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 Institute of Health Sciences, Chang Jung Christian University, Taiwan, ROC

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.

© 2008 Elsevier Inc. All rights reserved.

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) [7,17]. 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 detection rate 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.