Kinetic Polymerase Chain Reaction on Pooled DNA: A High-Throughput, High-Efficiency Alternative in Genetic Epidemiological Studies[1]

Jia Chen, Søren Germer, Russell Higuchi, Gertrud Berkowitz, James Godbold, and James G. Wetmur[2]

Departments of Community and Preventive Medicine [J. C., G. B., J. G.] and Microbiology [J. G. W.], Mount Sinai School of Medicine, New York, New York 10029, and Roche Molecular Systems, Inc., Alameda, California 94501 [S. G., R. H.]

Abstract

The ideal technology for screening single-nucleotide polymorphisms requires high throughput with minimal cost per sample, minimal usage of valuable DNA resources, and maximal flexibility for assessment of new polymorphisms. We demonstrate here the feasibility of kinetic allele-specific PCR with DNA pooling (S. Germer *et al.*, Genome Res., *10:* 258–266, 2000) in a population study that satisfies all of the mentioned criteria and offers a powerful new tool for detecting meaningful polymorphic differences in candidate gene association studies and genome-wide linkage dysequilibrium scans. Three individuals prepared pooled DNA samples from 269 individuals separated into three racial/ethnic groups: Caucasians ($n = 56$), African-Americans ($n = 86$), and Hispanics ($n = 127$). We used kinetic allele-specific PCR to determine the allele frequencies of the common paraoxonase 1 polymorphism, *PON1* Q191R, in these pools. Paraoxonase 1 is a critical enzyme for inactivating neurotoxic intermediates in the metabolism of organophosphates. In a blinded test of the technology, these nine pooled DNA samples were sent to Roche for genotyping by kinetic allele-specific PCR. The allele frequencies found were 0.266 ± 0.011, 0.386 ± 0.011, and 0.617 ± 0.010, respectively, which were comparable to the frequencies of 0.269, 0.403, and 0.622 determined by PCR-restriction fragment length polymorphism analysis. These same samples were genotyped on two kinetic PCR platforms from different manufacturers, using three different DNA polymerases. The results were comparable between both platforms and among all three polymerases. The results demonstrate a powerful new technology for determining frequencies of single-nucleotide polymorphisms in an epidemiological study.

Introduction

Most human diseases, such as cancer, cardiovascular disease, diabetes, and mental illness, are complex traits determined by the interplay of multiple genes and environmental factors with small to moderate effects. Identifying inherited and environmental markers and quantifying their risk associated with these complex diseases have been an enduring challenge. Such an effort requires the screening of a vast number of polymorphic markers in a vast number of individuals. Until recently, this task has been impossible because (*a*) only a small number of polymorphic markers have been identified, and (*b*) efficient technology capable of screening a large number of polymorphic markers in a large population has not been available. As the Human Genome Project nears its completion, a catalogue of human genome sequence variation is being established and made available to the public (1). A dense map of hundreds of thousands of SNPs,[3] a predominant (~90%) form of human sequence variation, will be constructed by the year 2003. This map will offer a powerful tool for identifying genes that make small but significant contributions to disease risk, for understanding relationships between genetic variation and diseases, and in turn for changing the future of disease prevention and treatment (2). Studying genome-wide sequence variations associated with human disease calls for the rapid development of efficient technologies that can identify subtle genetic risk factors that go undetected in existing study designs that use fewer markers and limited sample sizes. The ideal technology should provide for the rapid and efficient scoring of known SNPs in a large number of samples.

Although there are a number of high-throughput SNP analysis strategies in development (3), two competing molecular strategies dominate the field (4). One approach is to identify and/or type multiple polymorphisms one person at a time using, *e.g.*, high-density oligonucleotide hybridization arrays (5). Array hybridization, which relies on the difference between hybridization of matched and mismatched products to allele-specific oligonucleotides on the array, is powerful in SNP identification and has the advantage of maintaining individual information. However, the rate-limiting step for detecting SNPs is the PCR amplification, which has limited capacity for multiplexing. It addition, heterozygosity detection may not be completely foolproof for all SNPs (6), and the required amount of DNA for testing is substantial. An alternative high-throughput

Received 3/23/01; revised 10/18/01; accepted 10/26/01.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] This research was sponsored in part by grants from the NIH (ES09584), the Environmental Protection Agency (R827039), and Roche Molecular Systems, Inc. Dr. Chen was supported in part by a Career Development Award from the National Cancer Institute (CA81750).

[2] To whom requests for reprints should be addressed, at Department of Microbiology, Mount Sinai School of Medicine, 1 Gustave L. Levy Place, New York, NY 10029-6574. Phone: (212) 241-7685. Fax: (212) 534-1684; E-mail: james.wetmur@mssm.edu.

[3] The abbreviations used are: SNP, single nucleotide polymorphism; RFLP, restriction fragment length polymorphism.

strategy is to pool equal amounts of DNA from multiple individuals and then type one marker at a time. Pooled DNA samples have been used successfully with both microsatellite markers (7) and SNPs (8, 9), using fluorescent probes (10, 11), and capillary-based single-strand conformation polymorphism analysis (12).

Germer *et al.* (13) have developed a novel kinetic PCR method for pooled DNA that is capable of assessing SNP frequencies with high precision and efficiency. The method is accurate, time-saving, and inexpensive, requiring no labeled probes. It requires only a fraction of the genomic DNA from each individual needed by conventional genotyping methods without the need for SNP-specific optimization and post-PCR processing. It promises to be a highly efficient alternative that allows detection of the relatively weak but common genetic associations expected for complex diseases in genetic epidemiological studies. We demonstrate here the successful implementation of this technology in a population study in which we blind-tested the technology using a pool of DNA from 269 individuals that we had genotyped for the *PON1* Q191R polymorphism by the PCR-RFLP method. We pooled these individuals by their race/ethnicity to test the flexibility of various pooling strategies important in studying gene-environment interactions. We also discuss the importance and feasibility of applying this method to identifying disease genes as well as studying gene-environment and gene-gene interactions in genetic epidemiological studies.

Materials and Methods

Principle of Kinetic PCR with Pooled DNA. Kinetic PCR with pooled DNA has been described in detail elsewhere (13). In brief, the technology depends on allele-specific PCR, where the specificity of amplification results from placing the 3′ end of one of the primers directly over and matching one or the other variant nucleotides (14–17). Ideally, only the matched primer will be extended. In practice, the mismatched allele will also be extended, but with much less efficiency. The specificity can be improved by specific polymerases [*e.g.,* Stoffel fragment of Taq DNA polymerase (13, 18, 19)] and by use of a "hot start" method, which minimizes artifactual primer-dimer formation by preventing the DNA polymerase from initiating undesired primer extensions at low temperatures. It has been shown that mismatch amplification is frequently delayed by more than 10 cycles (20). To assess the allele frequency, pooled DNA samples with equal quantities of individual DNAs are analyzed independently with the two allele-specific primers. When the allele frequency is 50% in the pool, the yield of the two PCR reactions will reach a predetermined threshold value at the same cycle, assuming equal efficiency. When one allele is more common than the other, the amplification of the common allele will reach the threshold at an earlier cycle. A one-cycle delay implies that the ratio of two alleles in the pool is 1:2, assuming 100% PCR efficiency. Thus, the allele distribution can be calculated from the cycle delay. The cycle delay may be up to several cycles or any fraction thereof.

Study Population and DNA Samples. The study population was derived from prenatal patients in New York City who were participants in an ongoing cohort study on the effect of maternal exposure to pesticides and other toxicants on childhood neurodevelopment as part of the Mount Sinai Children's Environmental Health Center. All subjects gave informed consent for measurement of *PON1* genotypes as part of the study. The research protocol was approved by the Institutional Review Board of the Mount Sinai School of Medicine. Paraoxonase 1, the product of the *PON1* gene, is a critical enzyme for inactivating neurotoxic intermediates in the metabolism of organophosphate pesticides, and the activity on various substrates is affected by polymorphisms in the *PON1* gene. The study population consisted of 56 Caucasians, 86 African-Americans, and 127 Hispanics of Caribbean origin (mostly Puerto Ricans and Dominicans) from whom whole blood was collected. Genomic DNA was extracted from the buffy coat and purified with QIAamp blood kits (Qiagen) as described by the manufacturer.

Individual Genotypes. Individual genotypes of the *PON1* Q191R polymorphism were determined by a PCR-RFLP-based assay (21, 22). In brief, genomic DNA was amplified using PCR primers 5′-GTATGTTTTAATTGCAGTTTGAA-3′and 5′-TGAAATGTTGATTCCATTAGCAA-3′, where sequences with terminal AA sequences were chosen to suppress primer-dimer formation. Standard cycling conditions (1 min at 94°C, 1 min at 55°C, 3 min at 72°C) were used for Taq DNA polymerase in the buffer supplied by the manufacturer. The 207-bp PCR products were cleaved with *Alw*I and analyzed by fluorography after size fractionation on 1.2% agarose gels.

DNA Pooling. Individual DNA concentrations were determined from absorbance spectra measured with a Hewlett-Packard diode array spectrophotometer. Pooled DNA was generated by mixing 100 ng of DNA from individual samples. DNA pools were created for each of the three racial/ethnic groups (Caucasian, African-American and Hispanic), and each pooling was replicated independently by three investigators at the Mount Sinai School of Medicine. Thus, a total of nine pools were generated (3 races/ethnicities × 3 replicates); all subsequent measurements were carried out using these pools.

Kinetic PCR with Pooled DNA. In one set of experiments, kinetic PCR was carried out on aliquots of the nine DNA pools at Roche Molecular Systems in a GeneAmp 5700 Sequence Detection System (PE Applied Biosystems), using a "Gold" version of Stoffel Fragment DNA polymerase (23) as described previously (13). All other experiments were carried out at Mount Sinai School of Medicine on a LightCycler (Roche Molecular Biochemicals). Two different DNA polymerases were tested with the latter platform: FastStart Taq (Roche Molecular Biochemicals) and AmpliTaq Gold (PE Applied Biosystems). All three DNA polymerases required heat activation, one of the available methods for achieving hot starts and minimizing primer-dimer formation.

The primers for the *PON1* Q191R polymorphism were as follows:

5′-TATTTTCTTGACCCCTACTTACA-3′ (allele specific for 191Q)

5′-TTTCTTGACCCCTACTTACG-3′ (allele specific primer for 191R)

5′-CCACGCTAAACCCAAATACATCTC-3′) reverse common primer)

Reactions were assembled using micropipettors (Jencons, Ltd.). A basic master mix for the analyses on the LightCycler contained 1× AmpliTaq Gold buffer supplemented to final concentrations of 4 mM $MgCl_2$, 2% glycerol, 1× BSA (New England Biolabs), 5 units/20 μl of DNA polymerase, 0.5 μM reverse primer, 200 μM each deoxynucleotide triphosphate (with dUTP replacing dTTP), and 0.25× SYBR Green I (Molecular Probes). Two allele-specific master mixes were produced by addition of one or the other of the allele-specific primers to 0.5 μM in the basic master mix. Finally, individual 20-μl PCR solutions were prepared by the addition of 20 ng of pooled DNA template to an aliquot of one or the other of the allele-specific master mixes. The cycling condition included a

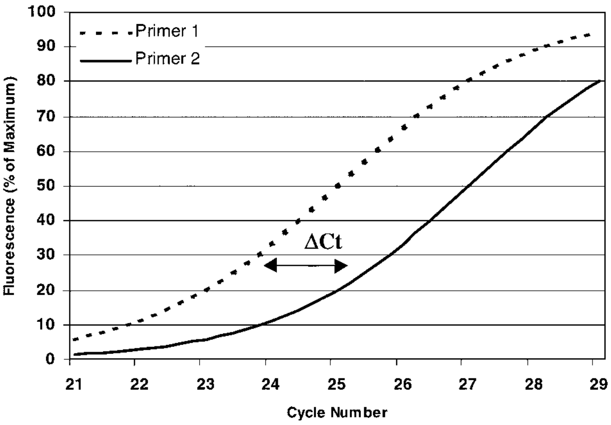

Fig. 1. Basis for determining allele frequency by kinetic allele-specific PCR. Aliquots of a pooled DNA sample are amplified using a primer specific for allele 1 or 2, respectively, and a common reverse primer. Fluorescence of SYBR Green I is measured as a function of cycle number. ΔCt is the difference in PCR cycle number at the point at which the two reactions have proceeded to the same extent toward completion.

heating step at 95° (4 min for FastStart Taq and 9 min for AmpliTaq Gold DNA) followed by 45 cycles of 25 s at 58°C, 25 s at 72°C, and 15 s at 95°C.

LightCycler Data Analysis. To determine the allele frequency in a pooled DNA sample, four kinetic PCR reactions were carried out in each of the two allele-specific master mixes. Ten replicate measurements of a heteroduplex sample were used to control for specificity of allele-specific PCR. The raw data were exported as an Excel spreadsheet, which gave the fluorescence as a function of cycle number (C) in each sample. The C value, reported as the average of four replicate runs, was determined as $(M - I)/S$, where M is the logarithmic mean of fluorescence signals, and I and S are the slope and intercept in the linear range of the PCR curve, respectively (Fig. 1). A spreadsheet patch to overlay onto LightCycler export data is available by e-mail upon request.[4]

Let C_1 and C_2 be the C values (average of 4 replicate runs) for kinetic PCR of a pooled DNA sample with allele-specific primers 1 and 2, and let $C_1{}'$ and $C_2{}'$ (average of 10 replicate runs) be the values for a heteroduplex sample used for calibration of allele specificity. In addition, let σ_1, σ_2, $\sigma_1{}'$, and $\sigma_2{}'$ be the corresponding SDs for the cycle numbers. The cycle difference then is $\Delta \mathrm{Ct} = (C_1 - C_2) - (\mathrm{C}_1{}' - \mathrm{C}_2{}')$ and the SD, in ΔCt, is $\sigma_{\Delta \mathrm{Ct}}$ and is calculated from the weighted root mean square of the SDs of the cycle numbers.

Allele Frequencies. Let F be the allele frequency matching primer 1 (191R allele):

$$F = \frac{100}{(2^{\Delta \mathrm{Ct}} + 1)} \qquad \text{(A)}$$

The SD of the measurement of F is σ_F:

$$\sigma_\mathrm{F} = \frac{F(100 - F)\, ln(2)\, \sigma_{\Delta \mathrm{Ct}}}{100}$$

The measurement error of F for M measurements is σ_m:

$$\sigma_m = \frac{\sigma_F}{\sqrt{M}} \qquad \text{(C)}$$

Sampling Error and SE of the Means. The sampling error in the allele frequency is defined here for a sample of size n (two alleles each) as (24):

$$\sigma_s = \sqrt{F(100 - F)/2n} \qquad \text{(D)}$$

The SE of the means is σ:

$$\sigma = \sqrt{\sigma_m{}^2 + \sigma_s{}^2} \qquad \text{(E)}$$

Correction for DNA Polymerase Allele Specificity. The maximum values for ΔCt, ΔCt(1), and ΔCt(2) were determined using primers 1 and 2, respectively, on 191R and 191Q homozygous controls. The signs of ΔCt(1) and ΔCt(2) were taken to be positive. Then:

$$F = \frac{100\,[2^{+\Delta \mathrm{Ct}} - 2^{-\Delta \mathrm{Ct}(2)}]}{\{2^{+\Delta \mathrm{Ct}}\,[1 - 2^{-\Delta \mathrm{Ct}(1)}] + [1 - 2^{-\Delta \mathrm{Ct}(2)}]\}} \qquad \text{(F)}$$

Results

Allele Frequencies of *PON1* Q191R by Kinetic PCR on Pooled DNA. Three different investigators at Mount Sinai School of Medicine (numbered 1–3) independently constructed DNA pools for each of three ethnic groups (56 Caucasians, 86 African-Americans, and 127 Hispanics of Caribbean origin), for a total of nine samples. Aliquots from these pools were coded and sent to investigators at Roche Molecular Systems for determination of *PON1* Q191R allele frequencies by their previously published kinetic PCR method (13). These results for the kinetic PCR method were compared with allele frequencies determined by PCR-RFLP analysis (Table 1). The agreement between the allele frequencies determined by kinetic PCR and the PCR-RFLP was quite good. Specifically, results from kinetic PCR *versus* PCR-RFLP analysis were 26.6 *versus* 26.9, 61.7 *versus* 62.2, and 38.6 *versus* 40.3 for the pooled samples from Caucasians, African-Americans, and Hispanics, respectively. The sampling errors for these same populations were 4.2, 3.7 and 3.1, respectively. On the other hand, the measurement errors for kinetic PCR were 0.6, 0.5 and 0.7, respectively, which were significantly smaller than the sampling errors.

Polymerase Dependence. To assess the polymerase dependence of the assay, kinetic PCR was carried out in a LightCycler on the same DNA pools, using two commercially available DNA polymerases, AmpliTaq Gold and FastStart Taq, both with enhanced PCR specificity because of a required heat activation of the polymerases. The assay at Roche Molecular Systems used a Gold version of Stoffel Fragment DNA polymerase, which also required heat activation but was not commercially available. Results are compared in Table 2. The point estimates were similar in all measurements. The measurement errors with Stoffel Fragment DNA polymerase (0.6–0.8) were slightly lower than those for FastStart Taq (1.0–1.6) or AmpliTaq Gold (0.8–1.1). Compared with the sampling error, the measurement errors were much smaller.

Platform Dependence. Kinetic PCR was performed on two different commercially available thermocyclers, a GeneAmp 5700 and a LightCycler. As can be seen in Table 3, the allele frequencies determined on these two platforms were quite comparable. The point estimates were similar in all measurements. The measurement errors with the GeneAmp 5700 (0.6–0.8) were comparable

[4] Send requests by e-mail to: james.wetmur@mssm.edu.

Table 1 Allele frequencies by race/ethnicity determined by kinetic PCR

	Investigator						Total		σ_m	RFLP
	1		2		3					
	Mean	σ_F	Mean	σ_F	Mean	σ_F	Mean	σ_F		
Caucasians	25.1	1.6	27.2	2.2	27.4	1.7	26.6	2.0	0.6	26.9
African-Americans	61.2	2.0	62.3	1.8	61.6	1.1	61.7	1.6	0.5	62.2
Hispanics	39.2	1.0	40.6	2.1	36.1	1.7	38.6	2.5	0.7	40.3

Table 2 Allele frequencies by race/ethnicity determined by kinetic PCR with three DNA polymerases

	DNA polymerases						Total	
	Stoffel		FastStart		AmpliTaq Gold			
	Mean	σ_m	Mean	σ_m	Mean	σ_m	Mean	σ_m
Caucasians	26.6	0.8	31.4	1.6	27.3	0.9	28.4	0.7
African-Americans	61.7	0.6	63.5	1.0	63.5	1.1	62.9	0.5
Hispanics	38.6	0.7	39.1	1.4	35.8	0.8	37.9	0.7

Table 3 Allele frequencies by race/ethnicity determined by kinetic PCR on two different instruments

	Platform				Total	
	GeneAmp 5700		LightCycler			
	Mean	σ_m	Mean	σ_m	Mean	σ_m
Caucasians	26.6	0.8	29.4	0.8	28.0	0.6
African-Americans	61.7	0.6	63.5	0.7	62.6	0.5
Hispanics	38.6	0.8	37.5	0.9	38.1	0.6

Table 4 Allele frequencies by race/ethnicity determined by kinetic PCR from three independent pools

	Investigator						Total	
	1		2		3			
	Mean	σ_m	Mean	σ_m	Mean	σ_m	Mean	σ_m
Caucasians	26.9	0.8	28.7	1.0	29.8	1.5	28.4	0.7
African-Americans	63.6	1.0	63.8	0.8	61.3	1.1	62.9	0.5
Hispanics	39.7	1.0	38.7	1.3	35.2	0.7	37.9	0.7

to those for the LightCycler (0.7–0.9). Again, the measurement errors were much smaller than the sampling errors.

Investigator Dependence. As can be seen in Table 4, the allele frequencies determined by kinetic PCR were reproducible among the pools produced by the three different investigators. Again the point estimates were similar in all measurements. Measurement errors, calculated from all measurements using the three different polymerases, were comparable for the three investigators (0.8–1.5); they were also smaller than the sampling errors.

Correction for DNA Polymerase Allele Specificity. Allele specificity was determined by amplification of eight homozygote DNA samples (four with the QQ and four with the RR genotype), each in quadruplicate. Although Stoffel Fragment is known to be more allele-specific than full-length Taq DNA polymerases, FastStart Taq and AmpliTaq Gold DNA polymerases performed quite well, having maximum ΔCt values of 5.7 and 6.0 for the Q allele and 7.5 and 7.5 for the R allele, respectively. Allele frequencies measured with these two polymerases on a LightCycler, uncorrected and corrected for DNA polymerase allele specificity, are presented in Table 5. The corrections in allele frequencies were small: 29.4 ± 1.6 to 28.5 ± 1.6, 63.5 ± 1.5 to 63.5 ± 1.5, and 37.9 ± 1.8 to 36.9 ± 1.8 (mean ± $2\sigma_m$) for Caucasians, African-Americans, and Hispanics, respectively.

A comparison of allele frequencies measured with highly allele-specific Stoffel Fragment (Table 1), corrected allele frequencies measured with the two less allele-specific Taq DNA polymerases (Table 5), and allele frequencies determined by RFLP analysis revealed good agreement, with values of 26.6 ± 1.2, 28.5 ± 1.6, and 26.9 for Caucasians; 61.7 ± 1.0, 63.5 ± 1.4, and 62.2 for African-Americans; and 38.6 ± 1.4, 36.9 ± 1.8, and 40.3 for Hispanics, respectively. The largest measurement errors corresponded to a sampling error in a population of ~1500 individuals.

Discussion

Complex genetic disorders such as cancer arise from genetic contributions of multiple genes interacting with environmental factors and among themselves. Unlike the Mendelian diseases, in which a single gene confers an enormous risk, the genetic contribution of any specific gene in a complex disease is likely to be small. To dissect a complex disease, one needs to screen numerous polymorphic markers in a large population. The human genome is polymorphic, with ~30,000 common coding SNPs in the genome (25). By comparing allele frequencies of SNPs between a diseased and a healthy population, one can assess whether the gene is related to the disease and the magnitude of the risk associated with it. There are four main reasons for the increasing popularity of SNPs as markers in genetic analysis: (*a*) compared with microsatellite markers, SNPs are far more prevalent in the genome; (*b*) some of the SNPs are located in functional domains of genes that directly affect protein structure or expression levels and may, therefore, represent candidate alterations for genetic mechanisms in disease; (*c*) SNPs are inherited stably compared with microsatellite markers; (*d*) SNPs are easily adaptable for high-throughput genotyping, offering sufficient power for genetic analyses.

One approach that can greatly reduce laboratory effort and increase efficiency is DNA pooling, in which DNA from multiple individuals is pooled before genotyping. Such a strategy has been shown to be effective in identifying disease-related genes in several settings, including Mendelian founder mutations (7, 26) as well as complex diseases (8). Rather than assessing each individual's genotype, allele frequency of the tested marker is measured in a pool of DNA. For example, affected individuals can be grouped, as can unaffected individuals. Allele frequencies of the tested SNPs can be ascertained in each group and compared. Association with disease is im-

Table 5 Allele frequencies by race/ethnicity determined by kinetic PCR: effect of correction for DNA polymerase allele specificity

	Investigator						Total		σ_m	RFLP
	1		2		3					
	Mean	σ_F	Mean	σ_F	Mean	σ_F	Mean	σ_F		
Corrected for allele specificity										
Caucasians	26.8	2.5	28.5	3.5	30.1	4.1	28.5	3.9	0.8	26.9
African-Americans	64.8	3.5	64.5	2.6	61.0	4.6	63.5	3.7	0.8	62.2
Hispanics	39.3	2.2	37.1	5.1	34.4	4.0	36.9	4.3	0.9	40.3
Uncorrected for allele specificity										
Caucasians	27.8	2.5	29.4	3.5	31.0	4.1	29.4	3.8	0.7	26.9
African-Americans	64.8	3.5	64.6	2.6	61.1	4.6	63.5	3.6	0.7	62.2
Hispanics	39.9	2.2	37.7	5.1	34.7	4.0	37.5	4.3	0.9	40.3

plied if a difference in allele frequency between the pools is detected. Estimation of allele frequency in only two pools of DNA rather than in a large number of subjects individually can achieve large savings in both labor and materials, especially individual DNA.

To successfully implement DNA pooling in association studies, it is crucial to develop a method that is capable of accurately measuring the allele frequency in a pooled sample. There are several kinetic PCR-based approaches that permit SNP frequency determination in a single PCR reaction, including TaqMan (10, 27) and molecular beacons probes (28, 29). The major disadvantage of these methods is that both require fluorescent labeling of probes, which significantly increases the expense of the assay. Another alternative is allele-specific kinetic PCR, first developed by Germer *et al.* (13), in which the allele frequency of a SNP is reflected by the difference in PCR cycles needed to generate detectable PCR product with wild-type- and variant-specific primers. This method does not require expensive fluorescent probes and, in turn, reduces costs. We have demonstrated the feasibility and flexibility of implementing this technology in a population study and have demonstrated that this technology offers a precise tool for conducting genetic epidemiological research. The advantages are as follows.

High Throughput. In this experiment, four PCR reactions were performed on three pools of individuals of different race/ethnicity from a total of 269 individuals to determine the allele frequency of for the *PON1* Q191R polymorphism, increasing the throughput by >20-fold. A much higher than 20-fold increase in throughput can be easily achieved. As sample size increases to ∼1000, the sampling error becomes similar in magnitude to the measurement error. As a result, we can easily achieve a 250-fold increase in throughput by creating a pool of 1000 or more individual DNAs.

Accuracy. We have demonstrated that allele-specific kinetic PCR is highly accurate for pooled DNA samples. Three DNA pools comprising samples from individuals of different ethnicity were prepared separately by three laboratory technicians, and the allele frequency of each pool was determined by four replicate kinetic PCRs (Table 4). The results for the three separate pools prepared by individual investigators were nearly identical and were highly comparable to those determined by PCR-RFLP analysis.

Conservation of Genomic DNA. With the conventional PCR-RFLP method, ∼20 ng DNA is needed from each individual for each SNP tested. A modest scan of 1000 SNPs requires 20,000 ng of DNA from each individual. On the other hand, with kinetic PCR of pooled DNA, the quantity of each DNA sample, which depends on the number of tested markers and the size of the pool, can be drastically reduced. It can be calculated as 160 ng of DNA (*i.e.,* 20 ng per reaction × 2 pools × 4 replicate reactions) multiplied by the number of SNPs, divided by the sample size of the pool. In the case of 1000 SNPs and 1000 samples, only 160 ng of DNA is needed. The amount of DNA can be decreased further by increasing the size of the pool. Conservation of valuable DNA resources is crucial for epidemiological studies, in which collecting blood is always difficult and the amount of genomic DNA is limited and precious.

Robustness and Flexibility. Kinetic PCR of pooled DNA is a homogeneous assay, meaning that it requires no post-PCR processing. The method is amenable to introduction of new markers. We have demonstrated that the method is compatible with at least two kinetic PCR platforms (Table 3) and three different thermostable DNA polymerases (Table 2). To facilitate the automation of SNP screening, a computerized primer design program can be implemented (30).

Implementation to Genetic Epidemiological Studies. An ideal methodology for dissecting a complex disease thus requires the capability and flexibility to study both gene-environment and gene-gene interactions. In this report, we have discussed strategies of applying the methodology of kinetic PCR of pooled DNA to genetic epidemiological studies as well as its feasibility and limitations.

Before any laboratory experiments are carried out, a set of *a priori* hypotheses with sound biological rationale should be proposed. A panel of candidate markers (*i.e.,* SNPs), either mechanism specific or genome wide, should be established. An initial screening of these SNPs would be performed, and their frequencies would be compared according to the disease status, *i.e.,* the “affected” *versus* “unaffected” pool. Once disease-associated markers have been detected with different allele frequencies in two pools, one would then study their interactions with the environmental or other genetic factors in relation to pathogenesis of human diseases. Meanwhile, we have to bear in mind that “false negatives” may arise because certain phenotypes (*e.g.,* disease) are only associated with genotypes (*e.g.,* homozygous variant) not directly represented by allele frequency.

Gene-Environment Interactions. The major downside of frequency determination by kinetic PCR on pooled DNA is the loss of individual information. Nevertheless, by stratifying samples according to potential risk factors, this technology is readily applicable to investigations of gene-environment interactions. On the basis of *a priori* hypotheses, pooling strategies would be designed according to exposure scenarios for the proposed environmental factors, *e.g.,* smoking status, use of hormone replacement therapy, and levels of alcohol intake. For example, within the affected and unaffected pools, DNA could be pooled based on an individual's environmental exposure (*e.g.*, "exposed" *versus* "unexposed"), which usually is available through questionnaires or, sometimes, biological markers. Thus, four pools would be created: "unaffected, unexposed"; "unaffected, exposed"; "affected, unexposed"; and "affected, exposed." Each pool should be matched for potential confounding factors, such as age and race/ethnicity. The importance of such matching is illustrated by the race/ethnicity distribution of *PON1* Q191R alleles uncovered in this investigation. Allele frequencies in each pool could then be determined by kinetic PCR and compared among the groups. The presence of gene-environment interactions is implied if a tested marker is enriched in a "multiplicative fashion" in the affected, exposed pool compared with the unaffected, unexposed pool. A formal statistical test for interactions needs to be developed for such an investigation.

Unlike the conventional genetic epidemiological studies in which investigation of gene-environmental interactions is a post-laboratory process and can easily be explored through data stratification on a computer, studying interactions using pooled DNA involves reanalyzing DNA samples in the laboratory. In this case, it becomes crucial to establish a set of *a priori* hypotheses with strong biological rationale and then to carefully design the pooling strategy accordingly. This approach leaves little room for fishing expeditions in the data set and prevents overexploitation. In this study, we have demonstrated that repooling of DNA (in our case, by ethnicity) can easily be achieved; it offers sufficient flexibility to explore various gene-environment interactions.

Gene-Gene Interactions. Loss of individual information through DNA pooling imposes difficulty in the investigation of gene-gene interaction. We would propose to first perform initial screening as discussed previously and select genes with different allele distribution between affected and unaffected pools. Because of this selection, a smaller panel of SNPs needs to be genotyped individually in the population. Depending on the number of disease-associated markers, scoring all of them individually still could be a formidable task. Most likely, however, a great majority of the markers tested in the initial screening will turn out be nonfunctional. The feasibility of the proposed strategy of combining kinetic PCR on pooled DNA and high-throughput genotyping needs to be demonstrated in further studies.

Acknowledgments

We thank Madhu Kumar, Caroline Palomeque, and Susana Gonzalez for preparing the pooled DNA samples.

References

1. Collins, A., Lonjou, C., and Morton, N. E. Genetic epidemiology of single-nucleotide polymorphisms. Proc. Natl. Acad. Sci. USA, *96:* 15173–15177, 1999.

2. Collins, F. S., Guyer, M. S., and Charkravarti, A. Variations on a theme: cataloging human DNA sequence variation. Science (Wash. DC), *278:* 1580–1581, 1997.

3. Kristensen, V., Kelefiosis, D., Kristensen, T., and Borrensen-Dale, L. High-throughput methods for detection of genetic variation. Biotechniques, *30:* 318–332, 2001.

4. Landegren, U., Nilsson, M., and Kwok, P. Y. Reading bits of genetic information: methods for single-nucleotide polymorphism analysis. Genome Res., *8:* 769–776, 1998.

5. Wang, D. G., Fan, J. B., Siao, C. J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubbell, E., Robinson, E., Mittmann, M., Morris, M. S., Shen, N., Kilburn, D., Rioux, J., Nusbaum, C., Rozen, S., Hudson, T. J., and Lander, E. S. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science (Wash. DC), *280:* 1077–1082, 1998.

6. Hacia, J. G., and Collins, F. S. Mutational analysis using oligonucleotide microarrays. J. Med. Genet., *36:* 730–736, 1999.

7. Barcellos, L. F., Klitz, W., Field, L. L., Tobias, R., Bowcock, A. M., Wilson, R., Nelson, M. P., Nagatomi, J., and Thomson, G. Association mapping of disease loci, by use of a pooled DNA genomic screen. Am. J. Hum. Genet., *61:* 734–747, 1997.

8. Arnheim, N., Strange, C., and Erlich, H. Use of pooled DNA samples to detect linkage disequilibrium of polymorphic restriction fragments and human disease: studies of the HLA class II loci. Proc. Natl. Acad. Sci. USA, *82:* 6970–6974, 1985.

9. Shaw, S. H., Carrasquillo, M. M., Kashuk, C., Puffenberger, E. G., and Chakravarti, A. Allele frequency distributions in pooled DNA samples: applications to mapping complex disease genes. Genome Res., *8:* 111–123, 1998.

10. Holland, P. M., Abramson, R. D., Watson, R., and Gelfand, D. H. Detection of specific polymerase chain reaction product by utilizing the 5′-3′ exonuclease activity of *Thermus aquaticus* DNA polymerase. Proc. Natl. Acad. Sci. USA, *88:* 7276–7280, 1991.

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

12. Sasaki, T., Tahira, T., Suzuki, A., Higasa, K., Kukita, Y., Baba, S., and Hayashi, K. Precise estimation of allele frequencies of single-nucleotide polymorphisms by a quantitative SSCP analysis of pooled DNA. Am. J. Hum. Genet., *68:* 214–218, 2001.

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

14. Newton, C. R., Graham, A., Heptinstall, L. E., Powell, S. J., Summers, C., Kalsheker, N., Smith, J. C., and Markham, A. F. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). Nucleic Acids Res., *17:* 2503–2516, 1989.

15. Sarkar, G., Cassady, J., Bottema, C. D., and Sommer, S. S. Characterization of polymerase chain reaction amplification of specific alleles. Anal. Biochem., *186:* 64–68, 1990.

16. Bottema, C. D., and Sommer, S. S. PCR amplification of specific alleles: rapid detection of known mutations and polymorphisms. Mutat. Res., *288:* 93–102, 1993.

17. Wu, D. Y., Ugozzoli, L., Pal, B. K., and Wallace, R. B. Allele-specific enzymatic amplification of β-globin genomic DNA for diagnosis of sickle cell anemia. Proc. Natl. Acad. Sci. USA, *86:* 2757–2760, 1989.

18. Lawyer, F. C., Stoffel, S., Saiki, R. K., Chang, S. Y., Landre, P. A., Abramson, R. D., and Gelfand, D. H. 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, 1993.

19. Tada, M., Omata, M., Kawai, S., Saisho, H., Ohto, M., Saiki, R. K., and Sninsky, J. J. Detection of *ras* gene mutations in pancreatic juice and peripheral blood of patients with pancreatic adenocarcinoma. Cancer Res., *53:* 2472–2474, 1993.

20. Higuchi, R., Fockler, C., Dollinger, G., and Watson, R. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. Biotechnology, *11:* 1026–1030, 1993.

21. Adkins, S., Gan, K. N., Mody, M., and La Du, B. N. Molecular basis for the polymorphic forms of human serum paraoxonase/arylesterase: glutamine or arginine at position 191, for the respective A or B allozymes. Am. J. Hum. Genet., *52:* 598–608, 1993.

22. Humbert, R., Adler, D. A., Disteche, C. M., Hassett, C., Omiecinski, C. J., and Furlong, C. E. The molecular basis of the human serum paraoxonase activity polymorphism. Nat. Genet., *3:* 73–76, 1993.

23. Birch, D. E. Simplified hot start PCR. Nature (Lond.), *381:* 445–446, 1996.

24. Glantz, S. Primer on Biostatistics. New York: McGraw-Hill, 1977.

25. Collins, F. S., Patrinos, A., Jordan, E., Chakravarti, A., Gesteland, R., and Walters, L. New goals for the U. S. Human Genome Project: 1998–2003. Science (Wash. DC), *282:* 682–689, 1998.

26. Carmi, R., Rokhlina, T., Kwitek-Black, A. E., Elbedour, K., Nishimura, D., Stone, E. M., and Sheffield, V. C. Use of a DNA pooling strategy to identify a human obesity syndrome locus on chromosome 15. Hum. Mol. Genet., *4:* 9–13, 1995.

27. Lee, L. G., Connell, C. R., and Bloch, W. Allelic discrimination by nick-translation PCR with fluorogenic probes. Nucleic Acids Res., *21:* 3761–3766, 1993.

28. Kostrikis, L. G., Tyagi, S., Mhlanga, M. M., Ho, D. D., and Kramer, F. R. Spectral genotyping of human alleles. Science (Wash. DC), *279:* 1228–1229, 1998.

29. Tyagi, S., and Kramer, F. R. Molecular beacons: probes that fluoresce upon hybridization. Nat. Biotechnol., *14:* 303–308, 1996.

30. Beasley, E., Myers, R., Cox, D., and Lazzaroni, L. Statistical refinement of oligonucleotide design parameters. *In:* M. Innes, D. Gelfand, and J. Sninsky (eds.), PCR Applications: Protocols for Functional Genomics, pp. 55–73. San Diego, CA: Academic Press, 1999.