

# DNA Analysis by Fluorescence Quenching Detection

Ming Xiao<sup>1</sup> and Pui-Yan Kwok<sup>1,2,3</sup>

<sup>1</sup>Cardiovascular Research Institute and the <sup>2</sup>Department of Dermatology, University of California, San Francisco, California 94143, USA

The analysis of human genetic variations such as single nucleotide polymorphisms (SNPs) has great applications in genome-wide association studies of complex genetic traits. We have developed an SNP genotyping method based on the primer extension assay with fluorescence quenching as the detection. The template-directed dye-terminator incorporation with fluorescence quenching detection (FQ-TDI) assay is based on the observation that the intensity of fluorescent dye R110- and R6G-labeled acycloterminators is universally quenched once they are incorporated onto a DNA oligonucleotide primer. By comparing the rate of fluorescence quenching of the two allelic dyes in real time, we have extended this method for allele frequency estimation of SNPs in pooled DNA samples. The kinetic FQ-TDI assay is highly accurate and reproducible both in genotyping and in allele frequency estimation. Allele frequencies estimated by the kinetic FQ-TDI assay correlated well with known allele frequencies, with an  $r^2$  value of 0.993. Applying this strategy to large-scale studies will greatly reduce the time and cost for genotyping hundreds and thousands of SNP markers between affected and control populations.

Single nucleotide polymorphisms (SNPs) are the most abundant sequence variation found in nature. When two human genomes are compared, single base-pair variations are found at approximately 1200-nucleotide intervals. Because of their abundance and low mutation rate, SNPs are the markers of choice in association studies to identify the genetic risk factors in common diseases (Risch and Merikangas 1996; Kruglyak 1999). As a result of several large initiatives, several million single base-pair variations have been deposited in public and commercial databases. Although many robust genotyping methods have been developed during the past decade, a major challenge remains regarding the cost and time needed to obtain the genotypes of numerous samples with hundreds of thousands of SNPs.

To reduce the time and cost associated with genotyping every individual in a study, it has been proposed that investigators work with pooled DNA samples constructed by mixing equal amounts of DNA from groups of individuals. The rationale is that, by definition, most of the SNPs in the genome are not associated with a disease phenotype, and so a cost-effective screening method that can identify significant divergence in allele frequency between case and control populations will not only shorten the time it takes to conduct an association study but also reduce the cost of the study substantially. A number of methods have been developed to estimate allele frequencies in pooled samples (Kwok et al. 1994; Breen et al. 2000; Germer et al. 2000; Giordano et al. 2001; Wolford et al. 2000; Zhou et al. 2001; Gruber et al. 2002; Werner et al. 2002). Because analysis of pooled samples requires only a handful of reactions (for pools of cases and controls plus references) per SNP, low assay-development cost is of paramount importance.

We have developed a genotyping assay based on allele-specific primer extension, termed template-directed dye-

terminator incorporation (TDI) assay with fluorescence quenching (FQ) detection. The FQ-TDI assay takes advantage of the fact that the fluorescence intensity of certain dyes is significantly quenched when they are attached to oligonucleotides. By monitoring changes in fluorescence intensity of dye-terminators in an allele-specific primer extension reaction, one can determine the genotype of a DNA sample. If one monitors the fluorescence intensity change in real time, the relative abundance of the alleles in a pooled sample can be determined accurately. We report here that the simple, extremely low-cost FQ-TDI assay yielded highly accurate genotypes (220 SNPs) and allele frequency estimates (six SNPs).

## RESULTS

### SNP Genotyping by the FQ-TDI Assay

The fluorescence of some conjugated fluorescent dyes is sensitive to the local environment of the oligonucleotides. The fluorescence quenching due to dye-nucleotide interactions has been reported for many different fluorescent dyes (Torimura et al. 2001; Nazarenko et al. 2002, and references within). A genotyping assay based on TDI with fluorescence polarization detection developed by our group (Chen et al. 1999; Hsu et al. 2001) was recently commercialized as AcycloPrime-FP by Perkin-Elmer. The AcycloPrime-FP SNP system utilizes dye-labeled acyclic nucleoside triphosphates (acycloterminators) and the AcycloPol DNA polymerase, which preferentially incorporates Acycloterminators over dideoxynucleotides (Perkin-Elmer Web site, <http://lifesciences.perkinelmer.com/products/snp.asp>). In the process of evaluating and using this AcycloPrime-FP genotyping assay, we noticed that the fluorescence of R110-labeled acycloterminators was heavily and universally quenched, once they were incorporated into the primers. Intrigued by this observation, we systematically investigated the quenching effects of oligonucleotides on four dye-labeled acycloterminators (R110, R6G, Tamra, and TexasRed), which have well separated excitation and emission spectra and could be

<sup>3</sup>Corresponding author.

E-MAIL [kwok@cvrmail.ucsf.edu](mailto:kwok@cvrmail.ucsf.edu); FAX (415) 476-2283.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.987803>.

used together in primer extension assays. Figure 1 shows the quenching patterns of these four dye-labeled acycloterminators for hundreds of randomly selected SNP primers. The quenching pattern of R110-acycloterminator indicates that 97% of the 667 SNP primers tested quenched the R110 fluorescence by at least 20%, with 54% of the SNP primers quenching R110 fluorescence by more than 50%.

The quenching pattern of R6G-acycloterminator is shifted somewhat lower than that of the R110-acycloterminator, with 88% of the 258 SNP primers tested quenching the R6G fluorescence by more than 20%. Tamra- and Texas Red-acycloterminators were quenched even less, with only 65% of the 667 SNP primer tested quenching the Tamra-acycloterminator by at least 20% and none of the 50 SNP primers tested quenching the Texas Red-acycloterminator more than 20%.

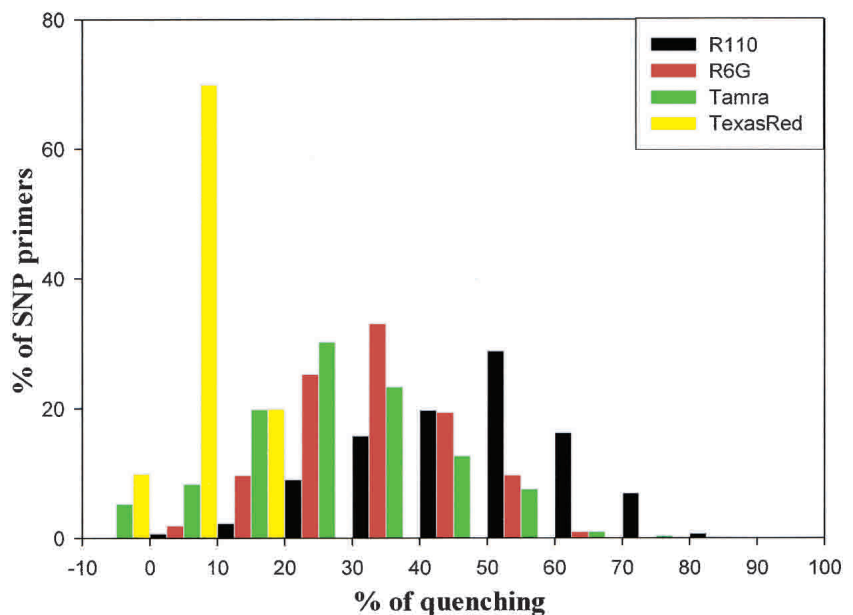
Considering the fact that over 88% of SNP primers show significant quenching effects (>20%) on R110- and R6G-acycloterminators, one can monitor the changes in fluorescence intensities during the primer extension step to determine which terminator(s) get incorporated and infer the SNP genotype. This can be done in real time by a fluorescence spectrophotometer connected to a thermal cycler or at its end-point by means of a fluorescence plate reader.

Figure 2 shows the real-time fluorescence intensity profiles of four representative samples tested for SNP marker rs154162 during linear amplification in the primer extension step of the TDI assay. The fluorescence readings correspond to the emission maxima for R110-G (525nm) and R6G-A (535 nm) acycloterminators normalized by multicomponent analysis. In Figure 2A, the G/G homozygous sample incorporates R110-G (but not R6G-A) and shows a progressive drop in R110 fluorescence (filled circles) but no change in R6G fluorescence (open circles) as thermal cycling proceeds. The homozygous A/A sample in Figure 2B incorporates R6G-A (but not R110-G) and shows a drop in R6G fluorescence but no change in R110 fluorescence. The heterozygous G/A sample in Figure 2C incorporates both R110-G and R6G-A and shows a drop in both R110 and R6G fluorescence. In contrast, the fluorescence intensity profile of a negative control sample (Fig. 2D) shows no change for both dyes, because neither is incorporated. The rate of change of the R110 and R6G fluorescence intensities is a reflection of the amounts of present alleles, with the initial slope of change for the heterozygote approximately half of that for the homozygote. By inspecting the changes in fluorescence intensity for R110 and R6G, one can assign the allelic status of each test sample with high confidence. We obtained 1920 genotypes across 258 SNPs with a conversion rate of 85% (220 SNPs were successfully genotyped, and >90% of the samples yielded high confidence genotype calls). Among the successfully genotyped SNPs, the average high confidence call rate is 93%, and the concordance between genotypes called by the FQ-TDI and FP-TDI assays was 99.8%.

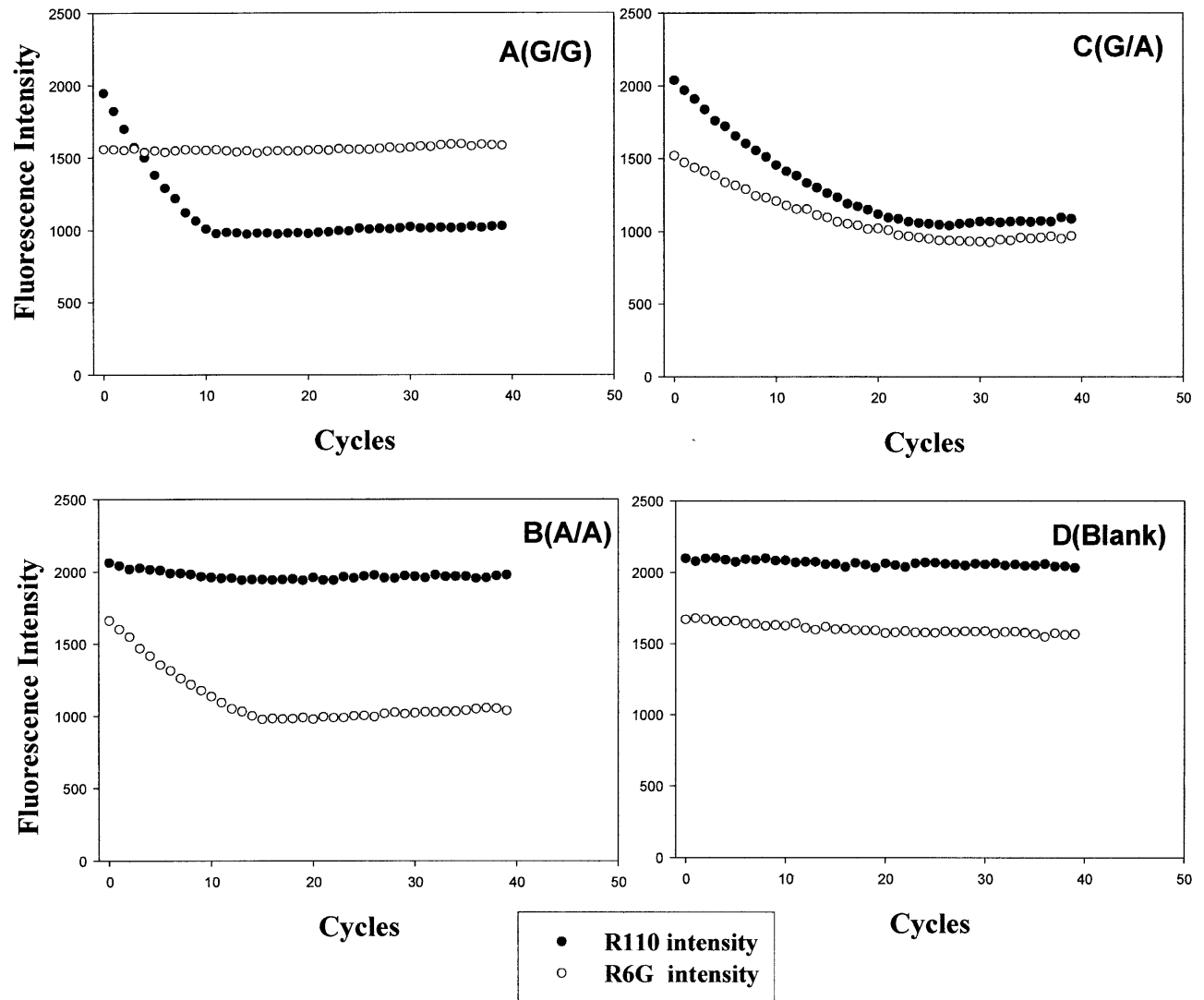
## Allele Frequency Determination in Pooled DNA Samples

The FQ-TDI assay can be extended to determine the allele frequency of pooled DNA samples. Because fixed amounts of dye terminators are used in the primer extension reaction, the larger the number of target DNA molecules containing a particular allele that is present, the faster the dye terminators are used up. The initial rate of change in fluorescence intensity (the steepness of the slope) of a dye-terminator is therefore a reflection of the amount of DNA molecule containing the allele corresponding to the dye-terminator incorporated. Assuming that the DNA polymerase incorporates the two dye-terminators with equal efficiency, the rate of change in fluorescence intensity of both dyes will be the same when the allele frequency is 50% or when the sample is from a heterozygote. To measure the allele frequencies of an SNP in a pooled DNA sample, the incorporation rates of the two dye-terminators are compared kinetically, cycle by cycle. These incorporation rates are then used in determining the relative amounts of each allele present in the pooled samples, using a heterozygote as the reference.

Figure 3 illustrates a typical reaction performed to determine allele frequency for pooled DNA samples. The top panel is a normalized real-time quenching curve for a heterozygous sample of rs922365 (see Methods section for normalization procedure). The linear regressions of the first eight cycles are shown in the inset, and the ratio of the two slopes is calculated as 0.93 (R6G/R110), which indicates that the AcycloPol enzyme incorporates R110-G slightly faster than R6G-A. The allele frequencies are obtained by comparing the ratios of the slopes for the pooled sample in the bottom panel of Figure 3



**Figure 1** The quenching patterns for four dye-labeled acycloterminators, R110, R6G, Tamra, and Texas Red. The quenching pattern for R110-acycloterminators (black bars) was derived from results of over 667 SNP primers. Red, green, and yellow bars represent the quenching patterns of R6G (258 SNP primers), Tamra (667 SNP primers), and Texas Red (50 SNP primers) acycloterminators, respectively. The x-axis is the percentage of quenching, which are grouped in 10% intervals from -10%–0% to 90%–100%. The y-axis is the percentage of SNP primers in each of the quenching fractions.



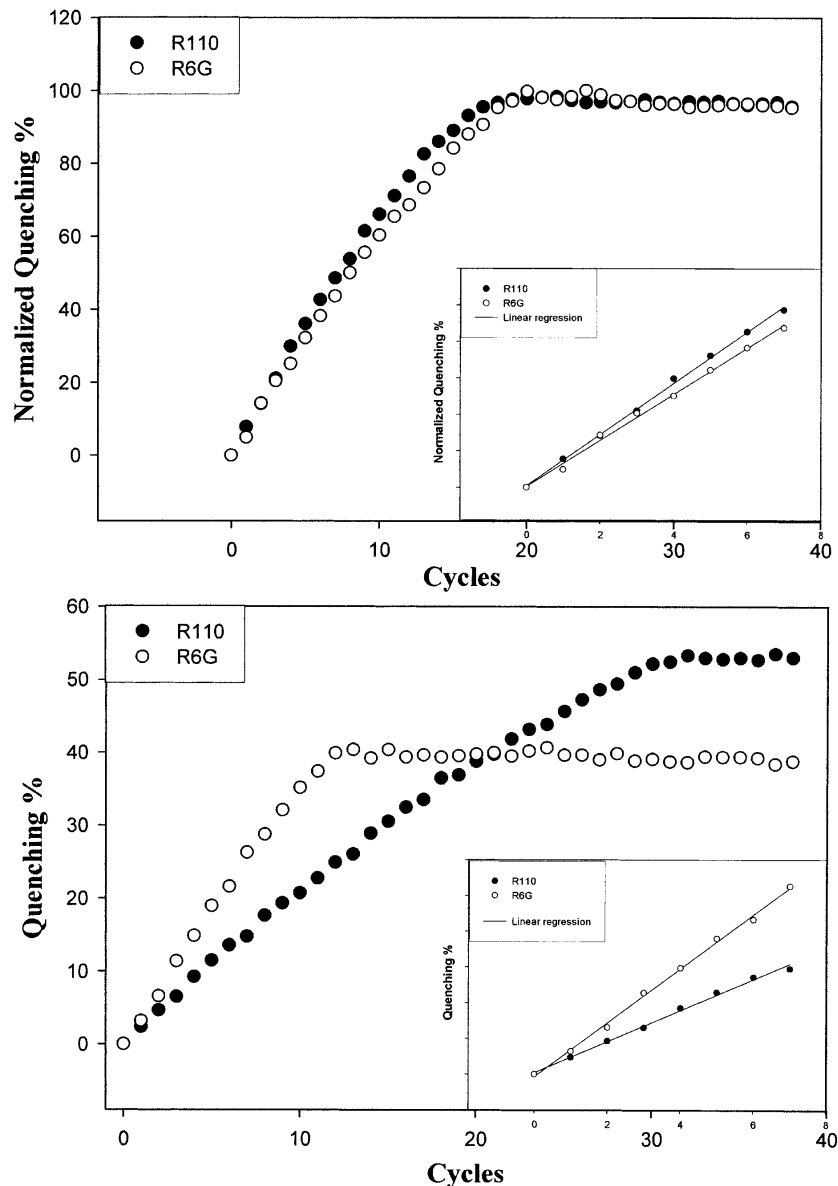
**Figure 2** The real-time fluorescence intensity profiles of four representative samples tested for SNP marker rs154162 during thermal cycling of the primer extension step of the TDI assay. The fluorescence readings were at the emission maxima for R110-G and R6G-A acycloterminators using multicomponent analysis to subtract contributions of fluorescence from the other dye at those wavelengths based on the pure spectra of R110 and R6G. (●) R110 fluorescence; (○) R6G fluorescence. The intensity profiles of (A) a G/G homozygous sample, (B) an A/A homozygous sample, (C) a G/A heterozygous sample, and (D) negative control.

and the ratio of the slopes for the heterozygous sample (see Methods).

To demonstrate the validity of the allele frequency estimation approach, we performed two sets of experiments. First, we constructed DNA mixtures containing varying amounts of the two alleles for the SNP marker rs922365 using DNA from two individuals homozygous for the two alleles. Mixtures with allele frequencies in 5% steps were constructed from 5%–95%. The calculated allele frequencies were strongly correlated with predetermined allele frequencies (Table 1). The calculations were based on the assumption that the relative terminator incorporation efficiency was 0.93 and the total quenching ratio was 0.4/0.55 for the two acycloterminators. The data show that this method can clearly distinguish between pools with allele frequencies that differ by less than 5% just by visual inspection. When plotted against the known allele frequencies of the mixtures, the calculated initial slope

ratios fit well with a hyperbolic curve represented by the equation  $y = 1.02/(1 + 0.66x)$ ; Fig. 4). This curve agrees well with the predicted equation of  $y = 1.0/(1 + 0.67x)$  used for allele frequency estimation. The results of this experiment confirm that the allele frequencies can be estimated accurately, with only slightly higher standard error at the ends of the allele frequency spectrum (the largest standard deviation is 2.8% for the 95% allele frequency mixture).

The second set of experiments was done to compare the estimated allele frequencies in 21 SNP-pooled sample combinations obtained by the kinetic FQ-TDI assay against the allele frequencies obtained by genotyping the individuals in these pools using the FP-TDI assay (Chen et al. 1999; Xiao et al., 2003). In all, five SNPs were genotyped across three pools consisting of 32 individuals each. In the allele frequency estimation experiments, four replicates were performed. SNP primers from both strands were used in two of the SNPs, yield-



**Figure 3** The basis of determining allele frequency for pooled DNA samples. The *top* panel is a normalized real-time quenching curve for a heterozygous sample of rs922365. The linear regression of the first eight cycles is shown in the *inset*. The *bottom* panel is the quenching curve for a mixture containing 75% A allele.

ing independent results in terms of allele frequency estimates. The results, summarized in Table 2, show that the estimated allele frequencies agree well with the known allele frequencies, with an average difference of 1.8%. Although there are occasional discrepancies between the pool estimates and real allele frequencies at the 3%–5% level, the relative allele frequency estimates between the pools are quite consistent. For example, the measured allele frequencies of marker rs3354 are 4%, 5%, and 3% lower than known allele frequencies for pool 1, pool 2, and pool 3, respectively, but the relative difference in allele frequency between the pools (pool 1 being 2% lower than pool 2 and 6% higher than pool 3) is preserved in the estimates (pool 1 being 1% lower than pool 2 and 5% higher than pool 3). Figure 5 is a scatterplot summarizing all of the

data in Tables 1 and 2. The measured allele frequency in pooled samples is highly correlated with known allele frequencies ( $r^2 = 0.993$ ), with a slope of 1.003 for the linear regression line.

## DISCUSSION

Fluorescence quenching is a well known phenomenon and is part of the detection method in several SNP genotyping assays, including molecular beacons, 5'-nuclease (Taqman) assay, and the Invader assay. Instead of using a special quencher moiety, our approach takes advantage of the inherent quenching properties of DNA. DNA does not quench all dyes effectively, however, and some DNA sequences do not quench very well. Of the fluorescent dyes tested, we have found that R110 and R6G, two commonly used dyes in nucleic acid labeling, are strongly quenched by DNA. We have also observed that >70% of the quenching is due to the primary structure of DNA and that effective quenching is always seen when there is at least one guanosine within 10 bases of the dye-terminator after it is incorporated (data not shown). Furthermore, the strongest quenching is observed when two consecutive guanines are found immediately adjacent to the incorporated dye-terminator. Our results confirm the important role guanosine plays as a quencher of fluorescence. As the base with the highest electron donating property, guanosine promotes the formation of charge transfer complexes between the fluorophores and nucleosides (Seidel et al. 1996; Steenken and Jovanovic 1997). Of the 1.7 million SNPs found in public databases for which we have designed TDI assays (Vieux et al. 2002), over 95% of the assays utilize at least one SNP primer containing a guanosine within 10 bases of the polymorphic site and are therefore suitable for fluorescence quenching detection. Based on our extensive experience with developing TDI assays from SNPs randomly chosen from dbSNP (the database of single nucleotide polymorphisms

at the National Center for Biotechnology Information), assays cannot be designed for 31% of the SNPs due to repetitive elements located at or near the SNPs. Coupled with the 5% of SNPs where there are no suitable primers that meet the quenching requirements, and an 85% conversion rate, the FQ-TDI assay will work for about 56% of the SNPs in dbSNP.

A feature of the FQ-TDI assay is that the dye-terminators are used up during the course of the reaction and quenching is complete when this occurs. Therefore, the initial incorporation rate of dye-terminators is the best indicator of the amount of DNA template present in the pooled sample. We only use the fluorescence readings during the first eight cycles of the primer extension reaction in calculating the initial slope of fluorescence intensity change because, under the

**Table 1.** Allele Frequency Estimates From Constructed Mixtures with Allele Frequency Ranging from 5% to 95%

Known allele frequency	$S_2/S_1$	S.D.	Estimated allele frequency	S.D.
0.05	34.33	2.47	0.09	0.022
0.1	13.25	4.82	0.13	0.023
0.15	8.63	1.43	0.16	0.025
0.2	7.12	0.95	0.19	0.014
0.25	4.36	0.20	0.27	0.009
0.3	3.54	0.13	0.31	0.008
0.35	3.01	0.17	0.35	0.012
0.4	2.64	0.20	0.38	0.018
0.45	1.94	0.11	0.45	0.014
0.5	1.61	0.042	0.50	0.007
0.55	1.34	0.091	0.55	0.017
0.6	1.10	0.064	0.60	0.015
0.65	0.92	0.058	0.64	0.016
0.7	0.70	0.060	0.70	0.019
0.75	0.56	0.043	0.75	0.016
0.8	0.42	0.030	0.80	0.013
0.85	0.28	0.041	0.86	0.018
0.9	0.23	0.012	0.89	0.006
0.95	0.098	0.046	0.96	0.028

$S_2/S_1$ , slope ratio; S.D., standard deviation.

standard conditions used, dye-terminator incorporation during the first eight cycles is linear for all the assays studied. Moreover, mis-incorporation of dye-terminators is minimal during the first eight cycles.

The data from the mixing experiments show that the FQ-TDI assay works well over the entire allele frequency spectrum, although slightly larger variations of the allele frequency are seen at both ends of the spectrum. This is largely due to the fact that we used equal amounts of R110- and R6G-acycloterminators for the primer extension reaction. For the allele with >90% allele frequency, the limited terminators will be used up too quickly to be sensitive to small allele frequency differences. The inaccuracy of estimating the slope of a flat line also contributes to the problem for the allele at <10% allele frequency. Instead of using equal amounts of two terminators, a 2:1 ratio of terminators (allele with high allele frequency vs. allele with low allele frequency) can be used to change the slopes and therefore improve the sensitivity for pools with extreme allele frequencies (data not shown).

In summary, we have developed an allele frequency estimation strategy that is both accurate and cost-effective. Without any dye-labeled probes used in the assay, the development cost is minimal. Furthermore, designs for the kinetic FQ-TDI assay using universal assay conditions have been made for >1.6 million publicly available SNPs. This low-cost assay is as accurate as any other currently available allele frequency estimation method. It is a viable approach to screen populations for small allele frequency differences in genome-wide case-control studies, thereby

reducing the time and cost associated with these large-scale studies.

## METHODS

### DNA Samples and Pool Construction

DNA samples of 96 anonymous individuals were obtained from the Coriell Institute for Medical Research. The pooled DNA samples included 32 individuals each from the African-American, Asian-American, and European-American panels. The concentration of individual DNA samples was determined by using both the absorbance at 260 nm and a DNA-specific fluorescence dye, PicoGreen (Molecular Probes).

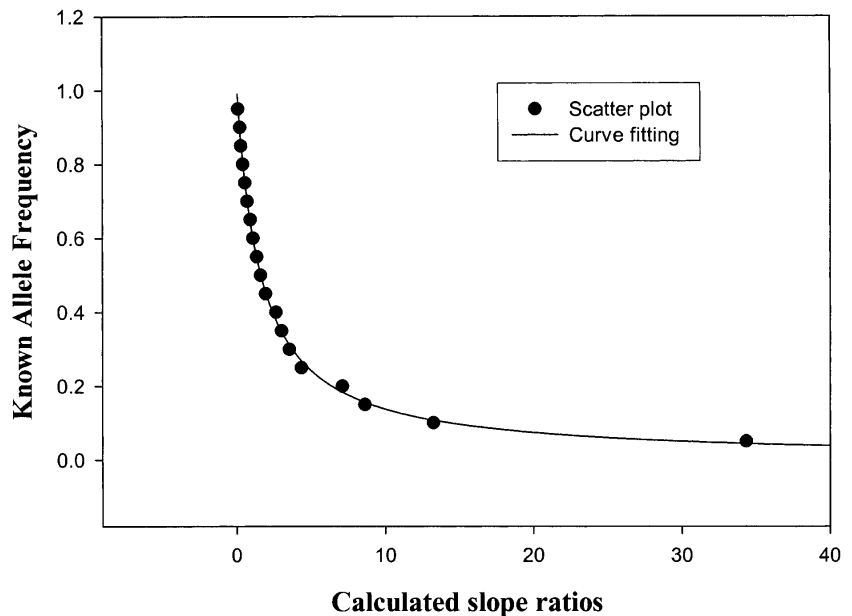
The population pool samples were constructed by adding equal amounts (300 ng) of DNA from each of the 32 individuals to yield a pooled sample containing a final DNA concentration of 10 ng/ $\mu$ L. The standard allele frequency samples were constructed by mixing two homozygous DNAs in various ratios.

### SNP Markers

Publicly available markers from the dbSNP database that were characterized by our group were used in this study.

### Assay Reagents

All reactions were run and read in 96-well black plates from MJ Research. PCR and SNP primers were obtained from Invitrogen. AmpliTaq Gold DNA polymerase was from Applied Biosystems. Exo-Sap (Exonuclease I and Shrimp Alkaline phosphatase) was purchased from USB. SNP detection kits (AcycloPrime) were generous gifts from Perkin-Elmer and included 10X reaction buffer, AcycloPol enzyme, and the AcycloTerminators mixture consisting of equal amounts of R110 and Tamra acycloterminators.



**Figure 4** The plot of slope ratio vs. allele frequency using data in Table 1. The scatterplot is the slope ratio vs. known allele frequency. The curve fitting is with the hyperbolic equation  $y = a/(1 + bX)$  with  $a$  and  $b$  being 1.02 ( $P < 0.0001$ ) and 0.66 ( $P < 0.0001$ ), respectively, which agrees well with predicted values of 1 and 0.67.

**Table 2.** Allele Frequency Estimates of Pooled DNA Sample Compared with Observed Allele Frequencies

Single nucleotide polymorphism markers (R110/Tamra)	Pool 1 32 African-American individuals		Pool 2 32 Asian-American individuals		Pool 3 32 European-American individuals	
	Known (allele)	Measured (mean ± SD)	Known (allele)	Measured (mean ± SD)	Known (allele)	Measured (mean ± SD)
rs3157 G/A	0.81 (A)	0.81 ± 0.007	0.55 (A)	0.53 ± 0.013	0.61 (A)	0.60 ± 0.012
rs3157 C/T		0.81 ± 0.007		0.52 ± 0.010		0.59 ± 0.018
rs688 G/A	0.48 (A)	0.47 ± 0.007	0.034 (A)	0.025 ± 0.015	0.45 (A)	0.44 ± 0.005
rs696 G/A		0.62 ± 0.006		0.44 ± 0.014		0.32 ± 0.024
rs696 C/T	0.61 (A)	0.61 ± 0.009	0.41 (A)	0.43 ± 0.006	0.31 (A)	0.35 ± 0.009
rs245 A/T	0.59 (A)	0.58 ± 0.015	0.55 (A)	0.54 ± 0.014	0.41 (A)	0.39 ± 0.010
rs3354 G/A	0.28 (A)	0.24 ± 0.018	0.30 (A)	0.25 ± 0.010	0.22 (A)	0.19 ± 0.009

SD, standard deviation of the mean: Four replicates were done for each pooled sample.

### PCR Amplification, and Degradation of Excess PCR Primers and dNTPs

DNA (15 ng) was amplified in 5- $\mu$ L reaction mixtures. The reaction mixture was heated for 12 min at 95°C for enzyme activation followed by 35 amplification cycles. Each cycle consisted of 30 sec of denaturation at 95°C, 40 sec of primer annealing at 58°C, and 40 sec of primer extension at 68°C. The reaction mixtures were then incubated at 72°C for 5 min for final primer extension. At the end of the reaction, the reaction mixtures were held at 4°C until further use. Two  $\mu$ L of Exo-Sap was added to the PCR reaction mixtures and incubated for 45 min at 37°C to degrade the excess PCR primers and dNTPs. The enzymes were then heat-inactivated at 95°C for 15 min.

### Primer Extension Reaction

SNP primers were designed using PrimerExpress to have a universal annealing temperature of 60°C. To the reaction mixtures from the previous step, 13  $\mu$ L of TDI cocktail was added according to the manufacturer's instructions (2  $\mu$ L 10X TDI reaction buffer, 0.5  $\mu$ L SNP primer [final concentration 2  $\mu$ M], 0.25  $\mu$ L Acyclo Enzyme, 2.25  $\mu$ L of the two-dye Acyclo-nucleotide terminators, and 8  $\mu$ L water). The reaction mixture was incubated at 93°C for 45 sec, 40 cycles of single-base extension at 93°C for 10 sec, and 58°C for 30 sec. For R110 and R6G dye-terminators, the fluorescence intensities were acquired during the annealing/extension phase of the primer extension cycles. The analysis was done using the multicomponent data from the Applied Biosystems 7700 Sequence Detector. For TexasRed and Tamra-terminators, the fluorescence intensities were read with an LJI Analyst (Molecular Devices).

### Theory of FQ Detection

In the FQ-TDI assay, two species with different fluorescence intensity values (quantum yield) exist for each of the two allelic dye-terminators: unincorporated dye-terminators ("free dye") with original high fluorescence intensity  $I_f$  and incorporated dye-terminators ("bound dye") into primers with quenched intensity  $I_b$ . For their intensity values at any particular cycle, we have the free dye at fraction  $f_f$  with a higher intensity value  $I_f$  and the bound dye at fraction  $f_b$  with a lower intensity value  $I_b$ . The observed intensity,  $I$ , at any given time is therefore:

$$I = f_f I_f + f_b I_b \quad (1)$$

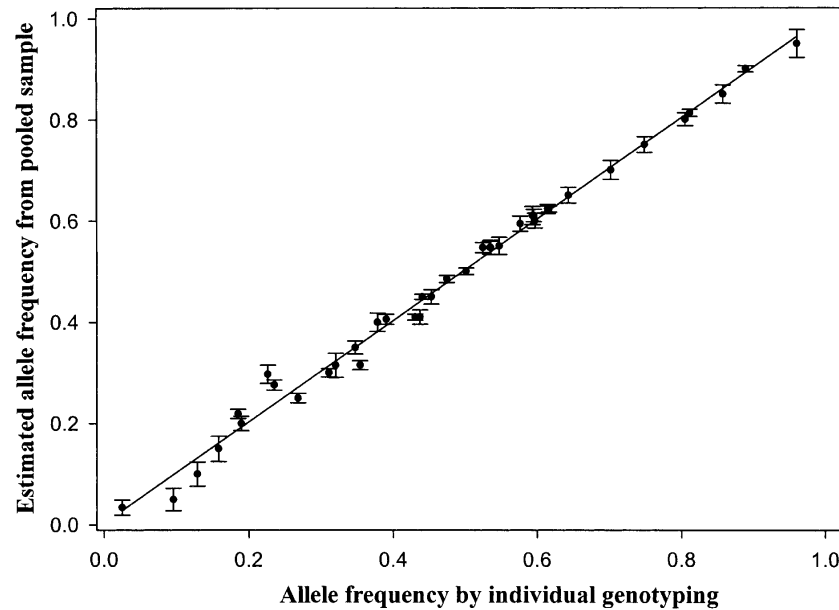
where  $f_f + f_b = 1$ .

Initially, all dye-terminators are unincorporated and the intensity is  $I_f$ , because  $f_f = 1$  and  $f_b = 0$ . When the reaction is

driven to completion, all dye-terminators are incorporated into primer and the observed intensity becomes  $I_b$ , with  $f_f = 0$  and  $f_b = 1$ . The fractions of free and bound dye-terminators change in opposite directions during the primer extension reaction, and the rate of this change depends on the amounts of the DNA template in the reaction.

In general, the rate of dye-terminator incorporation is determined by two factors: the amount of starting materials (PCR-amplified DNA fragments, SNP primers, and dye-terminators), and the incorporation efficiency of the dye-terminator by the DNA polymerase. Assuming that there are  $X$  copies of PCR-amplified fragments, then there are  $pX$  copies with the allele 1 and  $(1 - p)X$  copies with allele 2 (where  $p$  is the allele frequency of allele 1). Further assume that the two dye-terminators are present in equal amounts ( $Y$  molecules each). Because the SNP primers and the dye-terminators are in vast excess over the PCR products, dye-terminator incorporation is linear during initial cycles. As the dye-terminators and SNP primer are being consumed, dye-terminator incorporation becomes nonlinear, and the decrease of intensity observed reaches a plateau when dye terminators are all incorporated into primers.

During the first cycle of the primer extension reaction, each PCR-amplified fragment has one SNP primer hybridized to it. If every hybridized SNP primer gets extended, then  $pX$  and  $(1 - p)X$  dye-terminators are incorporated onto annealed SNP primers for the two dye-terminators, respectively. But in reality, neither the hybridization nor the primer extension reaction is 100% efficient. Dissimilar incorporation efficiencies of the dye-terminators will result in uneven incorporation of the two dye-terminators (Haff and Smirnov 1997). Denoting  $V_1$  as the incorporation efficiency for the dye-terminator corresponding to allele 1 and  $V_2$  for the dye-terminator corresponding to allele 2,  $pXV_1$  is the number of allele 1-specific dye-terminators incorporated, and  $(1 - p)XV_2$  is the number of allele 2-specific dye-terminators incorporated at the end of cycle one, because the incorporation of the two dye-terminators can be reasonably assumed to be independent. Accordingly,  $f_f$  can be represented as  $(Y - pXV_1)/Y$  for allele 1 and  $[Y - (1 - p)XV_2]$  for allele 2. Recalling that  $f_b$  is equal to  $1 - f_f$ ,  $f_f$  becomes  $(Y - npXV_1)/Y$  for allele 1 and  $[Y - n(1 - p)XV_2]$  for allele 2 after  $n$  cycles of linear incorporation. Substituting  $f_f$  and  $f_b$  of the two alleles into equation (1) separately, rearranging them and dividing both equations with  $I_f$ , equation (1) becomes equation (2) for allele 1 and equation (3) for allele 2.



**Figure 5** Allele frequency measurements in pools (y-axis) vs. allele frequency obtained by genotyping individuals (x-axis). A linear relation was observed with  $r^2 = 0.993$ , slope = 1.002 ( $P < 0.0001$ ).

$$(I_{1f} - I_1)/I_{1f} = np(X/Y)V_1(I_{1f} - I_{1b})/I_{1f} \quad (2)$$

$$(I_{2f} - I_2)/I_{2f} = n(1 - p)(X/Y)V_2(I_{2f} - I_{2b})/I_{2f} \quad (3)$$

Where  $(I_{1f} - I_1)/I_{1f}$  and  $(I_{2f} - I_2)/I_{2f}$  are the quenching percentage at a particular cycle, and  $(I_{1f} - I_{1b})/I_{1f}$  and  $(I_{2f} - I_{2b})/I_{2f}$  are the overall quenching for a particular marker. Both of these expressions are constants. Equations (2) and (3) have four variables each: X copies of PCR-amplified fragments, Y molecules of added dye-terminators, the incorporation efficiency V, and allele frequency p. Because X and Y vary from well to well, they are difficult to quantify. It is thus impossible to calculate the allele frequency value, p, by solving equations (4) and (5) (shown below) separately. By taking the ratio of equations (2) and (3), two of the variables (X, Y) are canceled out. If one estimates the ratio of the dye-terminator incorporation efficiency ( $V_1/V_2$ ) using a heterozygous individual (with “allele frequency” of each allele at 0.5 by definition) as the reference, the allele frequency, p, is now only related to the average intensity observed at the end of any particular cycle during the linear phase of the primer extension reaction.

Although taking just one intensity reading during the early phase of the primer extension reaction will yield a reasonable allele frequency estimate, the fact that this approach takes only one measurement during the presumed linear phase of the reaction makes it highly susceptible to random errors. Instead of taking the ratio of equations (2) and (3) directly, however, one can monitor the intensity at the end of every cycle and plot  $(I_f - I)$  against cycle n for each of the two dye-terminators. Two straight lines are obtained for the initial linear incorporation stage, and the slopes of the lines will be  $p(X/Y)V_1$  for allele 1 and  $(1 - p)(X/Y)V_2$  for allele 2. Because X, Y,  $V_1$ , and  $V_2$  remain constant for the same reaction, the ratio of the two slopes becomes equation (4):

$$\frac{\text{Slope for allele 1 (S}_1\text{)}}{\text{Slope for allele 2 (S}_2\text{)}} = \frac{pV_1(I_{1f} - I_{1b})/I_{1f}}{(1 - p)V_2(I_{2f} - I_{2b})/I_{2f}} \quad (4)$$

with X and Y canceled out as before. Rearranging equation (4) to solve for p, allele frequency can be obtained for the pooled DNA sample as equation (5):

$$p = \frac{1}{1 + \left\{ \frac{S_2/S_1}{I_{1b}/(I_{2f} - I_{2b})} \left[ \frac{V_1/V_2}{I_{2f}/I_{1f}} \right] \right\}} \quad (5)$$

$S_1$  and  $S_2$  are determined in the reaction involving the pooled sample, and  $V_1/V_2$  is the relative incorporation efficiency for two dye-acycloterminators obtained from a heterozygous sample.  $[(I_{1f} - I_{1b})/(I_{2f} - I_{2b})](I_{2f}/I_{1f})$  is the ratio of total quenching percentage for the two dye-terminators. We estimated the relative incorporation efficiency  $V_1/V_2$  for over 50 different SNPs using heterozygous samples, and the ratio ranges from 0.82–1.5 for R110- and R6G-labeled acycloterminators. Because the kinetic FQ-TDI approach includes many measurements during the primer extension reaction, the variation is reduced significantly.

### Data Analysis

Data were analyzed with Microsoft Excel and SigmaPlot software. The quenching percentage was calculated as the percentage of difference between original intensity and final intensity as dye-terminators incorporated into primers. Linear regression was performed, and results were discarded with  $r^2$  less than 0.99.

Linear regression was performed, and results were discarded with  $r^2$  less than 0.99.

### ACKNOWLEDGMENTS

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 USC section 1734 solely to indicate this fact.

### REFERENCES

Breen, G., Harold, D., Ralston, S., Shaw, D., and St Clair, D. 2000. Determining SNP allele frequencies in DNA pools. *Biotechniques* **28**: 464–466, 468–470.

Chen, X., Levine, L., and Kwok, P.Y. 1999. Fluorescence polarization in homogeneous nucleic acid analysis. *Genome Res.* **9**: 492–498.

Germer, S., Holland, M.J., and Higuchi, R. 2000. High-throughput SNP allele-frequency determination in pooled DNA samples by kinetic PCR. *Genome Res.* **10**: 258–266.

Giordano, M., Mellai, M., Hoogendoorn, B., and Momigliano-Richiardi, P. 2001. Determination of SNP allele frequencies in pooled DNAs by primer extension genotyping and denaturing high-performance liquid chromatography. *J. Biochem. Biophys. Methods* **47**: 101–110.

Gruber, J.D., Colligan, P.B., and Wolford, J.K. 2002. Estimation of single nucleotide polymorphism allele frequency in DNA pools by using pyrosequencing. *Hum. Genet.* **110**: 395–401.

Haff, L.A., and Smirnov, I.P. 1997. Single-nucleotide polymorphism identification assays using a thermostable DNA polymerase and delayed extraction MALDI-TOF mass spectrometry. *Genome Res.* **7**: 378–388.

Hsu, T.M., Chen, X., Duan, S., Miller, R.D., and Kwok, P.Y. 2001. Universal SNP genotyping assay with fluorescence polarization detection. *Biotechniques* **31**: 560, 562, 564–568, passim.

Kruglyak, L. 1999. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat. Genet.* **22**: 139–144.

Kwok, P.Y., Carlson, C., Yager, T.D., Ankener, W., and Nickerson, D.A. 1994. Comparative analysis of human DNA variations by fluorescence-based sequencing of PCR products. *Genomics* **23**: 138–144.

Nazarenko, I., Pires, R., Lowe, B., Obaidy, M., and Rashtchian, A.

2002. Effect of primary and secondary structure of oligodeoxyribonucleotides on the fluorescent properties of conjugated dyes. *Nucleic Acids Res.* **30**: 2089–2195.
- Risch, N., and Merikangas, K. 1996. The future of genetic studies of complex human diseases. *Science* **273**: 1516–1517.
- Seidel, C.A.M., Schulz, A., and Sauer, M.H.M. 1996. Nucleobase-specific quenching of fluorescent dyes. 1. Nucleobase one-electron redox potentials and their correlation with static and dynamic quenching efficiencies. *J. Phys. Chem. A* **100**: 5541–5553.
- Steenken, S. and Jovanovic, V. 1997. How easily oxidizable is DNA? One-electron reduction potentials of adenosine and guanosine radicals in aqueous solution. *J. Am. Chem. Soc.* **119**: 617–618.
- Torimura, M., Kurata, S., Yamada, K., Yokomaku, T., Kamagata, Y., Kanagawa, T., and Kurane, R. 2001. Fluorescence-quenching phenomenon by photo-induced electron transfer between a fluorescent dye and a nucleotide base. *Anal. Sci.* **17**: 155–160.
- Vieux, E.F., Kwok, P.Y., and Miller, R.D. 2002. Primer design for PCR and sequencing in high-throughput analysis of SNPs. *Biotechniques* **Suppl**: 28–32.
- Werner, M., Sych, M., Herbon, N., Illig, T., Konig, I.R., and Wjst, M. 2002. Large-scale determination of SNP allele frequencies in DNA pools using MALDI-TOF mass spectrometry. *Hum. Mutat.* **20**: 57–64.
- Wolford, J.K., Blunt, D., Ballecer, C., and Prochazka, M. 2000. High-throughput SNP detection by using DNA pooling and denaturing high performance liquid chromatography (DHPLC). *Hum. Genet.* **107**: 483–487.
- Xiao, M., Latif, S., and Kwok, P.Y. 2003. Kinetic FP-TDI assay for SNP allele frequency determination. *Biotechniques* **34**: 190–197.
- Zhou, G., Kamahori, M., Okano, K., Chuan, G., Harada, K., and Kambara, H. 2001. Quantitative detection of single nucleotide polymorphisms for a pooled sample by a bioluminometric assay coupled with modified primer extension reactions (BAMPER). *Nucleic Acids Res.* **29**: E93.

## WEB SITE REFERENCES

<http://lifesciences.perkinelmer.com/products/snp.asp>; Perkin-Elmer Web site.

Received November 12, 2002; accepted in revised form March 4, 2003.