Genomics 92 (2008) 429–435

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Applying polynomial standard curve method to correct bias encountered in estimating allele frequencies using DNA pooling strategy

Hsin-Hung Chen [a,c], Yuh-Shan Jou [a], Wei-Jei Lee [b], Wen-Harn Pan [a,*]

[a] Institute of Biomedical Sciences, Academia Sinica, Taiwan, ROC
[b] Department of Surgery, Min-Sheng General Hospital, Taiwan, ROC
[c] Department of Health Sciences, Chang Jung Christian University, Taiwan, ROC

ARTICLE INFO

Article history:
Received 17 March 2008
Accepted 18 August 2008
Available online 8 October 2008

Keywords:
Pooling
Association study
κ correction
Polynomial standard curves method (PSCM)
SNaPshot™

ABSTRACT

DNA pooling approach is a cost-saving strategy which is crucial for multiple-SNP association study and particularly for laboratories with limited budget. However, the biased allele frequency estimates cannot be completely abolished by κ correction. Using the SNaPshot™, we systematically examined the relations between actual minor allele frequencies (AMiAFs) levels and estimates obtained from the pooling process for all six types of SNPs. We applied principle of polynomial standard curves method (PSCM) to produce allele frequency estimates in pooled DNA samples and compared it with the κ method. The results showed that estimates derived from the PSCM were in general closer to AMiAFs than those from the κ method, particularly for C/G and G/T polymorphisms at the range of AMiAF between 20–40%. We demonstrated that applying PSCM in the SNaPshot™ platform is suitable for multiple-SNP association study using pooling strategy, due to its cost effectiveness and estimation accuracy.

Introduction

The pooling technique in genetic association studies reduces (by 100–1,000-fold) the number of required genotyping reactions and thus is a cost-effective mean to map loci that confer higher susceptibility to complex diseases [1–7], particularly for laboratories with limited resources. This technique has been applied to association studies that are based on unrelated individuals (case-control study) or families [8]. Efforts have been made to adopt DNA pooling scheme for varied platforms; these include SNaPshot™ [9–11], SNaPIT™ [1–4], Pyrosequencing [7], Invader assay [12], bioluminometric assay [13], TAQMAN™, and denaturing HPLC [14–16]. Recently, the GeneChip10K/100K (Affymetrix™) and Illumina Infinium I Chip have been applied in genome-wide association study with DNA pooling strategy [17–22].

However, there are two major problems associated with these techniques: differential efficiencies on the polymerase chain reaction (PCR) and allele-specific hybridization of heterozygote which may lead to biased findings with regard to the relations between genetic variants and diseases [22–24,175]. Therefore, a correction factor, generally expressed as "κ", is often applied to correct the biased estimates. [5,9,25].

To apply the κ correction on DNA pooling studies, one has to assume that the degree of preferential amplifications/hybridizations holds a linear trend (or an arithmetic progressive trend) with increments of allele frequency. However, the phenomenon was not carefully studied for all types of SNPs and with a wide range of AMiAFs (actual minor allele frequencies) [7,9,17]. Using κ correction, it has been observed by Gruber et al. [7] that three specific SNPs had the greater variability in allele frequency estimates of pooled DNA samples when their AMiAFs were more than 30%, indicating that the greater variability may be attributable to particular polymorphism. However, this report did not exactly denote which polymorphisms these SNPs were [7]. Moreover, it had been reported that the variations in κ were large enough to produce unacceptable error rates [26].

For GeneChip®, Brohede et al. [17] developed and applied a modified κ correction method for genome-wide DNA pooling study, integrating "κ correction" with second-degree polynomial standard curves constituted from 3 reference genotypes, 'aa', 'Aa' and 'AA' ('A' allele frequency: 0%, 50% and 100%, respectively). Using this same platform, Simpson et al. [24] addressed the κ variation by integrating κ values obtained from 10 probe-pair hybridizations. Recently, Macgregor, S. suggested that pooling variation in array-based platforms were mostly attributable to chip-to-chip variation [27]. Thus, it is very important to take into considerations of the κ variation and to find the optimal number of replicated genotyping when we apply array-based platforms to genome-wide DNA pooling study.

* Corresponding author. No. 128, Sec. 2, Academia Rd, Nankang, Taipei, 115 Taiwan, ROC. Fax: +886 2 27823047.
E-mail address: pan@ibms.sinica.edu.tw (W.-H. Pan).

0888-7543/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2008.08.010

At the same time, a breakthrough is urgently required to correct differential amplifications in pooling approach for fine-mapping or for candidate gene mapping endeavors with a suitable genotyping platform. A multiplex-genotyping platform, SNaPshot™, has been proven very useful for DNA pooling studies because it is relatively easy to use and economic [9,11]. In addition, the SNaPshot uses a "single-based extension" reaction which is DNA sequence and steric structure independent. Furthermore, SNaPshot kit uses AmpliTaq® DNA Polymerase, a modified polymerase optimized to minimize the dye-specific and base-specific differences on incorporation efficiency and each ddNTP in the kit is labeled with a fixed fluorescent dye. Thus luminescence ratio is almost fixed for a given polymorphism type irrespective of the dye used or the local sequence.

Currently, there were relative few studies systematically investigating the κ variations with this platform. Therefore, we intended to explore the degree of differential amplifications/hybridizations for six types of SNPs (A/T, A/C, C/T, A/G, C/G, and G/T polymorphisms) and at multiple AMiAF levels (5, 10, 20, 30, 40, 45, and 50%), using SNaPshot™. The aim of this study was to apply a second-degree polynomial standard curve method (PSCM) relating the seven reference AMiAFs with their relative signal intensity readings in SNaPshot™ reactions to obtain more accurate estimates of allele frequencies on pooled DNA samples. Furthermore, we also compared the accuracy and precision of estimates derived from the PSCM and the conventional κ correction method.

Materials and methods

This study was performed in preparation for our ongoing obesity genetic study. DNA samples from 125 participants of the study were used for this methodology investigation. The participants provided signed informed consent to the obesity study and related research. The study protocol and informed consent form were approved by the IRB (Institutional Review Board) of the Min-Sheng General Hospital.

DNA preparation and quantification

Genomic DNA was extracted from buffy coat layer. The quality of the DNA was assessed by the 260 nm to 280 nm ratio readings obtained from the spectrophotometer, and samples with ratios in the range of 1.8–2.0 were used. Each DNA sample was quantified using the PicoGreen dsDNA quantification reagent (Molecular Probes, Oregon, USA) and diluted to 6.5±0.5 ng/μl by using Tecan's Freedom EVOlyzer (TECAN Schweiz AG, Mannedorf, Switzerland).

Preparation of pooled DNA samples

Before pooling, the individual genotype of each DNA sample was validated for each SNP. For each SNP, different proportions of validated DNA samples, homozygote or heterozygote, were mixed according to the principle of Hardy–Weinberg equilibrium to construct pooled DNA samples with the exception for AMiAFs 5% and 50%. Table 1 shows the compositions of all the pooled DNA samples. Triplicate DNA pools were formed for each SNP at every designated AMiAF. One PCR reaction was carried out for each DNA pool formation, and then the SNaPshot™ reaction was executed on each of the PCR products. Signal peak heights (the signal intensities) of the major and minor peaks were subsequently measured on each PCR product after SNaPshot™ reaction.

PCRing SNPs

Ten SNPs (rs4844480, rs11761556, rs1938484, rs182052, rs3774262, rs2282739, rs1862513, rs9430012, rs696217, and rs11808690) with varied polymorphic types were studied and retrieved from the dbSNP database (http://www.ncbi.nlm.nih.gov/SNP). Primers were designed by the Web Primer program (http://seq.yeastgenome.org/cgi-bin/web-primer) and synthesized by Purigo Biotech, Inc. (Taipei, Taiwan) (Appendix 1). PCR reactions were carried out on the ABI-9700 (Applied Biosystems Inc., CA, USA) using 6.5±0.5 ng genomic DNA, 0.5 μM of each PCR primer, 150 μM dNTPs (Jena Bioscience), 25 mM $MgCl_2$, 0.4 U AmpiTaq Gold (Applied Biosystems), 1.5 μl AmpiTaq Gold buffer, and 7 μl distilled water for a total volume of 15 μl. The quality of the PCR products was assessed on a 2.5% agarose gel. PCR primers and dNTPs were removed from PCR products before the SNaPshot™ reaction (SNP genotyping); 7.5 μl of PCR products was incubated with 3 U of SAP (shrimp alkaline phosphatase) and 2 U of ExoI (USB Corporation) for 1 h at 37 °C for clean up, followed by 15 min at 75 °C for enzyme inactivation. Each cleaned PCR product was quantified by the PicoGreen method and then diluted to a range of 0.010–0.015 pmol for the SNaPshot™ reaction.

Table 1
Formations of pooled DNA samples with 7 designated minor allele frequencies on each SNP

AMiAFs[a]	Number of pipettes from the following 3 genotypes of DNA samples[b]		
m %	mm	Mm	MM
5%	1	8	91
10%	1	18	81
20%	4	32	64
30%	9	42	49
40%	16	48	36
45%	20	50	30
50%	25	50	25

[a] AMiAFs (m %): Actual Minor Allele Frequencies mm: homozygote of minor allele; Mm: heterozygote; MM: homozygote of major allele.
[b] An artificial pooled DNA sample was formed with a total of 100 pipettes (5 μl/pipette) from different validated DNA samples. The concentration of each DNA sample solution was quantified to 6.5±0.5 ng/μl.

The SNaPshot™ reaction and signal readings

The extension primers used for the SNaPshot™ reaction were designed according to the manufacturer's recommendation. In addition, we used the Oligo Calculator Server (http://genestamp.ibms.sinica.edu.tw/genestamp/oligo-tm-calc.htm) and on-line oligonucleotide tools (http://www.rnature.com/oligonucleotide.html) to assess the secondary structure and T_m of each extension primer. To reduce unwanted hybridization and to maximize the perfect match in the SNapPshot™ reaction, 1.5–2% Hi-Di formamide (Applied Biosystems), 6% glycerol were added for a final reaction volume of 6.5 μl, which contained 3.2 μl of SNapPshot™ multiplex mix reagent, 0.010–0.015 pmol cleaned PCR products, and 0.3–0.4 μM extension primer. The cycling program was: 25 cycles at 96 °C for 10 s, 52 °C for 6 s, and 60 °C for 30 s. The temperature was then maintained at 60 °C for 60 s, followed by a rapid thermal ramp to 4 °C to prepare for post-extension. After cycling, the unincorporated fluorescent ddNTPs were cleaned up by adding 1 U SAP and incubating for 1 h at 37 °C, followed by 15 min at 75 °C for enzyme inactivation. Subsequently, the SNaPshot™ reaction product was diluted fourfold. Approximately 1 μl of diluted SNaPshot™ reaction product and 0.3 μl LIZ-120 (Applied Biosystems) were added to a mixed aliquot of 9 μl Hi-Di formamide based on the fluorescent strength shown on the ABI-3700 instrument. Samples were then loaded onto an ABI-3700 sequencer using GeneScan-120 LIZ™ (Applied Biosystems) as a molecular weight standard in the POP6 polymer. The signal intensities (peak heights) of the two alleles were measured by using GeneScan v3.5.2 software (Applied

Table 2
Preparation of DNA standard solutions with various allele frequencies for constructing standard curves

DNA standard solutions	Standard Curve 1 (SC1)	
m %	DNA stock with heterozygous allele (Mm[a])	DNA stock with homozygous allele (MM[b])
5%	1 μl	9 μl
10%	2 μl	8 μl
20%	4 μl	6 μl
30%	6 μl	4 μl
40%	8 μl	2 μl
45%	9 μl	1 μl
50%	10 μl	0 μl

DNA standard solutions were constructed by mixing different volumes of heterozygous and homozygous DNA stocks as described above. For each SNP, the DNA stocks was a mixture of equal amounts of 3 to 8 validated DNA samples. The concentration of each validated DNA sample was 6.5±0.5 ng/μl measured by PicroGreen method.

[a] Mm: heterozygote.

[b] MM: homozygote of major allele.

Biosystems). And then the relative proportion of the two peak heights was calculated.

Estimating allele frequencies using κ correction method

For each SNP, peak heights of the two alleles, major and minor, were expressed as H_{maj} and H_{min}, respectively. Heterozygous individuals provide convenient reference samples. The κ value of the Xth individual with heterozygote was expressed as 'κ_X', which is equal to H_{min}/H_{maj}. For each SNP, κ_m was the mean of n κ_X calculated from peak height ratios obtained from n heterozygous individuals (n=8–12). The adjusted estimates of allele frequency were calculated by the equation of $H_{\text{min in pool}}/(H_{\text{min in pool}}+\kappa_m \times H_{\text{maj in pool}})$, as described by Norton [10].

Constructing standard curves (SC) and estimating allele frequencies by applying the polynomial standard curve method (PSCM)

Two standard curves, standard curve 1 (SC1) and standard curve 2 (SC2), were constructed for each SNP locus with 7 designated AMiAFs. Table 2 depicts how the DNA standard solutions were constructed for SC1 of each SNP. In short, DNA standard solutions of various AMiAFs were constructed by mixing different volumes of homozygous and heterozygous stock solutions. The process for constructing SC2 was similar to that for SC1; however, only homozygous DNA stocks were used ('MM' and 'mm'). The PCR, SNaPshot™ reaction, signal measurements and the calculation of relative signal proportion were performed for each of those DNA standard solutions as described above.

Based on AMiAF levels and correspondent signal proportions of the 7 DNA standard solutions, a first- or a second-degree polynomial standard curve through the coordinate origins was selected via regression model fitting. The polynomial standard curves generated were then used to extrapolate the allele frequencies of loci in the pooled DNA samples.

Calculation of variance due to DNAs pool formation (σ^2_{pf}) and variance due to measurement errors (σ^2_m)

DNA samples of four SNPs (rs4844480, rs3774262, rs11761556 and rs1862513) with AMiAF of 40% were used to estimate the variance attributed to the pooling formation and to measurement error. There were triplicate pool formations for each SNP. And one PCR followed by one SNaPshot™ reaction was carried out for each pool. Subsequently, 3 repeated signal intensity readings for each SNaPshot™ reaction were measured. For each SNP, the variance components for the DNA pool formations (σ^2_{pf}) and for the measurements (σ^2_m) were computed from the 9 estimates in a

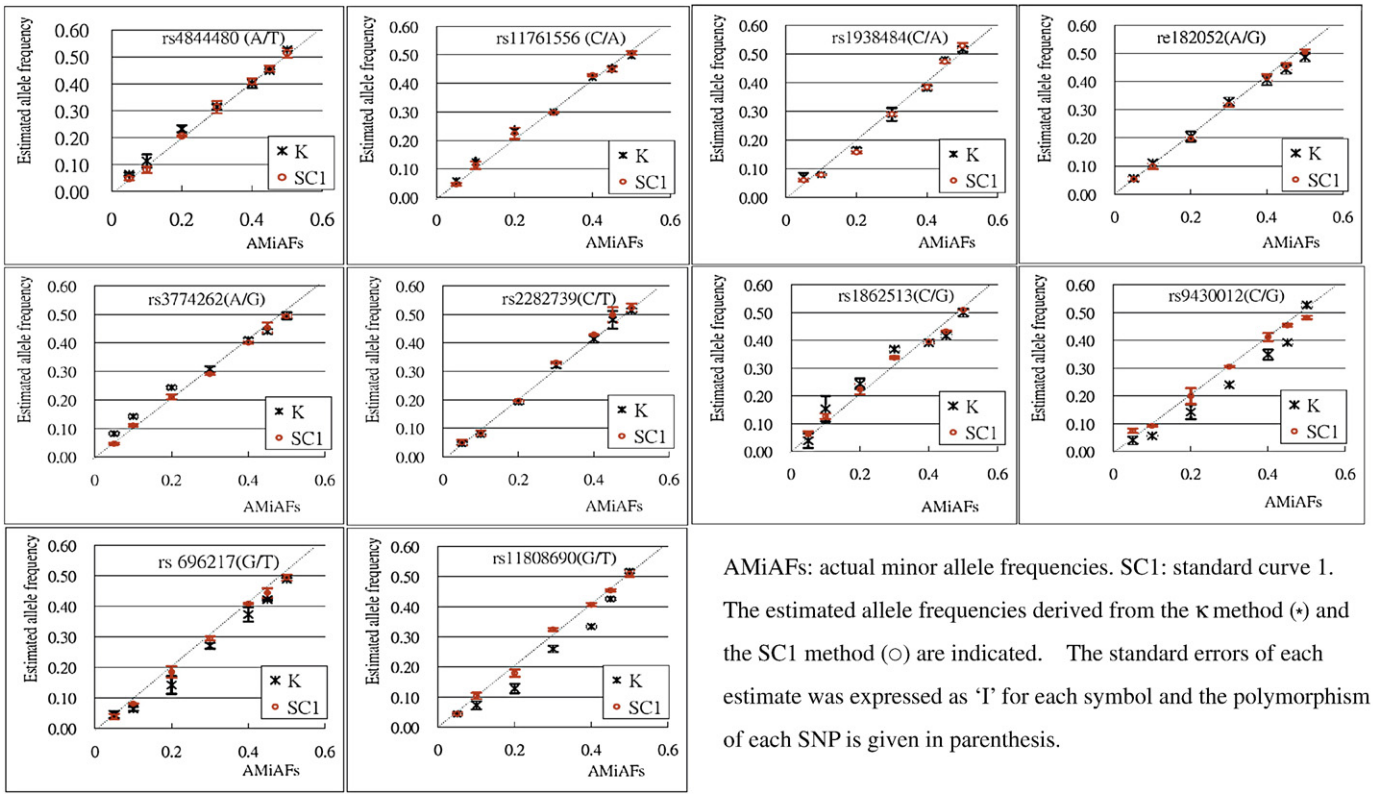

Fig. 1. Comparison of estimates derived from the polynomial standard curve and κ methods for 7 artificially pooled samples with the following actual minor allele frequencies: 5, 10, 20, 30, 40, 45, and 50%.

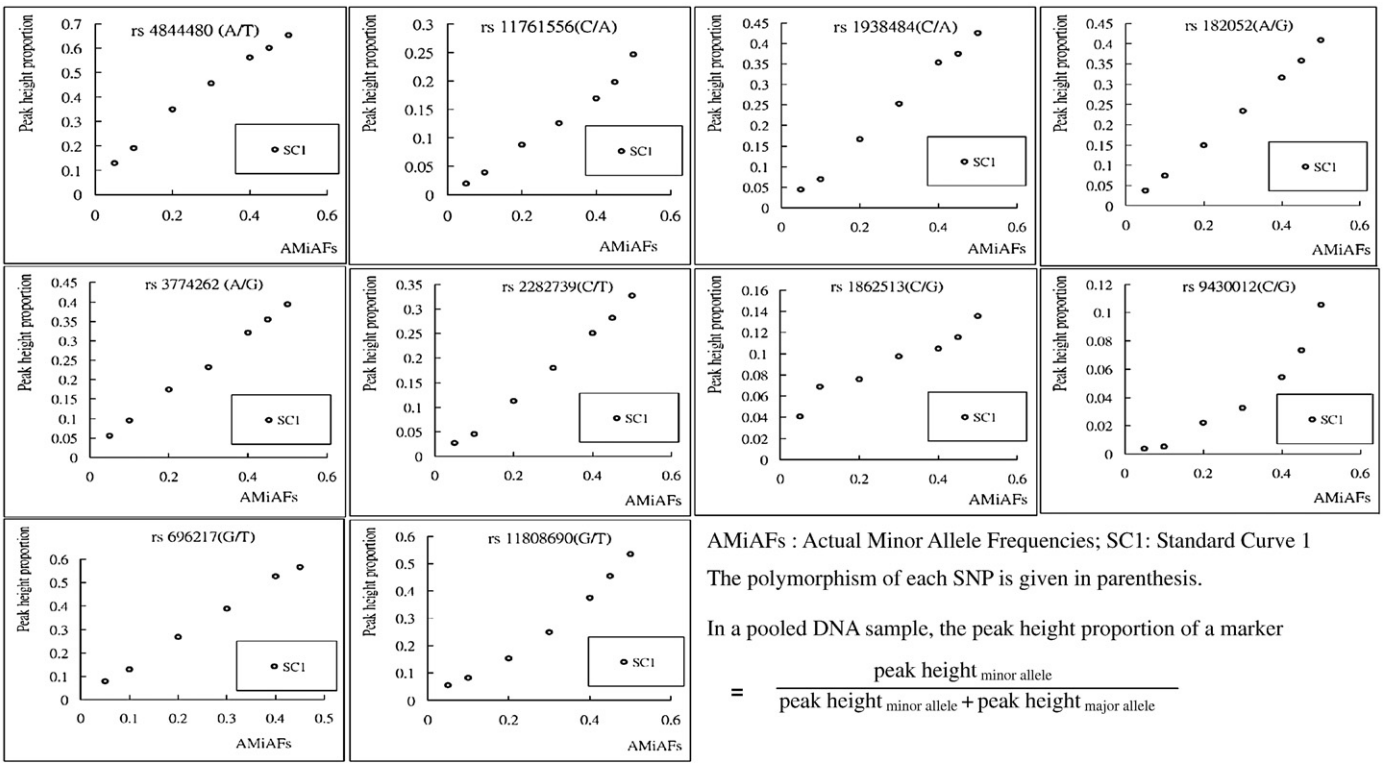

Fig. 2. The correlation between relative peak height proportions and actual minor allele frequencies of the 7 standard solutions in standard curve 1 (SC1).

hierarchical analysis of variance [25]. If there were M_m repeated allele frequency measurements for each PCR product and N_{pf} replicate pooling formations, denoting x_{ij} as the ith reading on each PCR product from the jth pool formation, then:

$$\sigma_m^2 = \frac{\sum(x_{ij}-\bar{x}_j)^2}{N_{pf}\times(M_m-1)}$$

$$\sigma_{pf}^2 = \frac{\left\{\frac{M_m\times\sum(\bar{x}_j-\bar{x})^2}{N_{pf}-1}-\sigma_m^2\right\}}{M_m}$$

Results

The relationship between AMiAFs and their relative signal intensity readings did not change whether the standard solutions contained heterozygous and homozygous DNA stocks (SC1) or homozygous stocks (SC2) (Appendix 2). The estimates of allele frequency obtained from SC1 were similar to those obtained from SC2 for each pooled DNA sample. Therefore, we only show the results obtained from SC1 in this article and refer to it as PSCM hereafter.

For each SNP-AMiAF, the peak heights of 2 alleles were measured and the relative peak height proportion was subsequently calculated for each of the triplicate DNA pools. Using the 3 peak height proportions, three estimates of allele frequency were obtained by applying the PSCM and κ correction methods, respectively (Fig. 1). In Fig. 1, the means and the standard errors of the means for the 3 estimated allele frequencies derived from either were plotted. For each of the six SNPs (rs4844480-A/T, rs11761556-A/C, rs1938484-A/C, rs182052-A/G, rs3774262-A/G and rs2282739-C/T), estimates of allele frequencies derived from the PSCM were similar to those from the κ correction method. For SNPs having the C/G polymorphism (rs1862513 and rs9430012), differences were observed between the two methods, particularly for frequencies in the range from 20 to 40% (Fig. 1). The largest discrepancy in allele frequency estimates was observed for rs9430012 at AMiAF 30%. The PSCM estimate was 30.4±0.38% and the κ method estimate was 23.9±1% (Fig. 1). A similar result was obtained for the two SNPs with the G/T polymorphism (rs696217 and rs11808690). For both the G/T and C/G polymorphisms, the PSCM was superior in both accuracy and precision compared with the κ method, particularly when the AMiAFs were at 20%, 30%, and 40%.

We plotted the peak height proportions versus AMiAFs, standard curve data for all the examined SNPs, in Fig. 2. There is an apparent curvilinear relationship between allelic peak height proportions and AMiAFs for the following 4 SNPs with C/G and G/T polymorphisms (rs1862513-C/G, rs9430012-C/G, rs696217-G/T, and rs11808690-G/T). These findings correlate with the poorer estimates for these SNPs obtained using the κ method for pooled DNA samples when compared with the PSCM which takes into consideration the curvilinear relationships.

The variances due to the DNA pool formation and measurement were analyzed and presented in Table 3. For each of the following 4 SNPs (rs4844480-A/T, rs3774262-A/G, rs11761556-A/C, and rs1862513-C/G), triplicate DNA pools were made and 3 repeated allele frequency measurements were subsequently performed for each pooled sample. The magnitude of the variances were in generally in the order of 10^{-4} to 10^{-5} for either pool formation or measurement.

Table 3
Estimation of variance due to pooling formation (σ^2_{pf}) and measurement errors (σ^2_m) for estimating allele frequencies of a pooled DNA sample with an actual minor allele frequency at 40%

SNP	Polymorphism	σ^2_{pf} [a]	σ^2_m [b]
rs4844480	A/T	6.05×10^{-5}	9.47×10^{-5}
rs3774262	A/G	7.69×10^{-5}	5.16×10^{-5}
rs11761556	A/C	3.93×10^{-5}	1.83×10^{-4}
rs1862513	C/G	3.76×10^{-5}	3.76×10^{-5}

[a] Triplicate pooled DNA samples were constructed for each SNP at actual minor allele frequencies of 40%.
[b] Three repeated measurements were performed for each pooled sample.

Discussion

With the SNaPshot™ platform, we have applied the principle of polynomial standard curves method (PSCM) to produce allele frequency estimates in pooled DNA samples and compared it with the κ method. The estimates derived from the PSCM were in general closer to AMiAFs than those from the κ method, particularly for C/G and G/T polymorphisms.

Specifically, our results showed an apparent discrepancy between the AMiAFs and the estimates determined using the κ method for pooled DNA samples with G/T and C/G polymorphisms, particularly when the AMiAFs were in the range of 20% to 40%. To date, there are relatively few studies to explore this phenomenon systematically. Previously, Gruber et al. [7] have also observed greater variability between estimated allele frequencies from pools and from individuals for certain SNPs with an actual minor allele frequency greater than 30%, but they did not clearly denote the polymorphic types of these SNPs. Furthermore, comprehensive analyses were not performed and no remedy was suggested. Our data provide an explanation to the above phenomenon, i.e., a curvilinear relationship exists between AMiAFs and signal intensity proportions for the SNPs with G/T and C/G polymorphisms (Fig. 2). To overcome this problem, we employed a second-degree polynomial standard curve model to relate signal intensity ratio with AMiAF in a comprehensive set of experiments with a total of 10 SNPs, including 6 polymorphic types of SNPs (A/C, A/T, C/T, A/G, C/G, and G/T polymorphisms) with AMiAFs ranging from 5 to 50%. Better performance was observed for SNPs with G/C and G/T polymorphisms than that for SNPs with other polymorphisms in pooled DNA samples using our approach than the κ method, especially when the AMiAFs ranged from 20 to 40%. To validate our proposed method, we genotyped the rs9430012 (C/G) for a DNA pool composed of 50 samples, using both the κ method and the PSCM and found that the estimated C-allele frequency derived from PSCM (22.3±2.3%) was closer to the real allele frequency (21.3%) than that from κ method (18.1±2.1%). Further research with a larger number of SNPs is needed to confirm our findings.

Although the DNA pooling technique has been applied in GWAS (Genome-Wide Association Studies) utilizing Affymetrix™ and Illumina Infinium I chips [17–21,28], cost is still an issue for fine-mapping candidate regions in the second-staged mapping of GWAS and for multiple candidate genes approach. Although the capillary electrophoresis system used by the SNaPshot™ may not be ideal in its throughput for individual genotyping for a study with large sample size and dense genetic markers [9], it is a substantially reasonable platform for researchers with limited budget to carry out candidate gene pooling research, since the per assay cost is relatively low compared to that of Pyrosequencing, GeneChip® and HPLC.

Early in our research, we observed a wide range of κ_X among individual heterozygotes using the SNaPshot™ reaction. Moskvina et al. [26] demonstrated that variations in κ values are large enough to result in unacceptable error rates when conducting association studies. Therefore, we modified the SNaPshot™ reaction by adding glycerol and Hi-Di formamide to the reaction, which lowered the variability of κ_X (data not shown). Our modified SNaPshot™ reaction is unique in providing a narrow range of κ_X for each SNP, which not only leads to more accurate but also more precise estimates of the allele frequencies.

For constructing the standard curve in PSCM, DNA standard solutions with various AMiAFs are usually prepared by combining different proportions of heterozygous and homozygous samples on hand. In general, there are more heterozygotes than homozygotes of minor alleles for most common SNPs in any give population. Our data indicated that the standard curves did not change regardless of whether standard solutions with varied AMiAFs were prepared from homozygote DNA stocks ('MM' and 'mm') or from a mixture of heterozygote and homozygote DNA stocks ('MM' and 'Mm'). We suggest that standard solutions can be mixed from DNA homozygote and heterozygote stock solutions as long as the AMiAFs are correct.

Conclusion

Our proposed method combines SNapPshot™ and the polynomial standard curve principle, and is conceptually simple and readily applicable without the need for complex statistical analysis. The PSCM method requires 7 more DNA pools than the κ correction method for constructing the standard curve for each SNP. It is important to consider the balance between accuracy and cost in estimating allele frequency for SNPs of G/C and G/T polymorphic types. To date, almost 9.3 million SNPs (including A/C, A/G, A/T, C/T, G/C and G/T polymorphisms) are included in the dbSNP database (http://www.ncbi.nlm.nih.gov/SNP). Among them, the percentage of G/C and G/T polymorphisms is greater than 18%. Our approach will facilitate to obtain more accurate results of these SNPs in association studies.

Acknowledgments

This study has been supported by Min-Sheng General Hospital and by Academia Sinica. The participation of all patients is greatly appreciated.

Appendix 1. Selected information for genotyped SNPs

SNP	Abbreviation of located gene	Chromosome location	Allele	Sequence of the primers
rs4844880	HSD11B1 (Hydroxysteroid (11-beta) dehydrogenase 1)	1q32.2	A/T	F:5′-TGGGCTACTCCCTTCATTATG-3′ R:5′-ACCATGCCTACTGATAGCCAT-3′ EXT:5′-ATGGAGAGCATTTCAATGGTGTTTTTATTC-3′
rs11761556	LEP (Leptin)	7q32.1	A/C	F:5′-TTTGAAGGAGGTGAGGGATGT-3′ R:5′-AAAGGCTTAAGACACCTCAGC-3′ EXT:5′-AGAGAAAGAAGAGACAGGAGGG-3′
rs1938484	LEPR (Leptin receptor)	1p31.2	A/C	F:5′-AATTCTGGGGTGCCATAGACA-3′ R:5′-TCCTCATCTAAAAAGCAAACA-3′ EXT:5′-CTTCACAGCTCTTGTTGTTGCTTTT-3′
rs182052	ADIPOQ (Adiponectin)	3q27.3	A/G	F:5′-ATAGCCTCTGGCTGGGATCA-3′ R:5′-TTGACTTTTCTGAAGCTGCC-3′ EXT:5′-GCATGGAACCATTCTGAATTTT-3′
rs3774262	ADIPOQ (Adiponectin)	3q27.3	A/G	F:5′-AATAGAGGAGGAGAGACATCCTAGA-3′ R:5′-GGACCAATAAGACCTAAGGAATGAG-3′ EXT:5′-CTGTGGGAGATATAGAAGGA-3′
rs2282739	HSD11B1 (Hydroxysteroid (11-beta) dehydrogenase 1)	1q32.2	C/T	F:5′-TTTCTTGTATGTGTAACATCCCG-3′ R:5′-TTTTAGCATGGGTGATGTGG-3′ EXT:5′-TCCTAGAGTGCTTGTTTACA-3′
rs1862513	RETN (Resistin)	19p13.2	C/G	F:5′-CAAATCCGGCACACGAATT-3′ R:5′-TCACTGTAGCTTCGAACTCCC-3′ EXT:5′-ACTTCCAACAGGGCCTCC-3′

Appendix 1 (*continued*)

SNP	Abbreviation of located gene	Chromosome location	Allele	Sequence of the primers
rs9430012	HSD11B1 (Hydroxysteroid (11-beta) dehydrogenase 1)	1q32.2	C/G	F:5′-TGAACGTGACCTTAAGTGGA-3′ R:5′-CCCATAAATAAACAGGGCTAG-3′ EXT:5′-GAAGGATTCCCTGCTATAGGTTTTG-3′
rs696217	GHRL (Ghrelin)	3p25.3	G/T	F:5′-GTTCTGGGGAGCTTGTAGTTG-3′ R:5′-ACCTCACTGTTTCTGGAAGGA-3′ EXT:5′-GCATGGAACCATTCTGAATTTT-3′
rs11808690	HSD11B1 (Hydroxysteroid (11-beta) dehydrogenase 1)	1q32.2	G/T	F:5′-ATGCAACTCCAGTCATTTGC-3′ R:5′-TGTGATATTCTGTTACATCCATAGA-3′ EXT:5′-GGGAAGAATGTCAGCGTGT-3′

F, R, and EXT: forward, reverse and extension PCR primers, respectively.

Appendix 2. Comparison of the accuracy of the three calculation models in estimating allele frequencies and standard deviations in pooled DNA samples

		0.05	0.10	0.20	0.30	0.40	0.45	0.50
rs4844480 A/T	SC 1	0.047 (0.0069)	0.080 (0.0128)	0.210 (0.0033)	0.313 (0.0220)	0.408 (0.0117)	0.455 (0.0120)	0.515 (0.0170)
	SC 2	0.047 (0.0034)	0.085 (0.0066)	0.223 (0.0036)	0.324 (0.0158)	0.401 (0.0208)	0.451 (0.0031)	0.499 (0.0200)
	κ	0.063 (0.0059)	0.112 (0.0242)	0.232 (0.0125)	0.314 (0.0096)	0.401 (0.0178)	0.449 (0.0044)	0.527 (0.0050)
rs11761556 A/C	SC 1	0.04 (0.0047)	0.112 (0.0132)	0.224 (0.0196)	0.297 (0.0022)	0.428 (0.0035)	0.449 (0.0091)	0.505 (0.0059)
	SC 2	0.039 (0.0045)	0.113 (0.0127)	0.221 (0.0192)	0.294 (0.0022)	0.428 (0.0037)	0.456 (0.0052)	0.512 (0.0066)
	κ	0.058 (0.0048)	0.124 (0.0130)	0.231 (0.0184)	0.299 (0.0020)	0.420 (0.0033)	0.451 (0.0138)	0.496 (0.0060)
rs1938484 A/C	SC 1	0.060 (0.0048)	0.078 (0.0010)	0.157 (0.0047)	0.290 (0.0065)	0.383 (0.0085)	0.472 (0.0056)	0.518 (0.0098)
	SC 2	0.060 (0.0025)	0.084 (0.0022)	0.157 (0.0045)	0.278 (0.0034)	0.381 (0.0112)	0.476 (0.0062)	0.512 (0.0098)
	κ	0.070 (0.0148)	0.079 (0.0014)	0.16 (0.0054)	0.289 (0.0227)	0.380 (0.0073)	0.480 (0.0018)	0.516 (0.0092)
rs182052 A/G	SC 1	0.058 (0.0013)	0.097 (0.0138)	0.198 (0.0058)	0.318 (0.0071)	0.419 (0.0167)	0.459 (0.0046)	0.512 (0.0230)
	SC 2	0.050 (0.0107)	0.107 (0.0129)	0.197 (0.0218)	0.332 (0.0025)	0.398 (0.0384)	0.453 (0.0118)	0.489 (0.0073)
	κ	0.055 (0.0092)	0.111 (0.0124)	0.204 (0.0171)	0.328 (0.0137)	0.420 (0.0192)	0.443 (0.0146)	0.486 (0.0147)
rs3774262 A/G	SC 1	0.046 (0.0037)	0.111 (0.0042)	0.210 (0.0084)	0.291 (0.0040)	0.410 (0.004)	0.453 (0.0176)	0.493 (0.0048)
	SC 2	0.054 (0.0024)	0.100 (0.0047)	0.206 (0.0050)	0.289 (0.0056)	0.390 (0.0065)	0.443 (0.0011)	0.503 (0.0143)
	κ	0.082 (0.0033)	0.142 (0.0035)	0.243 (0.0046)	0.307 (0.0097)	0.418 (0.0055)	0.440 (0.0080)	0.495 (0.0121)
rs2282739 C/T	SC 1	0.050 (0.0092)	0.082 (0.0090)	0.197 (0.0025)	0.330 (0.0040)	0.428 (0.0003)	0.498 (0.0261)	0.524 (0.0138)
	SC 2	0.049 (0.0020)	0.082 (0.0011)	0.187 (0.0019)	0.315 (0.0048)	0.413 (0.0071)	0.485 (0.0323)	0.520 (0.0085)
	κ	0.047 (0.0046)	0.078 (0.0012)	0.192 (0.0062)	0.320 (0.0090)	0.413 (0.0116)	0.480 (0.0310)	0.512 (0.0072)
rs1862513 C/G	SC 1	0.061 (0.0107)	0.126 (0.0101)	0.223 (0.0182)	0.337 (0.0042)	0.395 (0.0064)	0.432 (0.0026)	0.509 (0.0012)
	SC 2	0.049 (0.0062)	0.119 (0.0008)	0.241 (0.0062)	0.347 (0.0049)	0.403 (0.0031)	0.441 (0.0083)	0.520 (0.0127)
	κ	0.038 (0.0259)	0.153 (0.0460)	0.243 (0.019)	0.367 (0.0058)	0.391 (0.0035)	0.415 (0.0001)	0.500 (0.0128)
rs9430012 C/G	SC 1	0.073 (0.0076)	0.092 (0.0030)	0.199 (0.0287)	0.304 (0.0038)	0.421 (0.0144)	0.453 (0.0048)	0.481 (0.0058)
	SC 2	0.054 (0.0068)	0.084 (0.0031)	0.195 (0.0297)	0.320 (0.0119)	0.421 (0.0070)	0.457 (0.0011)	0.478 (0.0129)
	κ	0.040 (0.0125)	0.056 (0.0010)	0.141 (0.0159)	0.239 (0.0100)	0.348 (0.0077)	0.392 (0.0063)	0.527 (0.0070)
rs696217 G/T	SC 1	0.039 (0.0083)	0.081 (0.0003)	0.185 (0.0184)	0.295 (0.0067)	0.409 (0.0025)	0.444 (0.0152)	0.494 (0.0097)
	SC 2	0.038 (0.0025)	0.083 (0.0059)	0.203 (0.0035)	0.314 (0.0053)	0.407 (0.0215)	0.465 (0.0069)	0.517 (0.0132)
	κ	0.046 (0.0109)	0.064 (0.0062)	0.142 (0.0289)	0.271 (0.0097)	0.373 (0.0229)	0.422 (0.0028)	0.489 (0.0042)
rs11808690 G/T	SC 1	0.066 (0.0039)	0.105 (0.0006)	0.180 (0.0122)	0.324 (0.0047)	0.391 (0.0051)	0.454 (0.0034)	0.512 (0.0099)
	SC 2	0.066 (0.0039)	0.105 (0.0060)	0.177 (0.0073)	0.311 (0.0050)	0.420 (0.028)	0.457 (0.0020)	0.492 (0.0248)
	κ	0.043 (0.0005)	0.072 (0.0031)	0.129 (0.0153)	0.259 (0.0097)	0.334 (0.0030)	0.426 (0.0037)	0.517 (0.0055)

The three calculation models: Standard Curve 1 (SC1), Standard Curve 2 (SC2) and correction factorκ. Standard deviation (SD) is presented in parenthesis.
All of the estimates presented on this table are the average of three estimates for allele frequency from three independent pools, except that rs1862513 and rs696217 are the average of two estimates for allele frequency from two independent pools.

References

[1] J. Bader, A. Bansal, P. Sham, Efficient SNP-based tests of association for quantitative phenotypes using pooled DNA, GeneScreen 1 (2001) 143–150.

[2] S. Curran, L. Hill, G. O'Grady, D. Turic, P. Asherson, E. Taylor, P. Sham, I. Craig, P. Vaughan, Validation of single nucleotide polymorphism quantification in pooled DNA samples with SNaPIT. A glycosylase-mediated methods for polymorphism detection method, Mol. Biotechnol. 22 (2002) 253–262.

[3] P. Pacek, A. Sajantila, A.C. Syvanen, Determination of allele frequencies at loci with length polymorphism by quantitative analysis of DNA amplified from pooled samples, PCR Methods Appl. 2 (1993) 313–317.

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

[5] P. Sham, J.S. Bader, I. Craig, M. O'Donovan, M. Owen, DNA Pooling: a tool for large-scale association studies, Nat. Rev. Genet. 3 (2002) 862–871.

[6] P.M. Visscher, S. Le Hellard, Simple method to analyze SNP-based association studies using DNA pools, Genet. Epidemiol. 24 (2003) 291–296.

[7] J.D. Gruber, P.B. Colligan, J.K. Wolford, Estimation of single nucleotide polymorphism allele frequency in DNA pools by using Pyrosequencing, Hum. Genet. 110 (2002) 395–401.

[8] G. Kirov, N. Williams, P. Sham, N. Craddock, M.J. Owen, Pooled genotyping of microsatellite markers in parent–offspring trios, Genome Res. 10 (2000) 105–115.

[9] S. Le Hellard, S.J. Ballereau, P.M. Visscher, H.S. Torrance, J. Pinson, S.W. Morris, M.L. Thomson, C.A. Semple, W.J. Muir, D.H. Blackwood, D.J. Porteous, K.L. Evans, SNP genotyping on pooled DNAs: comparison of genotyping technologies and a semi automated method for data storage and analysis, Nucleic Acids Res. 30 (2002) e74.

[10] N. Norton, N.M. Williams, M.C. O'Donovan, M.J. Owen, DNA pooling as a tool for large-scale association studies in complex traits, Ann. Med. 36 (2004) 146–152.

[11] N. Norton, N.M. Williams, H.J. Williams, G. Spurlock, G. Kirov, D.W. Morris, B. Hoogendoorn, M.J. Owen, M.C. O'Donovan, Universal, robust, highly quantitative SNP allele frequency measurement in DNA pools, Hum. Genet. 110 (2002) 471–478.

[12] Y. Ohnishi, T. Tanaka, K. Ozaki, R. Yamada, H. Suzuki, Y. Nakamura, A high-throughput SNP typing system for genome-wide association studies, J. Hum. Genet. 46 (2001) 471–477.

[13] G. Zhou, M. Kamahori, K. Okano, G. Chuan, K. Harada, H. Kambara, 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 (2001) E93.

[14] W. Han, D.H. Lou, J. Wang, S. Yip, M. Yap, Application of DHPLC in human genomic SNP screening and genotyping for all-trans-retinol dehydrogenase, Zhejiang Da Xue Xue Bao Yi Xue Ban 34 (2005) 248–254.

[15] H. Tian, D.M. Liu, B. Xu, W.Y. Zheng, S.Y. Zhou, Analysis of sequence-tagged site in bcr and abl genes by DNA pooling and dHPLC, Zhongguo Shi Yan Xue Ye Xue Za Zhi 13 (2005) 468–471.

[16] B. Hoogendoorn, N. Norton, G. Kirov, N. Williams, M.L. Hamshere, G. Spurlock, J. Austin, M.K. Stephens, P.R. Buckland, M.J. Owen, M.C. O'Donovan, Cheap, accurate and rapid allele frequency estimation of single nucleotide polymorphisms by primer extension and DHPLC in DNA pools, Hum. Genet. 107 (2000) 488–493.

[17] J. Brohede, R. Dunne, J.D. McKay, G.N. Hannan, PPC: an algorithm for accurate estimation of SNP allele frequencies in small equiMol.ar pools of DNA using data from high density microarrays, Nucleic Acids Res. 33 (2005) e142.
[18] S. Macgregor, P.M. Visscher, G. Montgomery, Analysis of pooled DNA samples on high density arrays without prior knowledge of differential hybridization rates, Nucleic Acids Res. 34 (2006) e55.
[19] D.W. Craig, M.J. Huentelman, D. Hu-Lince, V.L. Zismann, M.C. Kruer, A.M. Lee, E.G. Puffenberger, J.M. Pearson, D.A. Stephan, Identification of disease causing loci using an array-based genotyping approach on pooled DNA, BMC Genomics 6 (2005) 138.
[20] G. Kirov, I. Nikolov, L. Georgieva, V. Moskvina, M.J. Owen, M.C. O'Donovan, Pooled DNA genotyping on Affymetrix SNP genotyping arrays, BMC Genomics 7 (2006) 27.
[21] A.E. Baum, N. Akula, I. Cardona, W. Corona, A. Singleton, J. Hardy, S. Detera-Wadleigh, F.J. McMahon, DNA pooling for whole genome association studies on the Illumina Infinium Assay arrays, 56th American Society of Human Genetics Annual Meeting (2006).
[22] H.C. Yang, C.C. Pan, C.Y. Lin, C.S. Fann, PDA: pooled DNA analyzer, BMC Bioinformatics 7 (2006) 233.
[23] G. Zou, H. Zhao, The impacts of errors in individual genotyping and DNA pooling on association studies, Genet. Epidemiol. 26 (2004) 1–10.
[24] C.L. Simpson, J. Knight, L.M. Butcher, V.K. Hansen, E. Meaburn, L.C. Schalkwyk, I.W. Craig, J.F. Powell, P.C. Sham, A. Al-Chalabi, A central resource for accurate allele frequency estimation from pooled DNA genotyped on DNA microarrays, Nucleic Acids Res. 33 (2005) e25.
[25] B.J. Barratt, F. Payne, H.E. Rance, S. Nutland, J.A. Todd, D.G. Clayton, Identification of the sources of error in allele frequency estimations from pooled DNA indicates an optimal experimental design, Ann. Hum. Genet. 66 (2002) 393–405.
[26] V. Moskvina, N. Norton, N. Williams, P. Holmans, M. Owen, M. O'Donovan, Streamlined analysis of pooled genotype data in SNP-based association studies, Genet. Epidemiol. 28 (2005) 273–282.
[27] S. Macgregor, Most pooling variation in array-based DNA pooling is attributable to array error rather than pool construction error, Eur. J. Hum. Genet. 15 (2007) 501–504.
[28] H.C. Yang, Y.J. Liang, M.C. Huang, L.H. Li, C.H. Lin, J.Y. Wu, Y.T. Chen, C.S. Fann, A genome-wide study of preferential amplification/hybridization in micro-array-based pooled DNA experiments, Nucleic Acids Res. 34 (2006) e106.