., *et al.* (2001) *Nature* **409**, 928–933.
2. Marth, G., Yeh, R., Minton, M., Donaldson, R., Li, Q., Duan, S., Davenport, R., Miller, R. D. & Kwok, P. Y. (2001) *Nat. Genet.* **27**, 371–372.
3. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., *et al.* (2001) *Science* **291**, 1304–1351.
4. Grupe, A., Germer, S., Usuka, J., Aud, D., Belknap, J. K., Klein, R. F., Ahluwalia, M. K., Higuchi, R. & Peltz, G. (2001) *Science* **292**, 1915–1918.
5. Hirschhorn, J. N., Lohmueller, K., Byrne, E. & Hirschhorn, K. (2002) *Genet. Med.* **4**, 45–61.
6. Templeton, A. R., Sing, C. F., Kessling, A. & Humphries, S. (1988) *Genetics* **120**, 1145–1154.
7. Kruglyak, L. (1999) *Nat. Genet.* **22**, 139–144.
8. Judson, R., Stephens, J. C. & Windemuth, A. (2000) *Pharmacogenomics* **1**, 15–26.
9. Martin, E. R., Gilbert, J. R., Lai, E. H., Riley, J., Rogala, A. R., Slotterbeck, B. D., Sipe, C. A., Grubber, J. M., Warren, L. L., Conneally, P. M., *et al.* (2000) *Genomics* **63**, 7–12.
10. Cardon, L. R. & Abecasis, G. R. (2003) *Trends Genet.* **19**, 135–140.
11. Hodge, S. E., Boehnke, M. & Spence, M. A. (1999) *Nat. Genet.* **21**, 360–361.
12. Clark, A. G. (1990) *Mol. Biol. Evol.* **7**, 111–122.
13. Stephens, M., Smith, N. J. & Donnelly, P. (2001) *Am. J. Hum. Genet.* **68**, 978–989.
14. Ruano, G. & Kidd, K. K. (1989) *Nucleic Acids Res.* **17**, 8392.
15. Ruano, G., Kidd, K. K. & Stephens, J. C. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 6296–6300.
16. Yan, H., Papadopoulos, N., Marra, G., Perrera, C., Jiricny, J., Boland, C. R., Lynch, H. T., Chadwick, R. B., de la Chapelle, A., Berg, K., *et al.* (2000) *Nature* **403**, 723–724.
17. McDonald, O. G., Krynetski, E. Y. & Evans, W. E. (2002) *Pharmacogenetics* **12**, 93–99.
18. Michalatos-Beloin, S., Tishkoff, S. A., Bentley, K. L., Kidd, K. K. & Ruano, G. (1996) *Nucleic Acids Res.* **24**, 4841–4843.
19. Tost, J., Brandt, O., Boussicault, F., Derbala, D., Caloustian, C., Lechner, D. & Gut, I. G. (2002) *Nucleic Acids Res.* **30**, e96.
20. Li, H., Cui, X. & Arnheim, N. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 4580–4584.
21. Stephens, J. C., Rogers, J. & Ruano, G. (1990) *Am. J. Hum. Genet.* **46**, 1149–1155.
22. Ding, C. & Cantor, C. R. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 3059–3064.
23. Daly, M. J., Rioux, J. D., Schaffner, S. F., Hudson, T. J. & Lander, E. S. (2001) *Nat. Genet.* **29**, 229–232.
24. Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., *et al.* (2002) *Science* **296**, 2225–2229.