Double genomic control is not effective to correct for population stratification in meta-analysis for genome-wide association studies

Shudong Wang1,2†, Wenan Chen1,†, Xiangning Chen3, Fengjiao Hu1, Kellie J. Archer1, Nianjun Liu4, Shumei Sun1 and Guimin Gao1*  

1 Department of Biostatistics, School of Medicine, Virginia Commonwealth University, Richmond, VA, USA  
2 College of Information Science and Engineering, Shandong University of Science and Technology, Qingdao, China  
3 Virginia Institute for Psychiatric and Behavioral Genetics, School of Medicine, Virginia Commonwealth University, Richmond, VA, USA  
4 Department of Biostatistics, University of Alabama, Birmingham, AL, USA

*Correspondence: Guimin Gao, Department of Biostatistics, School of Medicine, Virginia Commonwealth University, One Capitol Square, Room 710, 800 East Main Street, P.O. Box 980032, Richmond, VA 23298-0032, USA. e-mail: ggao3@vcu.edu

† Shudong Wang and Wenan Chen have contributed equally to this work.

Meta-analysis of genome-wide association studies (GWAS) has become a useful tool to identify genetic variants that are associated with complex human diseases. To control spurious associations between genetic variants and disease that are caused by population stratification, double genomic control (GC) correction for population stratification in meta-analysis for GWAS has been implemented in the software METAL and GWAMA and is widely used by investigators. In this research, we conducted extensive simulation studies to evaluate the double GC correction method in meta-analysis and compared the performance of the double GC correction with that of a principal components analysis (PCA) correction method in meta-analysis. Results show that when the data consist of population stratification, using double GC correction method can inflate type I error rates at a marker with significant allele frequency differentiation in the subpopulations (such as caused by recent strong selection). On the other hand, the PCA correction method can control type I error rates well and has much higher power in meta-analysis compared to the double GC correction method, even though in the situation that the casual marker does not have significant allele frequency difference between the subpopulations. We applied the double GC correction and PCA correction to meta-analysis of GWAS for two real datasets from the Atherosclerosis Risk in Communities (ARIC) project and the Multi-Ethnic Study of Atherosclerosis (MESA) project. The results also suggest that PCA correction is more effective than the double GC correction in meta-analysis.

Keywords: genome-wide association studies, meta-analysis, double genomic control correction, principal components analysis, population stratification

INTRODUCTION

Genome-wide association studies (GWAS) are an important approach for identifying genetic variants associated with complex human diseases. Recently, meta-analysis of GWAS has been used to obtain collective evidence from the multiple GWAS studies (Lohmueller et al., 2003; Houlston et al., 2008; Zeggini et al., 2008; Lindgren et al., 2009; Lin and Zeng, 2010; Stahl et al., 2010; Willer et al., 2010; Nalls et al., 2011; Qayyum et al., 2012). Unfortunately, population stratification in the studied samples can lead to spurious associations in disease studies (Cardon and Palmer, 2003; Freedman et al., 2004; Marchini et al., 2004; Price et al., 2006). To control the spurious associations caused by population stratification in meta-analysis of GWAS, genomic control (GC) correction within each study (which is referred to as single GC correction) has been used (Devlin and Roeder, 1999; Devlin et al., 2001; Reich and Goldstein, 2001; Devlin et al., 2004; Lindgren et al., 2009). However, the GC correction method cannot effectively control false positive rates and may lead to a loss in power (Price et al., 2006; Mägi and Morris, 2010; Willer et al., 2010). Therefore, a double GC correction method has been proposed and implemented in the widely used meta-analysis software METAL and GWAMA (Mägi and Morris, 2010; Willer et al., 2010). The double GC correction method adjusts the set of test statistics across all markers within each study by a GC inflation factor, calculates a combined statistic across studies at each marker, and then adjusts all combined statistics across the genome by the corresponding GC inflation factor. The double GC correction method has been used by many investigators in their meta-analyses (e.g., Lindgren et al., 2009; Mägi and Morris, 2010; Willer et al., 2010; Lettre et al., 2011). To adjust for stratification in meta-analysis of GWAS, another popular approach is the principal component analysis (PCA) correction method that adjusts for stratification by top principal components (PCs) of genotype data within each study (Price et al., 2006; Wang et al., 2009; Qayyum et al., 2012).

In this research, we compared the single GC correction, the double GC correction and PCA correction in meta-analysis by simulation studies and applied these methods to a meta-analysis of two real data sets from the Atherosclerosis Risk in Communities (ARIC) project and the Multi-Ethnic Study of Atherosclerosis (MESA) project. Results from simulations and real data analysis...
suggest that when population stratification exists, using double GC correction can have inflated false positive error rates at markers with significant allele frequency differentiation in the subpopulations (such as caused by recent strong selection), and can have lower power than using the PCA method for stratification correction in meta-analysis.

### MATERIALS AND METHODS

We consider $K$ case–control studies with $n_k$ individuals in the $k$th study ($k = 1, \ldots, K$) and $N$ single nucleotide polymorphisms (SNPs) in each study. In the $k$th study, let $Y_k$ denote the disease status ($1 = \text{disease}, 0 = \text{no disease}$) of the $i$th individual, let $g_{ik}$ denote the additive genotype coding value (the count of reference alleles) at the $j$th SNP of the $i$th individual, and let $X_{ik} = (x_{1k}, x_{2k}, \ldots, x_{nj})^T$ denote a vector of $s$ covariates. In this research, we focus on fixed-effect meta-analysis, i.e., we assume that the allelic effects at a test marker are the same across all studies. In the $k$th study, we use the following logistic regression model

$$
\log \text{Pr}(Y_k = 1) = a_k + q_k^T X_{ik} + \beta_k g_{ik},
$$

where $a_k$ is the study-specific intercept, $q_k$ is the coefficient vector corresponding to covariate vector $X_{ik}$, and $\beta_k$ denote the genetic effect (log odds ratio) of the reference allele at the $j$th SNP in the $k$th study. We can estimate $\beta_k$ and $\text{var}(\beta_k)$ at the $j$th SNP based on the logistic model (Eq. 1). Thus, the combined allelic effect $B_j$ at the $j$th SNP across $K$ studies, can be calculated as

$$
B_j = \frac{\sum_{k=1}^K w_k g_{ik} \beta_k}{\sum_{k=1}^K w_k},
$$

where $w_k$ is a weight for the genetic effect $\beta_k$, and $w_k = \frac{1}{\text{var}(\beta_k)}$ is the inverse of the variance of the estimated allelic effect in the $k$th study. The variance of $B_j$ is estimated by $V_j = \left(\sum_n w_n \right)^{-1}$ and its estimated standard error $SE_j$ is the square root of variance $V_j$, i.e., $SE_j = \sqrt{V_j}$. Therefore, the combined statistic $\chi^2_j = B_j^2 / V_j$ has an approximate $\chi^2$ distribution with one degree of freedom (Mägi and Morris, 2010; Willer et al., 2010).

### GC CORRECTION IN META-ANALYSIS

To adjust for population stratification, in the $k$th study, we can calculate a corrected weight at the $j$th SNP in the $k$th study as $w_{jk} = w_k / \lambda_k$ (see also Section “Materials and Methods” for the calculation of $w_k$) and calculate the combined allelic effect $B_j$ by using $w_{jk}$ to replace $w_k$ in Eq. 2 correspondingly, the variance of $B_j$ can be calculated as $V_j = \left(\sum_n w_{nk} \right)^{-1}$. The combined statistic $\chi^2_j = B_j^2 / V_j$ can be calculated for meta-analysis ($j = 1, 2, \ldots, N$).

### DOUBLE GC CORRECTION IN META-ANALYSIS

After GC correction in each individual study as described above, the combined statistic $\chi^2_j (j = 1, \ldots, N)$ may still deviate from the $\chi^2$ distribution which is caused by population stratification. Therefore, investigators proposed to further adjust the combined statistics $\chi^2_j$ by the corresponding GC inflation factor $\lambda$, which is the median of the combined statistics $\chi^2_j$ divided by 0.455 (Mägi and Morris, 2010; Willer et al., 2010). This process corrects for stratification twice and is called double GC correction.

### PCA CORRECTION IN META-ANALYSIS

Price et al. (2006) proposed to correct for stratification by PCA in GWAS analysis. For meta-analysis of GWAS with $K$ ($K > 1$) case–control studies of population stratification, PCA has been used to correct for stratification within each of the $K$ studies (Qayyum et al., 2012): for each of the $K$ studies, we can calculate the PCs (Jackson, 2005) of genome-wide genotype values for each individual and use the top 10 PCs as covariates, as suggested in the literature (Price et al., 2006; Liu et al., 2011; Qayyum et al., 2012), to correct for population stratification by incorporating these covariates (top 10 PCs) into the logistic regression model (Eq. 1). Meta-analysis for the $K$ studies can be conducted based on this modified model.

### SIMULATION STUDIES

#### DATA SIMULATION

To evaluate the performance of the single GC correction, double GC correction, and PCA correction in meta-analysis of data with population stratification, we simulated datasets on $K$ case–control studies in a similar way to those described in Pritchard and Donnelly (2001) and Price et al. (2006). Individuals in each of the $K$ case–control studies were sampled from two populations. For each dataset we simulated 100,000 independent random SNPs. To generate each of these SNPs, we used the Balding–Nichols model with $F_{ST} = 0.01$ (Balding and Nichols, 1995) to generate allele frequencies $f_j$ for populations $s = 1, 2$ which was drawn from a beta distribution with parameters $f_j (1 − F_{ST}) / F_{ST}$ and $(1 − f_j) (1 − F_{ST}) / F_{ST}$. $F_{ST} = 0.01$ usually leads to allele frequency differences under 0.10 for typical common SNPs (Price et al., 2006). At each of the random SNPs, we assumed Hardy–Weinberg equilibrium in each population, and individuals from populations were assigned genotypic values of 0, 1 or 2 with probabilities $(1 − f_j)^2$, $2f_j (1 − f_j)$, or $f_j^2$, respectively. For the PCA correction, we used these 100,000 random SNPs to calculate the first 10 PCs, which were used as covariates in the analysis.

#### Meta-analysis with two case–control studies ($K = 2$)

We first simulated two case–control studies. In each study, 60% of the cases and 40% of the controls were sampled from population 1 and the remaining cases and controls from population 2.

#### Meta-analysis with $K = 5$ case–control studies

To evaluate the performance of the three correction methods in meta-analysis with $K$ ($K > 2$) studies, we simulated datasets on five case–control studies. Each of the five studies had 1080 cases and 1000 controls. For each case–control study, a proportion of cases and a proportion of controls were sampled from population 1, and...
Table 1 | Type I error rates of meta-analysis with two case–control studies for different allele frequencies in the two subpopulations.

| Frequency | Study 1 | Study 2 | 10^-6 | 30^-4 |
|-----------|---------|---------|-------|-------|
| Population (1) | Case | Control | Case | Control | PCA | 1 GC | 2 GC | PCA | 1 GC | 2 GC |
| Random(2) | 1000 | 1000 | 1000 | 1000 | 9.51e-6 | 5.05e-6 | 5.34e-6 | 1.07e-6 | 3.90e-7 | 3.80e-7 |
| 1500 | 9.88e-6 | 5.03e-6 | 5.30e-6 | 9.19e-7 | 3.90e-7 | 4.60e-7 |
| 1500 | 1500 | 1500 | 750 | 1.04e-5 | 5.08e-6 | 5.50e-6 | 8.50e-7 | 5.30e-7 | 5.70e-7 |
| 0.4 (0.2) | 1000 | 1000 | 1000 | 1000 | 9.89e-6 | 1.25e-2 | 3.14e-2 | 9.20e-7 | 1.76e-3 | 1.86e-3 |
| 1500 | 1500 | 500 | 500 | 1.01e-5 | 8.44e-3 | 8.49e-3 | 1.01e-6 | 9.77e-4 | 1.06e-3 |
| 750 | 1500 | 1500 | 750 | 1.02e-5 | 1.26e-2 | 1.00e-2 | 1.00e-6 | 1.76e-3 | 1.86e-3 |
| 0.6 (0.2) | 1000 | 1000 | 1000 | 1000 | 9.67e-6 | 0.83e-6 | 0.83e-6 | 7.00e-7 | 0.58e-6 | 0.59e-6 |
| 1500 | 1500 | 500 | 500 | 9.43e-6 | 0.76e-6 | 0.78e-6 | 8.60e-7 | 0.48e-6 | 0.48e-6 |
| 750 | 1500 | 1500 | 750 | 9.88e-6 | 0.83e-6 | 0.84e-6 | 9.30e-7 | 0.58e-6 | 0.58e-6 |
| 0.8 (0.2) | 1000 | 1000 | 1000 | 1000 | 9.98e-6 | 1.00 | 1.00 | 1.14e-6 | 1.00 | 1.00 |
| 1500 | 1500 | 500 | 500 | 9.61e-6 | 1.00 | 1.00 | 9.50e-7 | 1.00 | 1.00 |
| 750 | 1500 | 1500 | 750 | 9.69e-6 | 1.00 | 1.00 | 9.20e-7 | 1.00 | 1.00 |

1. Allele frequencies in the two subpopulations from which cases and controls were sampled for each study.
2. Sample size.
3. Allele frequencies in the two populations of each study were randomly generated.
4. 0.576 = 5.01 × 10^-6

Type I error rates for meta-analysis with two case–control studies

Table 2 | Type I error rates for meta-analysis with five case–control studies

| Population (2) | Type I error rates for meta-analysis with five case–control studies (K = 5) |
|----------------|--------------------------------------------------------------------------|
|                | Type I error rates for meta-analysis with two case–control studies described above. |

We estimated the type I error rates for meta-analysis with K = 5 case-control studies. The detailed results are shown in Table 2. We can see the type I error rates had similar pattern to that for meta-analysis with two case–control studies described above.

POWER EVALUATION

To evaluate power of meta-analysis using the three stratification correction methods, we simulated 10,000 replicated datasets. Each data set contained 100,000 independent random SNPs (described above) and a causal SNP. We simulated the causal SNP by a similar method of Price et al. (2006). We used the Balding–Nichols method.

the remaining cases and controls were sampled from the population 2; in the first study, 55% of the cases and 45% of the controls were sampled from population 1; in the second study, 70% of the cases and 30% of the controls; in the third study, 60% of the cases and 40% of the controls.

TYPE I ERROR RATE EVALUATION

To verify the effectiveness of the three stratification correction methods in meta-analysis, we simulated replicated data sets. Each data set contained 100,000 independent random SNPs (as described above) and an additional test SNP (which is not associated with the disease). We simulated the test SNP by the following four scenarios: in the first scenario, the test SNP was simulated in each individual by the same way as that for the random SNPs described above using the Balding–Nichols model with θ = 0.01; for the second, third, and fourth scenarios, we assumed that the test SNP had significant different allele frequencies in populations 1 and 2, which could be caused by recent strong selection (see also Price et al., 2006). In scenario 2, the frequencies in populations 1 and 2 were 0.4 and 0.2, respectively; in scenario 3, frequencies were 0.6 and 0.2; in scenario 4, frequencies were 0.8 and 0.2. For each scenario, we computed type I error rates of meta-analysis using the three correction methods (PCA, single GC correction, and double GC correction) separately, under a set of combination of parameters (numbers of cases and controls in the two studies). For each combination of these parameters, we simulated 10^6 replicated data sets to estimate the type I error rates at significance levels of 10^-5 and 10^-4. We did not use the significance level of 10^-3 because estimating accurate type I error rates at this level requires at least 10^8 replicated data sets. This is computationally intensive.
model with $F_{ST} = 0.01$ again to generate allele frequencies for populations 1 and 2. We generated genotypes at the causal SNP by using a multiplicative disease risk model with a relative risk above. Case individuals were assigned genotypic values 0, 1, or 2 respectively, where

$$\text{genotype coding value at each SNP was defined as the count of the reference allele.}$$

In meta-analyses of GWAS for the two real data sets, we used PCA correction, single GC correction, and double GC correction for stratification, separately. The top-eight SNPs with $p$-values $<10^{-5}$ in at least one of the three meta-analysis methods are displayed in Table 5. At each of the eight SNPs, meta-analysis with PCA correction had much smaller $p$-value than that with double GC. This may indicate that meta-analysis with double GC correction for stratification is not as effective as that with PCA.

**DISCUSSION**

In this research, we evaluated the performance of three methods for correcting for population stratification (PCA correction, single GC correction, and double GC correction) in meta-analysis of GWAS by simulation studies. Our results demonstrate that both the single GC correction and the widely used double correction cannot control type I error rates in meta-analysis, when the test SNP has significant allele frequency differentiation in the subpopulations in the case–control data. On the other hand, the PCA correction can control type I error rates well. In addition, when population stratification exists in the case–control data, using double GC method usually results in much lower power than using the PCA correction in meta-analysis, even though the causal SNP does not have significant allele frequency differentiation in the subpopulations. We note that Price et al. (2006) reported similar results in single GWAS analysis. Therefore, the double GC method is not effective to correct for stratification in meta-analysis. Although the PCA correction method works well for correction for stratification in meta-analysis, it may need more than 5,000 SNPs to calculate PCs for each individual (Price et al., 2006). This may be challenging in replication studies with meta-analysis in which usually only a small set of significant SNPs identified from previous meta-analysis are collected and tested. If population stratification exists in the data for replication meta-analysis, we would suggest collecting additional 3,000 or 10,000 independent SNPs to calculate PCs for each individual, and use these PCs to correct for population stratification when testing the small set of SNPs not concordant in terms of positive or negative strand in the two data sets.
### Table 3 | Power of meta-analysis of GWAS for two studies using PCA correction, single GC correction (1 GC), and double GC correction (2 GC) at significance levels of $10^{-5}$, $10^{-6}$, and $10^{-7}$.

| Study 1 | Study 2 |
|---------|---------|
| R Case | Control | R Case | Control | 10$^{-5}$ | 10$^{-6}$ | 10$^{-7}$ | 10$^{-5}$ | 10$^{-6}$ | 10$^{-7}$ |
| 1.2 | 1000 | 1000 | 1.2 | 1000 | 1000 | 0.1901 | 0.0462 | 0.0425 | 0.0908 | 0.0113 | 0.0124 | 0.0394 | 0.0041 | 0.0045 |
| 1000 | 1000 | 1.3 | 1000 | 1000 | 0.4449 | 0.1207 | 0.1194 | 0.2797 | 0.0511 | 0.0502 | 0.1637 | 0.0199 | 0.0196 |
| 1500 | 1500 | 1.4 | 500 | 500 | 0.4232 | 0.1368 | 0.1450 | 0.2631 | 0.0604 | 0.0645 | 0.1437 | 0.0238 | 0.0253 |
| 1500 | 1500 | 1.5 | 750 | 1500 | 0.8332 | 0.4306 | 0.4368 | 0.7221 | 0.2724 | 0.2787 | 0.5959 | 0.1549 | 0.1602 |
| 1.3 | 1500 | 1500 | 1.3 | 500 | 500 | 0.7029 | 0.2412 | 0.2375 | 0.5467 | 0.1228 | 0.1205 | 0.3934 | 0.0552 | 0.0535 |
| 1000 | 1000 | 1.4 | 1000 | 1000 | 0.8768 | 0.4703 | 0.4866 | 0.7738 | 0.3033 | 0.3085 | 0.6386 | 0.1709 | 0.1680 |
| 1000 | 1000 | 1.5 | 1000 | 1000 | 0.9425 | 0.6377 | 0.6403 | 0.8635 | 0.4683 | 0.4702 | 0.8060 | 0.3128 | 0.3153 |
| 1.4 | 750 | 1500 | 1.4 | 1500 | 750 | 0.9951 | 0.6532 | 0.6570 | 0.9946 | 0.4740 | 0.4752 | 0.8251 | 0.3190 | 0.3215 |
| 1500 | 1500 | 1.5 | 500 | 500 | 0.9700 | 0.7008 | 0.6986 | 0.9302 | 0.5442 | 0.5400 | 0.8733 | 0.3819 | 0.3773 |
| 1.5 | 1500 | 1500 | 1.5 | 750 | 1500 | 0.9952 | 0.8942 | 0.8930 | 0.9842 | 0.8044 | 0.8028 | 0.9686 | 0.6844 | 0.6824 |

Boldface values denote the highest power among meta-analyses using the three methods (PCA, 1 GC, and 2 GC).

### Table 4 | Power of meta-analysis of GWAS for five studies using PCA correction, single GC correction (1 GC), and double GC correction (2 GC) at significance levels of $10^{-5}$, $10^{-6}$, and $10^{-7}$.

| Study 1 | Study 2 | Study 3 | Study 4 | Study 5 |
|---------|---------|---------|---------|---------|
| R Case | Control | PCA 1 GC | 2 GC | 1 GC | 2 GC | 1 GC | 2 GC | 1 GC | 2 GC | 1 GC | 2 GC | 1 GC | 2 GC |
| 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 0.8284 | 0.3120 | 0.3148 | 0.6912 | 0.1683 | 0.1703 | 0.5421 | 0.0623 | 0.0641 |
| 1.3 | 1.3 | 1.3 | 1.3 | 1.3 | 0.9985 | 0.8735 | 0.8796 | 0.9937 | 0.7586 | 0.7670 | 0.9873 | 0.6101 | 0.6226 |
| 1.4 | 1.4 | 1.4 | 1.4 | 1.4 | 1.000 | 0.9951 | 0.9951 | 1.000 | 0.9628 | 0.9632 | 1.000 | 0.9539 | 0.9545 |
| 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.000 | 0.9997 | 0.9997 | 1.000 | 0.9991 | 0.9991 | 1.000 | 0.9976 | 0.9979 |
| 1.2 | 1.2 | 1.3 | 1.3 | 1.3 | 0.9940 | 0.7678 | 0.7779 | 0.9829 | 0.6153 | 0.6310 | 0.9552 | 0.4548 | 0.4693 |
| 1.3 | 1.3 | 1.4 | 1.4 | 1.4 | 0.9998 | 0.9843 | 0.9854 | 0.9998 | 0.9598 | 0.9625 | 0.9996 | 0.9164 | 0.9202 |
| 1.3 | 1.4 | 1.4 | 1.5 | 1.5 | 1.000 | 0.9928 | 0.9931 | 1.000 | 0.9772 | 0.9783 | 1.000 | 0.9453 | 0.9472 |
| 1.2 | 1.2 | 1.4 | 1.4 | 1.5 | 0.9997 | 0.9267 | 0.9299 | 0.9985 | 0.8498 | 0.8560 | 0.9960 | 0.7414 | 0.7514 |

**Figure 1** | The stratification structure plotted by the first PC versus the second PC for the two data sets ARIC (A) and MESA (B).
Table 5 | Top SNPs with p-values < 10^{-3} in a meta-analysis with PCA correction, single GC correction (1 GC), or double GC correction (2 GC)

| Chromosome | SNP Reference allele | p-Value |
|------------|---------------------|-------|
| 1          | rs6686672 T         | 6.63e-6   |
| 1          | rs657751 G          | 3.46e-6   |
| 1          | rs1049566 C         | 6.64e-4   |
| 4          | rs767213 G          | 4.06e-5   |
| 4          | rs3782685 A         | 6.44e-6   |
| 5          | rs1477626 A         | 7.29e-5   |
| 5          | rs1090562 A         | 4.07e-6   |
| 9          | rs1090592 T         | 5.50e-6   |

of promising SNPs identified from previous meta-analysis. As can be seen from the results described above, using PCA correction controls type I error rates well, and more importantly can have increased power in meta-analysis compared to the widely used double GC correction.

When evaluating the three methods for correcting for stratification in meta-analysis for case-control studies, we only considered the fixed-effect model and the additive genetic model for case-control designs with binary traits. We expect that the conclusion will also hold for quantitative traits. We plan to conduct stratification correction used in meta-analysis in our future study.

In our simulation studies, we did consider cryptic relatedness among individuals. Since GC method can be useful for controlling spurious association caused by relatedness in the data, investigators proposed a “PCA + double GC” method to control the spurious association findings in meta-analysis as follows: 1) to perform PCA adjustment for population stratification in the individual study association analysis, followed by a GC correction on the genome-wide results, 2) to perform a GC correction on the combined statistics over all studies by the corresponding GC inflation factor λ (see Materials and Methods). Although this “PCA + double GC” method may control type I error rate well, it may not maximize power to detect true associations in meta-analysis (Price et al., 2010). To adjust for the cryptic relatedness and population stratification in meta-analysis, a better choice may be using the mixed model approaches, such as the EMMAX and TASSEL software (Kang et al., 2010; Zhang et al., 2010).

ACKNOWLEDGMENTS

This research was supported by grants GM073766 and GM081488 from the National Institute of General Medical Science and U01HL101064 from the National Heart, Lung, and Blood Institute (NHLBI). Shuodong Wang was partially supported by grant 61170183 from National Natural Science Foundation of China, B2011JS0025 from Excellent Young and Middle-Aged Scientists Fund of Shandong Province of China and SDUST Research Fund of China (2010 KYJQ104). The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by NHLBI contracts (HHSN268201100009C, HHSN268201100008C, HHSN268201100007C, HHSN268201100006C, HHSN268201100005C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C). The authors thank the staff and participants of the ARIC study for their important contributions. Funding for CARE genotyping was provided by NHLBI contract N01-HC-95159 through N01-HC-95169 and RR-024156. Funding for genotyping was provided by NHLBI contracts N02-HL-6-4278 and N01-HC-65226. MESA and the MESA SHARE project are conducted and supported by the NHLBI in collaboration with MESA investigators. Support for MESA is provided by contracts N01-HC-95159 through N01-HC-95169 and RR-024156. Funding for genotyping was provided by NHLBI contracts N02-HL-6-4278 and N01-HC-65226.

REFERENCES

Balding, D. J., and Nichols, R. A. (1995). A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. Genet. Epidemiol. 12, 585-94.

Cardon, L. R., and Palmer, L. J. (2003). Population stratification and spurious allelic association. Lancet 361, 598-604.

Devlin, B., Bacanu, S. A., and Roeder, K. (2004). Genomic control to the extreme. Nat. Genet. 36, 1129-1130.

Devlin, B., Roeder, K. (1999). Genomic control for association studies. Biometrics 55, 997-1004.

Devlin, B., Roeder, K., and Wasserman, L. (2001). Genomic control: a new approach to genome-wide association studies. Theor. Popul. Biol. 60, 155-166.

Freidman, M. L., Reich, D., Penney, K. L., McDonald, G. J., Mignone, A. A., Patterson, N., et al. (2006). Assessing the impact of population stratification on genetic association studies. Nat. Genet. 38, 898-903.

Hedrick, R. S., Webb, E., Bruderer, P., Pittman, A. M., De Bernardo, M. C., Lathrop, S., et al. (2008). Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. Nat. Genet. 40, 1437-1439.

Jackson, J. E. (2001). A User’s Guide to Principal Components. New York: John Wiley & Sons.

Kang, H. M., Sul, J. H., Service, K. S., Zaitlen, N. A., Kong, S., Freimer, N. B., et al. (2010). Vari- ant component model to account for sample structure in genome-wide association studies. Nat. Genet. 42, 348-354.

Krumh, H., Han, C., Post, W., Geff, D., Duro-Brown, A., Cooper, R., et al. (2010). Racial/Ethnic difference in Hypertension and Hypertension treatment and control in the Multi-Ethnic Study of Atherosclerosis (MESA). Ann. J. Hypertens. 27, 965-970.

Letts, G., Palmer, C. D., Young, T., Eebe, K. G., Allman, H., Bertram, E. J., et al. (2011). Genome-wide association study of coronary heart disease and its risk factors in 8,090 African Americans the NHLBI CARE Project. PLoS Genet. 7:e1001905. doi: 10.1371/journal.pgen.1001905

Liu, Y., Zhao, H., Patki, A., Linds, N. A., and Allison, D. B. (2011). Controlling population structure in human genetic association studies with samples of unusual individuals. Stat. Interface 4, 317-326.

Lohmueller, K. E., Franco, C. L., Piko, M., Lander, E. S. and Hirschhorn, J. N. (2003). Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat. Genet. 35, 177-182.

Mige, R., and Morris, A. F. (2010). GWAMA: software for genome-wide association meta-analysis.
Double GC is not effective in meta-analysis studies. Nat. Rev. Genet. 11, 459–63.

Price, J. K., and Donnelly, P. (2001). Case-control studies of association in structured or admixed population. Theor. Popul. Biol. 60, 227–237.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38, 904–909.

Price, A. L., Zaitlen, N. A., Reich, D., and Patterson, N. (2010). New approaches to population stratification in genome-wide association studies. Nat. Rev. Genet. 11, 5–6.

Pritchard, J. K., and Donnelly, P. (2001). Theoretical basis for case-control studies of association in structured or admixed population. Theor. Popul. Biol. 60, 227–237.

Qayyum, R., Snively, B. M., Ziv, E., Nalls, M. A., Liu, Y. M., Tang, W. H., et al. (2012). A meta-analysis of genome-wide association studies identifies seven new rheumatoid arthritis risk loci. Nat. Genet. 44, 508–514.

Wang, D., Sun, Y., Steng, P., Berlin, J. A., Wilson, M. A., and Li, Q. Q. (2009). Comparison of methods for correcting population stratification in a genome-wide association study of rheumatoid arthritis: principal-component analysis versus multidimensional scaling. BMC Proc. 3(Suppl. 7):S109. doi: 10.1186/1753-6561-3-S7-S109

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 18 September 2012; paper pending published: 08 October 2012; accepted: 04 December 2012; published online: 24 December 2012.

Citation: Wang S, Chen W, Chen X, He F, Archer EJ, Liu N, Sun S and Gao G (2012) Double genomic control is not effective to correct for population stratification in meta-analysis for genome-wide association studies. Front. Gene. 3:300. doi: 10.3389/fgene.2012.00300

This article was submitted to Frontiers in Applied Genetic Epidemiology, a specialty of Frontiers in Genetics.

Copyright © 2012 Wang, Chen, Chen, He, Archer, Liu, Sun and Gao. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.

www.frontiernan.org

December 2012 | Volume 3 | Article 300 | 7

“fgene-03-00300” — 2012/12/22 — 10:43 — page 7 — #7