

High-throughput SNP genotyping by single-tube PCR with T_m -shift primers

Jun Wang¹, Karen Chuang¹, Mandeep Ahluwalia¹, Sarika Patel², Nanette Umblas¹, Daniel Mirel¹, Russell Higuchi¹, and Soren Germer^{1,2}

BioTechniques 39:885-893 (December 2005)
doi 10.2144/000112028

Despite many recent advances in high-throughput single nucleotide polymorphism (SNP) genotyping technologies, there is still a great need for inexpensive and flexible methods with a reasonable throughput. Here we report substantial modifications and improvements to an existing homogenous allele-specific PCR-based SNP genotyping method, making it an attractive new option for researchers engaging in candidate gene studies or following up on genome-wide scans. In this advanced version of the melting temperature (T_m)-shift SNP genotyping method, we attach two GC-rich tails of different lengths to allele-specific PCR primers, such that SNP alleles in genomic DNA samples can be discriminated by the T_m s of the PCR products. We have validated 306 SNP assays using this method and achieved a success rate in assay development of greater than 83% under uniform PCR conditions. We have developed a standalone software application to automatically assign genotypes directly from melting curve data. To demonstrate the accuracy of this method, we typed 592 individuals for 6 SNPs and showed a high call rate (>98%) and high accuracy (>99.9%). With this method, 6–10,000 samples can be genotyped per day using a single 384-well real-time thermal cycler with 2–4 standard 384-well PCR instruments.

INTRODUCTION

The identification of genes affecting complex traits is a very difficult and challenging task (1–3). A key development in complex trait analysis is the establishment of large collections of closely spaced genetic markers, including single nucleotide polymorphisms (SNPs). The quickly rising number of publicly available, validated SNPs has made the task of performing genome-wide association studies feasible. Furthermore, many new SNP genotyping technologies have been developed in the past few years. These technologies are based on various methods of allele discrimination, target amplification, and detection platforms (4–6). The detection mechanisms roughly fall into two categories: solid-phase-mediated detection and homogeneous detection. Solid-phase-mediated detection uses a solid support in the detection step, including mass spectrometry (7–9), micro-

arrays (10–12), microbeads (13), and electrophoresis (14). In homogeneous detection, genotyping is performed in solution from beginning to end, and no separation or purification steps are needed. Homogeneous detection methods frequently rely on fluorescence detection, such as fluorescence resonance energy transfer (FRET) detection (15), including TaqMan[®] (16) and molecular beacons (17,18), or allele-specific PCR with real-time fluorescence detection (19–21).

Most recently, several highly multiplexed and ultra-high-throughput genotyping systems have become generally available (22–24). However, most of these systems are based on solid-phase-mediated detection and are not very flexible or require significant up-front instrument investments. For projects focusing on a limited number of SNPs in candidate gene studies or following up on genome-wide scans, or projects that do not warrant large-scale initial investments, there is still

a need for very inexpensive, highly flexible SNP genotyping methods with a reasonable throughput.

We have previously reported a very inexpensive, homogeneous melting temperature (T_m)-shift genotyping method that relies on allele-specific PCR without labeled oligonucleotides (25). The T_m -shift method uses (i) two allele-specific primers, each of which contains a 3'-terminal base that corresponds to one of the two SNP allelic variants, (ii) a reverse primer that amplifies both alleles, and (iii) a fluorescent dye, SYBR[®] Green I (Molecular Probes, Eugene, OR, USA). Depending on the sample genotype, either one or the other or both allele-specific primer(s) amplify. With GC-rich tails of unequal length attached to allele-specific primers, the PCR product has a distinct T_m that depends on which of the two primers is responsible for the amplification, and genotypes can be determined by inspection of a melting curve on a real-time PCR instrument.

¹Roche Molecular Systems, Alameda, CA and ²Hoffman-La Roche, Nutley, NJ, USA

Table 1. GC-Tailed Primers

SNP	Primer Type	GC-Tailed Primer Sequence
rs1367117	AS1	5'- <u>GCGGGCAGGGCGGC</u> GAAAGACCAGCCAGTGCAC-3'
	AS2	5'- <u>GCGGGC</u> TGAAGACCAGCCAGTGCAT-3'
	CP	5'-CAAGGCTTTGCCCTCAGGGTT-3'
rs734351	AS1	5'- <u>GCGGGCAGGGCGGC</u> TGCACCTTTTGGACATTC-3'
	AS2	5'- <u>GCGGGC</u> TGCACCTTTTGGACATTT-3'
	CP	5'-GAACGCACGTCCCTTGTC-3'
rs2585	AS1	5'- <u>GCGGGCAGGGCGGC</u> CCCGTGTGGACTCAACTCAG-3'
	AS2	5'- <u>GCGGGC</u> CCCGTGTGGACTCAACTCAA-3'
	CP	5'-TTTGCCGAAATATTAGCGT-3'
rs1004446	AS1	5'- <u>GCGGGCAGGGCGGC</u> TTGTGCAGCCCTAAAGG-3'
	AS2	5'- <u>GCGGGC</u> TTGTGCAGCCCTAAAGA-3'
	CP	5'-CAGGAAAGGCCATTGTGGAGA-3'

The GC tails are bolded and underlined. AS1 and AS2, allele-specific primers; CP, common primer.

Here we describe an advanced single-tube T_m -shift SNP genotyping method that allows a >98% genotype call rate, >99.9% accuracy, and >83% success rate in novel assay design under uniform assay conditions. The genotyping assay is very inexpensive, in that it requires only three standard (unlabeled) PCR primers, such that the cost per new assay is less than US \$30.

MATERIALS AND METHODS

Genomic DNA

Genomic DNA samples from the human cell line for assay validation were purchased from the Coriell Institute for Medical Research (Camden, NJ, USA). The clinical genomic DNA samples were obtained from the Genomics Collaborative (Boston, MA, USA), as described in Ardlie et al. (26).

PCR Amplification and Melting Analysis

PCRs were performed with 10 ng genomic DNA in 15- μ L volumes for the data shown here. However, we have also tested assay performance with lower amounts of input DNA and now routinely use 4 ng genomic DNA for genotyping. We use a heat-activatable modified (Gold) version of Stoffel fragment polymerase (27) to

provide a simplified hot start as previously described (available from the authors upon request) (19). PCRs were set up as follows: 0.2 μ M each of the three primers; 1.8 U of Stoffel Gold polymerase (gift from David Birch, Roche Molecular Systems, Alameda, CA, USA); 10 mM Tris-HCl, 40 mM KCl, pH 8.0; 2 mM $MgCl_2$; 50 μ M each dATP, dCTP, and dGTP; 25 μ M dTTP; 75 μ M dUTP; 0.2 \times SYBR Green I; 5% dimethyl sulfoxide (DMSO); 2.5% glycerol; 0.25 μ M ROX dye (optional) (Sigma-Aldrich, St. Louis, MO, USA).

All PCRs for genotyping experiments were set up automatically by using a Biomek[®] 2000 or Biomek FX robotic system (Beckman Coulter, Fullerton, CA, USA). PCR was performed on the GeneAmp[®] PCR System 9700 (dual 384-well blocks) or on the ABI PRISM 7900 HT Sequence Detection System (thermal cyclers from Applied Biosystems, Foster City, CA, USA) in initial optimization experiments. The melting curve analysis was performed on the ABI PRISM 7900 HT. PCR was performed in standard 384-well plates (ISC Bioexpress, Salt Lake City, UT, USA), sealed with optical adhesive film (Applied Biosystems). An initial enzyme heat-activation step of 12 min at 95°C was followed by 40 cycles of a 3-step amplification profile of 20 s at 95°C for denaturation, 1 min at 58°C for annealing, and 30 s at 72°C for extension. Melting curves were obtained using the ABI PRISM 7900

HT onboard software (SDS 2.1) by using the default "dissociation step" to measure the fluorescence intensity of the PCR product in a linear denaturation ramp from 60°C–92°C. We do not use uracil DNA glycosylase (UNG) to sterilize carryover contamination because UNG might degrade amplification products during the time interval between PCR amplification and melting curve analysis. However, we continue to use dUTP in the PCRs to permit UNG-mediated sterilization if necessary, although carryover contamination is very unlikely because these are closed-tube assays that do not require any post-PCR liquid handling or transfer. We have also performed PCR amplification and melting analysis in a 96-well format on the cheaper thermal cyclers, ABI GeneAmp 5700 and ABI PRISM 7000 Sequence Detection Systems (Applied Biosystems) with comparable results (data not shown).

Primer Design

For each SNP, two forward allele-specific primers with the 3' base of each primer matching one of the SNP allele bases, and a reverse common primer, were first designed as previously described (25,28). Primers were designed using either Oligo[®] 6 (Molecular Biology Insights, Cascade, CO, USA) or an in-house primer design program. Allele-specific primers were designed to have a T_m between 58.3°C–63.5°C (optimally between 59°C–62°C) and a length ranging from 14–27 bp (optimally from 15–22 bp). Common primers were designed with a T_m from 62°C–75°C (optimally between 63°C–70°C) with a length ranging from 22–30 bp (optimally from 22–27 bp). The common primer was typically placed no more than 20 bp downstream from the SNP, giving relatively short PCR products that allow reasonably good amplification efficiency with the Stoffel polymerase under conditions that favor allele discrimination. To each of the two allele-specific primers, GC tails of different lengths were then added (Table 1). The long 14-bp GC tail has the sequence 5'-GCGGGCAGGGCGGC-3' and the short 6-bp GC tail has the sequence 5'-GCGGGC-3'. To achieve a good

success rate, we routinely attach the long tail to an allele-specific primer that has the higher T_m base (G or C) at its 3' end, and the short tail to the other allele-specific primer with the lower T_m base (A or T). The allele-specific primers can be designed for either strand of DNA, and for particularly important SNPs, assays for both strands can be designed at once (thus increasing the chance of a successful assay).

Data Analysis and Genotype Calling

The melting curve analysis was performed on an ABI PRISM 7900 HT using the onboard software (SDS 2.1) to calculate the negative derivative of the change in fluorescence. The melting analysis values were exported from the onboard software as a text file and imported into a customized Microsoft® Excel® spreadsheet or a standalone software that performs a series of normalization and quality-control algorithms and allows a semiautomated scoring of the genotypes (details in the Results section). In brief, we plotted each sample on a scatter plot of their peak areas in the two relevant temperature ranges in the melting profiles. We then used a *k*-means clustering algorithm to automatically classify samples into four genotype categories based on nearest-centroid sorting or three if the rare allele homozygote is not present in a particular run. The software flags samples with an extremely high *t*-statistic (implying low conditional probability) belonging to a particular cluster and samples that are shifted more than 1°C (default) during the temperature normalization as outliers.

RESULTS

Advanced T_m -Shift SNP Genotyping Method

The basic concept of T_m -shift genotyping has been previously described (25). In that earlier version of the technology, a GC tail of 26 GC-rich bases was added to one of the allele-specific primers, and no tail was added to the other allele-specific primer.

However, when testing this system on a larger number of SNPs, we found the success rate to be low because the amplification products from the primers with the 26-bp GC tail tended to dominate the PCR, resulting in melting curve peaks of very different height. In the advanced version reported here, we have changed the length of the GC tails such that both allele-specific primers now have a GC-rich tail (Figure 1A). The two GC tails are based on the sequence of the original 26-bp tail (25), with the shorter tail containing 6 bp and the longer tail containing 14 bp from the 5' end of the 26-bp tail (Table 1). This difference of 8 bp allows sufficient discrimination between the two allele-specific products in the melting profiles but has a minimal impact on priming and amplification efficiencies. In Figure 1B, we show examples of T_m -shift genotyping assays for 4 SNPs on samples of known genotype previously determined by real-time allele-specific PCR. The three genotypes for each SNP locus can be clearly distinguished by their melting profiles.

We tested other sets of GC tails with various lengths derived from the 26-bp tail sequence,

each with a 7 to 8 GC of base-length difference. The 6 GC/14 GC tail set reported here was the combination that gave the highest success rate in terms of providing adequate temperature

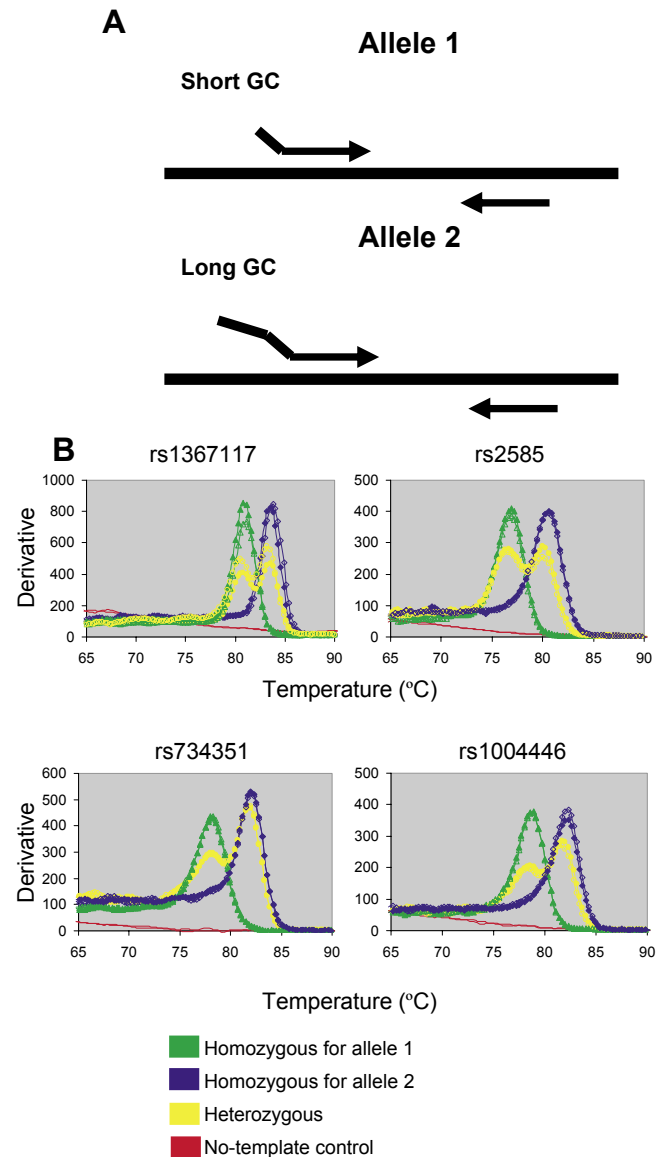


Figure 1. T_m -shift SNP genotyping. (A) Principle of method. Genomic DNA is amplified in a multiplex reaction with two forward allele-specific primers and a common reverse primer. Samples homozygous for allele 1 will be amplified with the short GC-tailed (6 bp) primer and only give a product with lower temperature peak in a melting curve. Samples homozygous for allele 2 will be amplified with the long GC-tailed (14 bp) primer and only give a higher temperature peak. The heterozygous samples will be amplified with both GC-tailed primers, and the melting curves will have both peaks. (B) Method validation and melting curve analysis. The melting curve of the four representative SNPs (Table 1) that were genotyped with the T_m -shift method are shown here. Each SNP has typed samples in duplicate with three different cell line DNA samples, one of each genotype, and a no-template control. T_m , melting temperature; SNP, single nucleotide polymorphism.

Table 2. SNP Genotyping Results Comparison

SNP	T_m -Shift Genotyping				Molecular Inversion Probes (ParAllele BioScience)				Comparison		
	No. of Sample ^a	Calls	No Calls ^b	Call Rate (%)	No. of Sample ^a	Calls	No Calls ^b	Call Rate (%)	Overlapping Calls	Nonmatching Calls	Concordance (%)
SNP1	592	574	18	97.0	600	600	0	100	574	0	100
SNP2	592	589	3	99.5	600	592	8	98.7	581	1	99.8
SNP3	592	592	0	100	600	597	3	99.5	590	0	100
SNP4	592	592	0	100	600	596	4	99.3	588	0	100
SNP5	592	592	0	100	600	598	2	99.7	590	0	100
SNP6	592	592	0	100	600	598	2	99.7	590	0	100
Averages	592	588	3.5	99.4	600	597	3	99.5			
Total									3513	1	99.97

SNP, single nucleotide polymorphism; T_m , melting temperature.
^aThe ParAllele BioScience-generated data set includes 600 samples, but 8 of these had insufficient DNA for subsequent analysis by T_m -shift genotyping.
^b"No Calls" means a lack of call due to assay failure or inferior data quality.

discrimination while maintaining equal amplification efficiency.

Assay Optimization

We observed that some genotyping assays designed with GC tails of 6 and 14 bp displayed poor discrimination in T_m between the two alleles' melting curves, such that one of the peaks in the heterozygote melting curve looked more like a shoulder than an independent second peak. Surprisingly, this problem can be overcome for transition SNPs (AT to GC changes) by attaching the long tail (14 bp) to the allele-specific primer with the higher T_m base (G or C) at the 3' end, and the short tail to the other allele-specific primer with the lower T_m base (A or T). The problem remains only for a few transversion SNPs (approximately 15% of all human SNPs, dbSNP Build 124). Fewer than 3%–5% of all the assays we have tested lack sufficient temperature discrimination for accurate genotype scoring.

A second pattern of melting curves that can make genotype scoring difficult is when one primer amplifies substantially more efficiently than the other, giving rise to melting peaks of very uneven height, despite the presence of GC tails on both primers. To resolve this issue, we have instituted a second round of design/optimization for these assays, in which the allele-specific primer with the higher peak height is added at half of the original

concentration (0.1 μ M). This is not always strictly necessary—because genotypes can be called even under the original conditions—but it often facilitates rapid genotype scoring. In our experience, 20%–25% of all assays can be rescued or improved with this second round of assay optimization. Overall, the main mode of assay failure is near-complete lack of amplification with either one or both allele-specific primers. Some proportion of failed assays may be rescued through further optimization (e.g., lowering of annealing temperatures during PCR). However, the assay success rate and all the assays described here are the product of assay validation using a single, uniform set of conditions, with a second round of optimization for a minority of borderline assays (again under a uniform set of alternate conditions).

Assay Validation

To increase the allele specificity during PCR, we use a truncated version of *Taq* DNA polymerase, the Stoffel fragment (25,29), which has been shown to enhance discrimination of 3' primer-template mismatches (30). To further increase the specificity, we used a modified "Gold" version of Stoffel fragment polymerase (Stoffel Gold) to provide a simplified hot start as previously described (19,27) and to minimize the potential for primer-dimer formation (31). For the Stoffel

polymerase, we have found that adding 40 mM KCl to the mastermix achieves the maximal specificity (and prevents the extension of mismatched primers) without sacrificing amplification efficiency. For most assays, the nonmodified version of the Stoffel fragment polymerase (AmpliTaq[®] DNA Polymerase, Stoffel Fragment; Applied Biosystems) will work equally well, although greater care should be taken in primer design to eliminate primer-primer homology and the potential for primer-dimer formation. In addition, anti-*Taq* antibodies can be used for hot start to achieve results similar to Stoffel Gold.

We have designed more than 306 GC-tailed primer assays to date using the conditions described here. A total of 48 (15.6%) of these assays failed, in almost all cases due to amplification failure of one or both allele-specific primers in the assay. For an additional 19 (6.2%) assays, no variant alleles were observed; however, the number of control samples used for assay validation was small ($n = 22$) and the allele frequencies for these SNPs were, when known, very low. Thus, 239 assays (78% of the total) worked well, but excluding those assays for which no minor allele was observed, the effective success rate was above 83% [239/(48 + 239)]. This is comparable to the results obtained by many other genotyping platforms (12).

SNP Genotyping with Offline Amplification/End-Point Analysis

In order to improve the throughput of our method by leveraging the full capacity of the relatively expensive 384-well real-time thermal cyclers, we tested whether we could amplify multiple plates in dual-block thermal cyclers (GeneAmp 9700s) and then transfer the plates to the real-time thermal cycler (ABI PRISM 7900 HT) for the melting curve analysis. Ninety genomic DNA samples from control cell lines with known genotypes for SNPs rs734351 and rs1004446 (Table 1) were amplified in a 384-well plate on the GeneAmp 9700 with four replicates for each sample. After amplification, the 384-well plate was transferred to an ABI PRISM 7900 HT to analyze the melting curve with onboard software. The melting curve values were exported as a text file and imported to a customized Excel spreadsheet for analysis (Figure 2). All the genotypes of the tested samples generated from the clustering algorithm matched the previous known genotypes for both SNPs.

In the first step of the genotyping analysis procedure, we determine the overall data area of interest corresponding to the outer temperature boundaries of the two peaks of the melting curves to exclude artificial signal arising either from instrument noise or spurious, nonspecific amplification (e.g., primer-dimer). We then determine a center temperature point that unambiguously separates the two melting curve peaks. Using this center temperature, we then perform a normalization such that for each sample we calculate whether the maximum peak height falls into the lower temperature or higher temperature peak area and then normalize all samples falling into either area to the mean temperature maxima for samples in that area. The well-to-well temperature difference for each fluorescence measurement is typically less than 0.5°C. If the peak maximum is shifted more than 1°C after normalization, we flag that sample as an outlier. We have also implemented a quality-control algorithm that averages the first two data points of the raw fluorescence values (at 60°C) for

any sample designated as a no-template control and excludes any sample with a lower initial raw fluorescence value from genotype calling. Samples with a lower initial raw fluorescence reading than the highest no-template control at 60°C are assumed to contain insufficient amounts of PCR product to warrant genotype calling. This prevents samples that fail PCR amplification (e.g., due to failure of adding genomic DNA to the well) from being assigned a genotype based on the melting curve of nonspecific amplification products (e.g., from primer-dimer formation). We next integrate the data points of the melting curves within each of two peak areas. The analysis software defines default peak areas for integration as $\pm 1^\circ\text{C}$ of the average peak maxima for the left and the right peak, respectively. It is sometimes necessary to tune these peak areas to maximize discrimination and improve clustering. The resulting

values for each sample are then graphed in a scatter plot with coordinates corresponding to the integrated peak areas for the first and the second alleles. Then the arctangent value (y/x) of each point to the origin (0, 0) is calculated where x and y were the two peak integration values.

We then use the k -means (32) clustering algorithm to automatically classify samples into four genotype categories based on nearest-centroid sorting. The algorithm employs the arctangent values of each sample as the metrics during the clustering process. Samples that are either predefined as no-template controls or have been filtered by the previously described quality control are automatically assigned an “NA” label. Then data are classified into a predetermined number of clusters (i.e., four clusters: three genotype clusters and one NA cluster, or three clusters, if

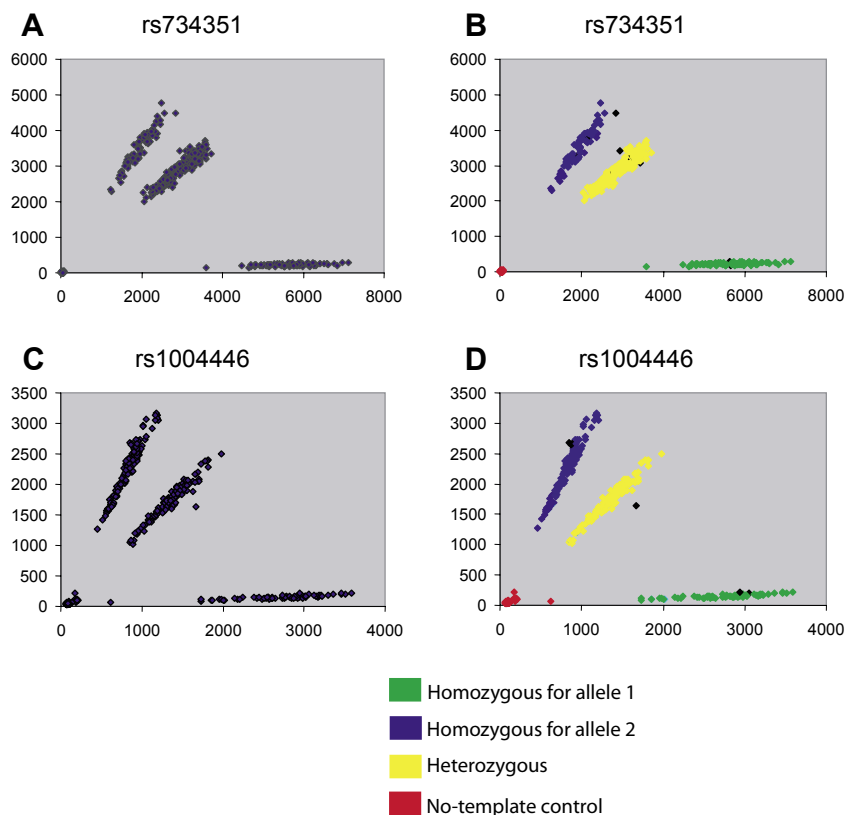


Figure 2. Genotyping results for two representative SNPs. Data for 368 genotypes are shown (90 genomic DNA samples and 2 no-template controls with 4 replicates each). The melting curve data were analyzed using the k -means algorithm in a Microsoft Excel spreadsheet. The x-axis is the integration value of peak 1 areas, and the y-axis is the integration value of peak 2 areas. (A and C) The scatter plot of the integrated data points from two peak areas for each SNP. (B and D) Output of the k -means clustering. Colored dots correspond to the indicated genotypes; black dots are samples with low conditional probability of belonging to a particular cluster. SNP, single nucleotide polymorphism.

the variant homozygote is not present in a particular run), and samples are assigned to the cluster with the smallest distance between the sample data and the center of the cluster (centroid). The genotypes are then assigned automatically to samples in each cluster.

To further increase the throughput of the genotype calling process, we have developed a standalone application (GCSNPclust; software and accompanying user manual are available online at www.BioTechniques.com as Supplementary Material) written in Java using the same *k*-means algorithm. This software is capable of simultaneously analyzing melting curve results from multiple 384-well plates for a given SNP and automatically assigning genotypes such that up to 1536 genotypes (four 384-well plates) can be called in a matter of minutes. The software performs the above-mentioned

calculations based on the user input parameters described as shown in Figure 3. The software flags genotypes that appear to be outliers. First, it flags samples (dots encircled in magenta) that are shifted more than 1°C (default) during the temperature normalization. Genotypes are still assigned to these samples, and the default-allowable temperature shift can be increased for assays that have greater variability in peak maxima temperatures. Second, the software flags samples with a low conditional probability of belonging to a particular cluster. These samples are colored according to the cluster to which they are assigned, but displayed with a white outer circle. Flagged genotypes are easily visualized in a filtered display option, and these genotype calls can then be manually changed or excluded after inspection of the melting curves.

To test the accuracy of our method, we genotyped 592 human genomic DNA samples on 6 SNPs that had previously been genotyped using a different method (as part of a highly multiplexed molecular inversion probe genotyping assay) by ParAllele BioScience (San Francisco, CA, USA) (13). Genotypes (3513) were called by both methods, and only one discordant genotype call was observed for a concordance rate of 99.97%. The call rate for T_m -shift genotyping was >98%. This indicates that the accuracy of T_m -shift genotyping is very high (Table 2).

DISCUSSION

General criteria for selecting SNP genotyping methods include accuracy/sensitivity, success rate, flexibility, cost, and throughput. Here we describe an advanced single-tube SNP genotyping technology using the T_m -shift mechanism that offers an easy and inexpensive method to determine SNP genotypes.

T_m -shift genotyping as described here is a very inexpensive system. In a high-throughput mode, it only requires a liquid-handling robot and several standard thermal cyclers, as well as a 384-well real-time thermal cycler. These instruments, including the real-time thermal cycler, are used for many applications and are becoming standard equipment in many laboratories and core facilities. T_m -shift genotyping assays require no fluorescently labeled or otherwise modified oligonucleotide primers because the real-time fluorescence signal is generated by a nonspecific DNA binding dye, SYBR Green I. We used standard, non-high-performance liquid chromatography (HPLC)-purified, desalted primers, ordered lyophilized in a 96-well plate format (Integrated DNA Technologies, Coralville, IA, USA), so the cost per new assay is less than US \$30. The reagent cost per genotype for a 15- μ L reaction is about 30 cents, primarily due to the cost of polymerase, but this is in the range of standard PCR costs. Obviously, the cost per genotype is substantially lower for some highly multiplexed genotyping systems; however, such methods generally

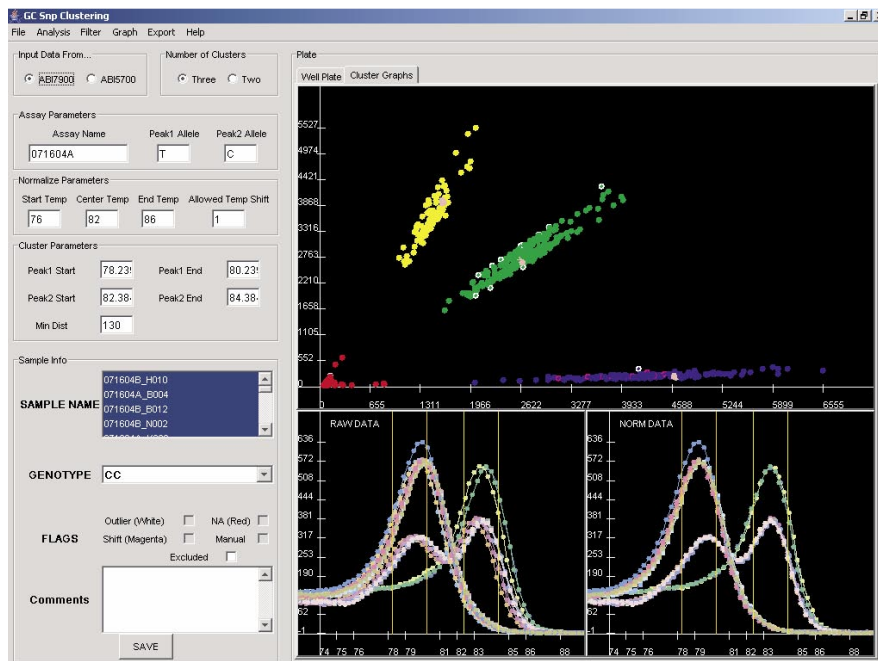


Figure 3. T_m -shift genotyping with standalone software (GCSNPclust). After checking the melting curves of several samples on plates, we determined the outer boundary of the two peak areas by defining the start temperature of the first peak and the end temperature of the second peak, and set the center temperature separating the two peak areas. The software processed the *k*-means clustering algorithm and automatically assigned the genotypes to each sample. Top panel, clustering results of the SNP genotyping data of 600 genomic DNA samples and controls from two 384-well plate data files are shown. The x-axis is the integration value of peak 1 areas, and the y-axis is the integration value of peak 2 areas. Red dots, no-template controls or failed samples; blue dots, homozygotes of allele 1; yellow dots, homozygotes for allele 2; green dots, heterozygotes; white-encircled dots, outlier samples with low conditional probability of belonging to that particular cluster; magenta-encircled dots, sample with the T_m shifted more than the “allowed temperature shift”; pink dots, samples selected for inspection. Bottom-left panel, raw melting curves of selected (i.e., pink-highlighted) samples in top panel. Bottom-right panel, melting curves of the same selected samples after normalization. The y-axis is derivative fluorescence change, and the x-axis is the T_m (°C) in both bottom panels. Curve colors are matched between the two lower panels but are otherwise arbitrary. T_m , melting temperature; SNP, single nucleotide polymorphism.

require an investment in expensive, dedicated instrumentation that is only cost-efficient when it can amortize over millions of genotypes. To further reduce the costs of the T_m -shift genotyping, preliminary data (data not shown) indicate that many SNP genotyping assays may be performed in a total volume of 10 or 5 μ L.

The T_m -shift genotyping assay is also highly flexible. To date, we have validated over 300 SNP assays for 230 different SNPs under uniform reaction conditions with a success rate of >83%. (For some SNPs, we tested assays to both strands to increase the chance of achieving at least one working assay for a given SNP.) For assay validation, 16 new SNP assays can be tested on 24 samples (or 22 samples with 2 no-template controls) in one 384-well plate. We recently demonstrated the flexibility of the T_m -shift method by genotyping 9 SNPs across 596 samples in only 5 full working days from the receipt of the SNP sequences to the final genotyping report. This included primer design, assay validation, genotyping, and genotype scoring, and resulted in an overall 98.5% call rate and 5364 completed genotypes.

Another advantage of the T_m -shift method is that, as with other homogeneous detection methods, all reactions and measurements are performed in a single closed tube in T_m -shift assays. This eliminates complicated post-PCR manipulation and significantly reduces the risk of PCR contamination and the need for separate pre- and post-PCR laboratories, which is recommended for many highly multiplexed genotyping systems. Because both alleles are detected in a single reaction, the T_m -shift assay also encompasses an internal amplification control and eliminates the risk of false-negative results that are associated with some other genotyping methods.

The throughput of T_m -shift genotyping makes it suitable for small to medium-sized projects, such as those focusing on following up on genome-wide scans or on candidate genes, SNPs, or linkage regions. A key advantage of the T_m -shift method is the use of melting curves to examine PCR products as an end-point analysis. Because it is not necessary to monitor

the fluorescence signals in real time, the offline amplification enables one to better leverage the capacity of the most-expensive instrument (the real-time thermal cycler). With a relatively simple laboratory setup, including one Biomek 2000, two GeneAmp 9700 (dual 384-well) thermal cyclers, and one ABI PRISM 7900 HT real-time thermal cycler, one technician in our laboratory routinely processes 8–10 plates for roughly 3500 genotypes per day. This throughput could be improved with a faster liquid-handling robot bearing a 96-probe tool (e.g., the Biomek Fx) and additional GeneAmp 9700s. In addition, we have tested the stability of the 384-well reaction plates in the interval of post-PCR and premelting curve analysis. The reaction plates were very stable at room temperature for up to 2 days and gave consistent genotype calling, demonstrating that it is feasible to have the amplified reaction plates be automatically loaded onto the ABI PRISM 7900 HT via a plate stacker. This should decrease the necessary hands-on time required for plate processing and enable a throughput of 6000–10,000 genotypes per day. The throughput could be further scaled up by the use of additional instruments (thermal cyclers and liquid-handling robots).

The T_m -shift genotyping method enables the high-throughput, easy, accurate, and inexpensive typing of tens to hundreds of SNPs across hundreds to thousands of DNA samples.

ACKNOWLEDGMENTS

We thank David Birch for supplying the Stoffel Gold polymerase; Huyen Phan and Christina Kerksteck for providing control DNA samples; Brian Rhees, Mitchell Martin, Jia Li, and Henry Erlich for support and encouragement.

COMPETING INTERESTS STATEMENT

All authors except S.P. declare that they are or were employed by Roche Molecular Systems, a supplier of a product mentioned in this article. S.P. declares no competing interests.

REFERENCES

1. Mackay, T.F. 2001. The genetic architecture of quantitative traits. *Annu. Rev. Genet.* 35:303-339.
2. Botstein, D. and N. Risch. 2003. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat. Genet.* 33(Suppl):228-237.
3. Carlson, C.S., M.A. Eberle, L. Kruglyak, and D.A. Nickerson. 2004. Mapping complex disease loci in whole-genome association studies. *Nature* 429:446-452.
4. Kwok, P.Y. 2001. Methods for genotyping single nucleotide polymorphisms. *Annu. Rev. Genomics Hum. Genet.* 2:235-258.
5. Twyman, R.M. and S.B. Primrose. 2003. Techniques patents for SNP genotyping. *Pharmacogenomics* 4:67-79.
6. Chen, X. and P.F. Sullivan. 2003. Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput. *Pharmacogenomics J.* 3:77-96.
7. Fu, D.J., K. Tang, A. Braun, D. Reuter, B. Darnhofer-Demar, D.P. Little, M.J. O'Donnell, C.R. Cantor, and H. Koster. 1998. Sequencing exons 5 to 8 of the p53 gene by MALDI-TOF mass spectrometry. *Nat. Biotechnol.* 16:381-384.
8. Buetow, K.H., M. Edmonson, R. MacDonald, R. Clifford, P. Yip, J. Kelley, D.P. Little, R. Strausberg, et al. 2001. High-throughput development and characterization of a genome-wide collection of gene-based single nucleotide polymorphism markers by chip-based matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Proc. Natl. Acad. Sci. USA* 98:581-584.
9. Mohlke, K.L., M.R. Erdos, L.J. Scott, T.E. Fingerlin, A.U. Jackson, K. Silander, P. Hollstein, M. Boehnke, and F.S. Collins. 2002. High-throughput screening for evidence of association by using mass spectrometry genotyping on DNA pools. *Proc. Natl. Acad. Sci. USA* 99:16928-16933.
10. Hacia, J.G., J.B. Fan, O. Ryder, L. Jin, K. Edgemon, G. Ghandour, R.A. Mayer, B. Sun, et al. 1999. Determination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide arrays. *Nat. Genet.* 22:164-167.
11. Fan, J.B., X. Chen, M.K. Halushka, A. Berno, X. Huang, T. Ryder, R.J. Lipshutz, D.J. Lockhart, and A. Chakravarti. 2000. Parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays. *Genome Res.* 10:853-860.
12. Hardenbol, P., J. Baner, M. Jain, M. Nilsson, E.A. Namsaraev, G.A. Karlin-Neumann, H. Fakhrai-Rad, M. Ronaghi, et al. 2003. Multiplexed genotyping with sequence-tagged molecular inversion probes. *Nat. Biotechnol.* 21:673-678.
13. Ferguson, J.A., F.J. Steemers, and D.R. Walt. 2000. High-density fiber-optic DNA random microsphere array. *Anal. Chem.* 72:5618-5624.
14. Medintz, I., W.W. Wong, L. Berti, L. Shiw, J. Tom, J. Scherer, G. Sensabaugh, and R.A. Mathies. 2001. High-performance multiplex SNP analysis of three hemochroma-

- tosis-related mutations with capillary array electrophoresis microplates. *Genome Res.* 11:413-421.
15. **De Angelis, D.A.** 1999. Why FRET over genomics? *Physiol. Genomics* 1:93-99.
 16. **Livak, K.J., J. Marmaro, and J.A. Todd.** 1995. Towards fully automated genome-wide polymorphism screening. *Nat. Genet.* 9:341-342.
 17. **Tyagi, S., D.P. Bratu, and F.R. Kramer.** 1998. Multicolor molecular beacons for allele discrimination. *Nat. Biotechnol.* 16:49-53.
 18. **Piatek, A.S., S. Tyagi, A.C. Pol, A. Telenti, L.P. Miller, F.R. Kramer, and D. Alland.** 1998. Molecular beacon sequence analysis for detecting drug resistance in *Mycobacterium tuberculosis*. *Nat. Biotechnol.* 16:359-363.
 19. **Germer, S., M.J. Holland, and R. Higuchi.** 2000. High-throughput SNP allele-frequency determination in pooled DNA samples by kinetic PCR. *Genome Res.* 10:258-266.
 20. **Newton, C.R., A. Graham, L.E. Heptinstall, S.J. Powell, C. Summers, N. Kalsheker, J.C. Smith, and A.F. Markham.** 1989. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res.* 17:2503-2516.
 21. **Wu, D.Y., L. Ugozzoli, B.K. Pal, and R.B. Wallace.** 1989. Allele-specific enzymatic amplification of beta-globin genomic DNA for diagnosis of sickle cell anemia. *Proc. Natl. Acad. Sci. USA* 86:2757-2760.
 22. **Matsuzaki, H., S. Dong, H. Loi, X. Di, G. Liu, E. Hubbell, J. Law, T. Berntsen, et al.** 2004. Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nat. Methods* 1:109-111.
 23. **Hardenbol, P., F. Yu, J. Belmont, J. Mackenzie, C. Bruckner, T. Brundage, A. Boudreau, S. Chow, et al.** 2005. Highly multiplexed molecular inversion probe genotyping: over 10,000 targeted SNPs genotyped in a single tube assay. *Genome Res.* 15:269-275.
 24. **Murray, S.S., A. Oliphant, R. Shen, C. McBride, R.J. Steeke, S.G. Shannon, T. Rubano, B.G. Kermani, et al.** 2004. A highly informative SNP linkage panel for human genetic studies. *Nat. Methods* 1:113-117.
 25. **Germer, S. and R. Higuchi.** 1999. Single-tube genotyping without oligonucleotide probes. *Genome Res.* 9:72-78.
 26. **Ardlie, K.G., K.L. Lunetta, and M. Seielstad.** 2002. Testing for population subdivision and association in four case-control studies. *Am. J. Hum. Genet.* 71:304-311.
 27. **Birch, D.E.** 1996. Simplified hot start PCR. *Nature* 381:445-446.
 28. **Germer, S. and R. Higuchi.** 2002. Homogeneous allele-specific PCR in SNP genotyping. *In* P.-Y. Kwok (Ed.), *Methods in Molecular Biology*, vol. 212: Single Nucleotide Polymorphisms: Methods and Protocols. Humana Press, Totowa, NJ.
 29. **Lawyer, F.C., S. Stoffel, R.K. Saiki, S.Y. Chang, P.A. Landre, R.D. Abramson, and D.H. Gelfand.** 1993. High-level expression, purification, and enzymatic characterization of full-length *Thermus aquaticus* DNA polymerase and a truncated form deficient in 5' to 3' exonuclease activity. *PCR Methods Appl.* 2:275-287.
 30. **Tada, M., M. Omata, S. Kawai, H. Saisho, M. Ohto, R.K. Saiki, and J.J. Sninsky.** 1993. Detection of ras gene mutations in pancreatic juice and peripheral blood of patients with pancreatic adenocarcinoma. *Cancer Res.* 53:2472-2474.
 31. **Chou, Q., M. Russell, D.E. Birch, J. Raymond, and W. Bloch.** 1992. Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplifications. *Nucleic Acids Res.* 20:1717-1723.
 32. **Sharma, S.** 1996. Cluster analysis, p. 185-236. *In* *Applied Multivariate Techniques*. John Wiley & Sons, New York.

Received 4 May 2005; accepted 4 August 2005.

Address correspondence to:

Jun Wang
 Human Genetics Department
 Roche Molecular Systems
 1145 Atlantic Avenue
 Alameda, CA 94501, USA
 e-mail: jun.wang@roche.com

To purchase reprints
 of this article, contact

Reprints@BioTechniques.com