Genomics 105 (2015) 191–196

Contents lists available at ScienceDirect

Genomics

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

A mixed model reduces spurious genetic associations produced by population stratification in genome-wide association studies

Jimin Shin, Chaeyoung Lee *

Department of Bioinformatics and Life Science, Soongsil University, Seoul 156-743, Republic of Korea

ARTICLE INFO

Article history:
Received 4 July 2014
Accepted 23 January 2015
Available online 30 January 2015

Keywords:
False discovery
Genomic control
Mixed model
Population stratification
Statistical power

ABSTRACT

Population stratification can produce spurious genetic associations in genome-wide association studies (GWASs). Mixed model methodology has been regarded useful for correcting population stratification. This study explored statistical power and false discovery rate (FDR) with the data simulated for dichotomous traits. Empirical FDRs and powers were estimated using fixed models with and without genomic control and using mixed models with and without reflecting loci linked to the candidate marker in genetic relationships. Population stratification with admixture degree ranged from 1% to 10% resulted in inflated FDRs from the fixed model analysis without genomic control and decreased power from the fixed model analysis with genomic control ($P < 0.05$). Meanwhile, population stratification could not change FDR and power estimates from the mixed model analyses ($P > 0.05$). We suggest that the mixed model methodology was useful to reduce spurious genetic associations produced by population stratification in GWAS, even with a high degree of admixture (10%).

1. Introduction

Recent identifications of nucleotide sequence variants underlying human complex diseases have greatly relied on genome-wide association studies (GWASs). Inconsistency of the results across various GWASs might be attributed to the heterogeneity of populations. False discovery might be a possible cause, although many studies have attempted to employ a conservative multiple testing method, the Bonferroni adjustment, to avoid it. This false discovery might come from population stratification; that is, the different allele frequencies between cases and controls are attributed to spurious genetic associations caused by systematic differences in ancestry [1,2]. Researchers have used a variety of methods to solve this problem. Genomic control is the most common method for dealing with population stratification, by uniform adjustment of all the association statistics using the genomic inflation factor (λ) [3]. This method has the critical limitation of inflated type II error because the uniform adjustment might be an improper account for rare variants and variants with largely heterogeneous frequency in populations. Principal component analysis (PCA) [4] and structured association analysis [5] are often used to correct for stratification, but have suffered shortcomings. These approaches should be conducted by a subjective determination of subpopulations within the study population, with a high complexity.

A recent advancement in the correction for population stratification was achieved by employing a mixed model [6–10]. When associations between genetic variants and complex diseases are tested, the mixed model methodology reflects the polygenic effects explained by the genetic relationships among individuals using genomic information. Use of mixed models has restricted overestimation of statistics by population stratification in testing genetic associations [9,11]. However, the effectiveness of the mixed model in identifying genetic associations against population stratification may depend on admixture degree of populations. The objectives of this study were to explore false discovery and statistical power using GWAS data simulated with population stratification and to compare the empirical estimates obtained by mixed models to those by fixed models.

2. Results

2.1. False discovery rate

FDRs were empirically estimated with the simulated populations, and their means are presented by heritability, prevalence, degree of population stratification, and analytical method in Fig. 1. Simulated data are also available at http://clee11.cafe24.com/mmreducesfalsediscovery.html. Use of mixed models reduced the FDRs, compared to those obtained with fixed models ($P < 0.05$). The difference increased with a heritability of 0.5, and a prevalence of 0.05. The FDRs obtained in mixed models were consistent, regardless of the degree of admixture with respect to population stratification. In contrast, FDRs increased using the fixed models with an increase in the degree of admixture. As a result, the difference was dramatically increased for populations with a large degree

* Corresponding author at: Department of Bioinformatics and Life Science, Soongsil University, 511 Sangdo-dong, Dongjak-gu, Seoul 156-743, Republic of Korea. Fax: +82 2 824 4383.
E-mail address: clee@ssu.ac.kr (C. Lee).

http://dx.doi.org/10.1016/j.ygeno.2015.01.006
0888-7543/

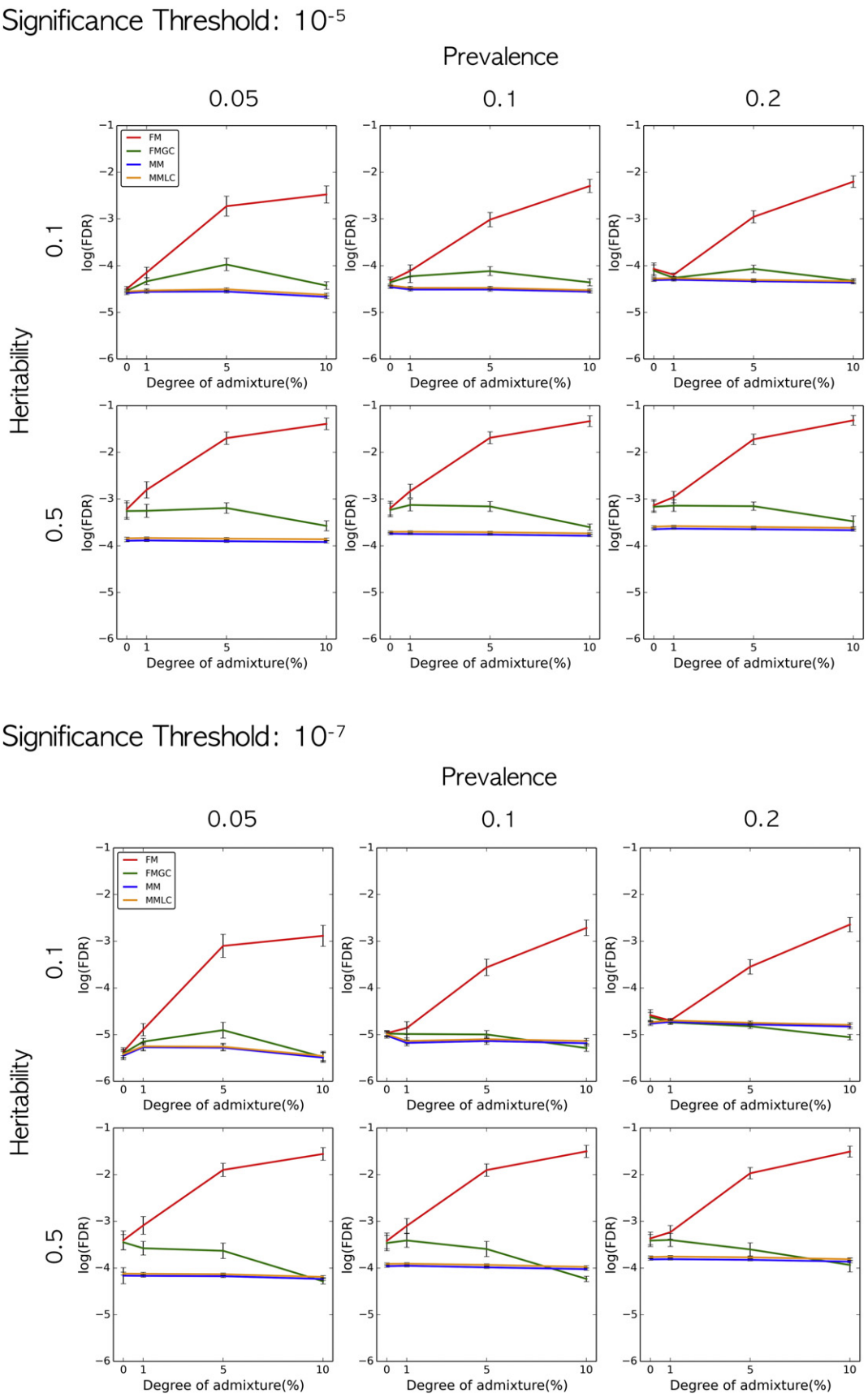

Fig. 1. False discovery rate (FDR) in the genome-wide association study (GWAS) for dichotomous traits using the fixed model with (FMGC) and without (FM) genomic control, and using the mixed model with (MMLC) and without (MM) reflecting loci linked to the candidate marker in the genetic relationship between individuals. The FDR was estimated with a false positive significance threshold of 10^{-5} or 10^{-7}. The mean estimate is accompanied with the vertical bar presenting its standard error empirically obtained from 50 replicates.

of admixture, with respect to population stratification. The fixed model analysis showed that the tests adjusted by employment of the genomic inflation factor reduced the errors, compared to those without the adjustment. The mixed model analysis showed a negligible difference between the errors with and without the reflecting loci linked to the candidate marker in the genetic relationship between individuals called the leaving-one-chromosome-out (LOCO) analysis.

2.2. Statistical power

Empirical mean estimates of statistical powers by heritability, prevalence, degree of population stratification, and analytical method were also obtained from the simulation study (Fig. 2). The statistical power estimates for data without population stratification were corresponding regardless of the analytical methods used. The statistical power, however, decreased with an increase in the admixture of the population in the fixed model analysis with genomic control. The statistical power was also observed to be lower with a larger heritability and a larger prevalence in that model. On the other hand, the power did not decrease in the other analyses. The fixed model analysis without genomic control resulted in a dramatically larger power than that with genomic control. Also, use of mixed models improved statistical powers in comparison with the genomic control approach using a fixed model ($P < 0.05$). The statistical powers were improved with a large heritability (0.5) and a large prevalence (0.2). The power estimates from the mixed model analysis showed little difference with and without LOCO.

3. Discussion

Population stratification has been a great concern in uncovering nucleotide variants for complex disease in GWAS. The mixed model is expected to lessen the problem by explaining the genetic relationship of individuals using genomic information. This might lead to a larger statistical power and a smaller FDR [12]. The simulations in the current study showed that mixed models perform better in reducing false discovery and retaining statistical power than fixed models, especially for populations with a large degree of admixture.

The results concurred with those from previous studies which had shown some improvement with mixed models in reducing false positive and negative associations. When single-trait and multiple-trait mixed models were employed in a GWAS for human cholesterol levels, quantile–quantile plot did not show any evidence of confounding produced by population stratification [9]. Their simulation study showed that multiple-trait mixed model performed better than single-trait mixed model in improving statistical power and FDR. Quite recently, Tucker et al. [11] proposed a modified approach called PC-Select, which required a principal component analysis with genotype matrix prior to the mixed model analysis using a genetic relationship matrix constructed with top associated SNPs (FaST-LMM Select [13]). The PC-Select was efficient with simulated dichotomous data in avoiding inflation of statistics produced by population stratification, a problem partially remained with the FaST-LMM Select. They also suggested that the PC-Select was even better than the mixed model used in this study, which employed genetic relationship matrix constructed with all the SNPs under the assumption of infinitesimal model. Nevertheless, if the number of genetic variants influencing complex disease is not small (e.g. >100), the infinitesimal model might be an excellent choice to explain population stratification explicitly. This is because a previous simulation study [14] showed that the analysis under the wrong assumption of infinitesimal model negligibly influenced heritability estimates, thus might have limited influence on causal genetic effects and polygenic effects. Also, the study of Tucker et al. showed that the difference was not found ($P > 0.05$) between the infinitesimal model and the PC-Select in terms of inflation of statistics [11]. But the statistical power was improved with the PC-Select. This might be because their phenotypes were simulated by adding 0.25 times of the first principal component of the genotype matrix, and this must have shown a better fit with the PC-Select.

The current study also revealed that the mixed model worked even better than the fixed model with a genomic control, the most common method known to correct population stratification, in respect to both statistical power and false discovery rate. The genomic control reduced false discovery, but caused a simultaneous decrease in statistical power. This concurred with results of previous simulation studies [2,4]. The current study suggested that the statistical power dramatically decreased with an increase in the degree of admixture in the population. It was also found that genomic control led to a dramatic decrease in statistical power, compared to that from the mixed model analysis. Although the false discovery can decrease more with the use of the genomic control than with the mixed model, the mixed model is preferable because the statistical power and false discovery are mutually compensatory. That is, use of a more stringent false positive threshold value leads to a smaller statistical power [15], as shown in the current study in which statistical power was larger with the threshold value of 10^{-7} than that with the threshold value of 10^{-5} (Fig. 2). Thus, using stringent multiple testing reduced the error (Fig. 1), but it could obstruct identification of truly causal variants.

Statistical power and false discovery rate were not influenced by an increase in the degree of admixture in the population using a mixed model. On the other hand, obvious changes were observed using the fixed model with or without the genomic control. This implies that population stratification can be appropriately controlled by employing a mixed model.

A mixed model analysis performed including the candidate marker in the genetic relationship matrix (GRM) could lead to loss in power. The decreased power is caused by redundant use of the candidate marker in the analytical model, both as a fixed effect tested for association and as a random effect as part of the GRM [12,13,16]. However, the differences in statistical power and false discovery rate by the analyses performed here with and without LOCO were negligible. This implied that the use of LOCO in the mixed model framework might not be essential for identifying genetic associations with the false positive threshold of $P = 10^{-5}$ or $P = 10^{-7}$. We suspect that the result might not have come from the absence of LOCO effects, but from compensation between two forces. One force is to increase individual variant effects by prohibiting them to be absorbed to polygenic effects using LOCO. The other is to reduce individual variant effects as well as polygenic effects with insufficient correction for population stratification by ignoring all the genetic information in one chromosome. Of course, all the genetic information is not necessarily considered for polygenic effects [14].

The results from the current study were novel especially in simulating population stratification by various degrees of admixture of major and minor populations in contrast with the previous studies [9,11] simulating a population stratification suspected in real data without a definite admixture degree. As a result, the current study could show influences of population stratification more explicitly by comparing data with and without population stratification. Furthermore, the superiority of mixed model and the trend in statistical power and FDR by degree of admixture would provide more insights into GWAS where they most likely consider a study population within a country and have intrinsic problems with migrants and/or mixed-blood individuals. The current findings might be applied to many complex diseases under the assumption of normally distributed underlying liability for dichotomous phenotypes. Although the current simulation study did not deal with any specific covariates in a variety of diseases, we found negligible differences in powers/FDRs from data simulated with and without age as a covariate for a few designs of this study ($P > 0.05$, data not shown). However, attention should be given for the diseases with additional covariates influential to them. We should also be cautious for the use of admixed populations. This study implies that

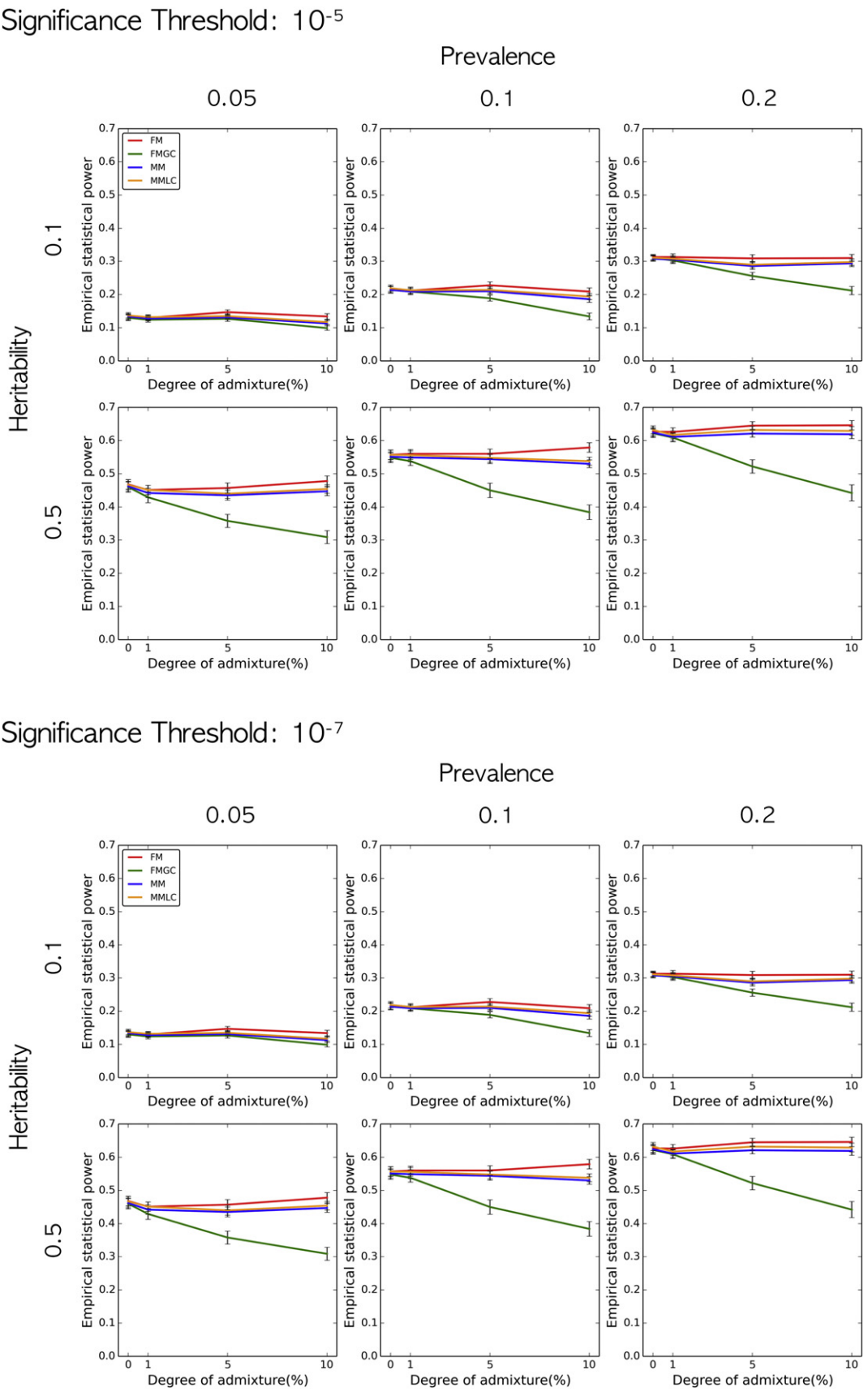

Fig. 2. Statistical power in genome-wide association study (GWAS) for dichotomous trait using the fixed model with (FMGC) and without (FM) genomic control, and the mixed model with (MMLC) and without (MM) reflecting loci linked to the candidate marker in the genetic relationship between individuals. The statistical power was estimated with a false positive significance threshold of 10^{-5} or 10^{-7}. The mean estimate is accompanied with the vertical bar presenting its standard error empirically obtained from 50 replicates.

employing the mixed model can appropriately explain genome-wide linkage disequilibrium produced by population stratification, but not explain potential bias produced by environmental effects associated with the admixed populations.

The current study demonstrated that the mixed model with a genomic covariance structure could improve both the false discovery rate and statistical power with respect to population stratification with admixture degree of 1–10%. The mixed model methodology was particularly useful in reducing spurious genetic associations in GWAS.

4. Materials and methods

4.1. Genotypic data

Subjects with genotypes from different ethnic populations were used in order to simulate dichotomous traits with population stratification. The major subjects were selected from a Korean population of the Korea Association REsource (KARE) Analysis Consortium. The population was recruited for a large-scale GWAS which included 8842 individuals at 351,677 single nucleotide polymorphisms (SNPs) of the Affymetrix Genome-Wide Human SNP Array 5.0 (Affymetrix, Inc., Santa Clara, CA, USA) after quality assurance screening [17]. All the SNPs were selected by genotype call rate > 95%, minor allele frequency > 0.01, and Hardy–Weinberg equilibrium, and all the individuals were selected by genotype call rate > 95% and pairwise genetic relationship coefficients < 0.025 [18]. Deviation from the Hardy–Weinberg equilibrium was determined by Pearson's chi-square test with a significance threshold of 10^{-6}. Genomic data were also used from the Phase 1 release of the 1000 Genomes Project [19]. We used populations of 379 Europeans, 246 Africans, and 75 Americans (55 Puerto Ricans and 20 Columbians). The data contained about 38 million SNPs.

After excluding sex chromosomal SNPs, all the populations shared 342,220 autosomal SNPs. Among them, only 186,314 unlinked variants with $r^2 < 0.8$ in sliding 50 SNP windows were used in the simulation.

4.2. Simulation of phenotypes with population stratification

A Monte Carlo simulation was conducted for dichotomous phenotypic data based on genomic information reflecting population stratification. To mimic a population stratification model, four scenarios were devised with different components of ethnic populations. The populations were composed as follows: 1) only 7000 Koreans, under the assumption of 0% admixture; 2) 6930 Koreans and 70 others, under the assumption of 1% admixture; 3) 6650 Koreans and 350 others, under the assumption of 5% admixture; 4) 6300 Koreans and 700 others, under the assumption of 10% admixture. Individuals were randomly selected from the Korean population and from the 3 other ethnic populations included in the 1000 Genome Project.

In order to simulate phenotypes, we first simulated the genetic effects of the 186,314 unlinked SNPs based on observed genotype data. We randomly selected 20 out of 186,314 SNPs and assigned them as causal variants. The remaining 186,294 variants were assigned as polygenic variants. Their genetic effects were randomly generated from Normal distribution under 1:1 variance ratio of causal genetic effects to polygenic effects. The causal genetic effect was generated from $N\left(0, \frac{0.5}{20}\right)$, and the polygenic effect was generated from $N\left(0, \frac{0.5}{186{,}294}\right)$. Simulated genetic effects were all added under the assumption of additive genetic model. Then, environmental effect was generated from the Normal distribution with the mean of 0 and the variance of $\frac{1-h^2}{h^2}$ where h^2 is heritability with a value of 0.1 or 0.5 reflecting both causal and polygenic effects. The environmental variance was equal to 9 for $h^2 = 0.1$ and to 1 for $h^2 = 0.5$. Simulated environmental effect was added to the genetic effects. The phenotype of a dichotomous trait was determined as a case or a control. This assumed a disease liability with the threshold of the Normal distribution determined by a disease prevalence of 0.05, 0.1, or 0.2. The control was an individual with a smaller disease liability than the threshold, whereas the case was an individual with a larger disease liability than the threshold. All the input values used in this simulation are summarized in Table 1. A total of 24 designs were simulated with a variety of population stratification (4 levels), heritability (2 levels), and prevalence (3 levels). For each design, populations were simulated in 50 replicates, which had been proven sufficient in our preliminary analysis to obtain estimates of false discovery and statistical power with reasonable sampling error to be tested for comparison.

4.3. Analysis with fixed models

Genome-wide genetic associations were analyzed with the simulated data using a fixed model or a mixed model. The fixed model included a fixed additive allele effect to be tested for genetic associations. The fixed model analysis was conducted with and without a genomic control method for correcting for population stratification [3]. Genomic inflation factor (λ) was estimated as the mean of χ^2 statistics calculated with 186,314 variants by the Cochran–Armitage trend test [20]. The association analysis with the fixed model was conducted using PLINK v1.06 software [21].

4.4. Analysis with mixed models

The simulated data were also analyzed using the following mixed model:

$$\boldsymbol{y} = \mu 1 + \mathbf{x}\beta + \boldsymbol{g} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{y}$ is the vector of dichotomous phenotypes; μ is the overall mean; 1 is the vector of 1's; β is the fixed effect for the minor allele of the candidate SNP; and $\mathbf{x}$ is the vector with elements of 0, 1, and 2 for the homozygote of the minor allele, heterozygote, and homozygote of the major allele. $\boldsymbol{g}$ is the vector of random polygenic effects ($\boldsymbol{g} \sim N(0, \boldsymbol{G}\sigma_g^2)$), where σ_g^2 is the polygenic variance component, and $\boldsymbol{G}$ is the $n \times n$ genomic relationship matrix of which elements consist of pairwise genetic relationship coefficients estimated using genotypes of 186,314 SNPs in linkage equilibrium ($r^2 > 0.8$). The pairwise relationship coefficient was calculated as:

$$\frac{1}{186{,}314}\sum_{i=1}^{186{,}314}\frac{\left(x_{ij}-2p_i\right)\left(x_{ik}-2p_i\right)}{2p_i(1-p_i)}$$

where x_{ij} and x_{ik} represent the number (0, 1, or 2) of the minor allele for the ith SNP of the jth and kth individuals, and p_i is the frequency of the minor allele. $\boldsymbol{\varepsilon}$ is the vector of random environmental effects ($\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{I}\sigma_\varepsilon^2)$), where σ_ε^2 is the environmental variance

Table 1
Input values for simulating dichotomous phenotypes in this study.[a]

Item	Input values/description
Population stratification	Level of admixture degree: 0%, 1%, 5%, 10%
Prevalence	Rate: 0.05, 0.1, 0.2
Heritability	0.1, 0.5
Causal genetic effect	Generated from Normal distribution for 20 SNPs Causal genetic variance: 0.5[a]
Polygenic effect	Generated from Normal distribution for 186,294 SNPs Polygenic variance: 0.5[a]
Environmental effect	Generated from Normal distribution Environmental variance[b]: 9,1

[a] Variances for individual SNP are $\frac{0.5}{20}$ for causal genetic effect and $\frac{0.5}{186{,}294}$ for polygenic effect.
[b] Determined by heritability, i.e. 9 for heritability = 0.1 and 1 for heritability = 0.5.

component, and $\boldsymbol{I}$ is the $n \times n$ identity matrix. The polygenic and environmental variance components were estimated using the restricted maximum likelihood (REML). We first estimated polygenic and environmental variance components by EM-REML, then the estimates were used as initial values in order to obtain their AI-REML estimates. The fixed and random effects were then solved with the estimated variance components under the mixed model equations.

Additional analysis was conducted with the genetic relationship coefficients estimated by excluding the chromosome in which the candidate SNP was located. The following modified mixed model was used for the leaving-one-chromosome-out (LOCO) analysis.

$$\boldsymbol{y} = \mu 1 + \mathbf{x}\beta + \boldsymbol{g}^{-} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{g}^{-}$ is the vector of random polygenic effects using the genome without the chromosome where the candidate SNP is located. That is, $\boldsymbol{g}^{-} \sim N\left(0, \boldsymbol{G}^{-}\sigma^2_{g^-}\right)$ where $\boldsymbol{G}^{-}$ is the $n \times n$ genomic relationship matrix without one chromosome, and $\sigma^2_{g^-}$ is the polygenic variance component that should be re-estimated whenever a specific chromosome is excluded from calculating the genetic relationship matrix.

We employed two values (10^{-5} and 10^{-7}) as the false positive threshold. Empirical false discovery rate (FDR) was calculated as the probability of the identification of simulated non-causal SNPs, and empirical statistical power was calculated as the probability of the identification of simulated causal SNPs. The association analysis with the mixed model was conducted using the GCTA v2.0 freeware [22]. Significant differences in the empirical FDR and empirical statistical power among the analytical methods were determined by the paired t-test, using SPSS v21.0 (SPSS Inc., Chicago, IL, USA).

4.5. Comparison of false discovery rate and statistical power

We compared the FDR estimates obtained from different analytical methods. Influence of population stratification on the FDR estimates was determined by comparing the estimates from data simulated with and without population stratification. Statistical power estimates were also compared in the same way. All the comparisons were tested by t statistic with false positive error of 0.05.

Acknowledgments

The valuable and constructive comments of anonymous reviewers on an earlier version of this article are greatly appreciated. We are grateful to the National Institute of Health in Korea for providing the genotypic and epidemiological data to the KARE Analysis Consortium. This study was supported by the Basic Science Research Program of the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (Grant No. 2012002096).

References

[1] L.R. Cardon, L.J. Palmer, Population stratification and spurious allelic association, Lancet 361 (2003) 598–604.
[2] J. Marchini, L.R. Cardon, M.S. Phillips, P. Donnelly, The effects of human population structure on large genetic association studies, Nat. Genet. 36 (2004) 512–517.
[3] B. Devlin, K. Roeder, L. Wasserman, Genomic control, a new approach to genetic-based association studies, Theor. Popul. Biol. 60 (2001) 155–166.
[4] A.L. Price, N.J. Patterson, R.M. Plenge, M.E. Weinblatt, N.A. Shadick, D. Reich, Principal components analysis corrects for stratification in genome-wide association studies, Nat. Genet. 38 (2006) 904–909.
[5] J.K. Pritchard, M. Stephens, N.A. Rosenberg, P. Donnelly, Association mapping in structured populations, Am. J. Hum. Genet. 67 (2000) 170–181.
[6] J. Yu, G. Pressoir, W.H. Briggs, I.V. Bi, M. Yamasaki, J.F. Doebley, M.D. McMullen, B.S. Gaut, D.M. Nielsen, J.B. Holland, S. Kresovich, E.S. Buckler, A unified mixed-model method for association mapping that accounts for multiple levels of relatedness, Nat. Genet. 38 (2006) 203–208.
[7] H.M. Kang, J.H. Sul, S.K. Service, N.A. Zaitlen, S.Y. Kong, N.B. Freimer, C. Sabatti, E. Eskin, Variance component model to account for sample structure in genome-wide association studies, Nat. Genet. 42 (2010) 348–354.
[8] Z. Zhang, E. Ersoz, C.Q. Lai, R.J. Todhunter, H.K. Tiwari, M.A. Gore, P.J. Bradbury, J. Yu, D.K. Arnett, J.M. Ordovas, E.S. Buckler, Mixed linear model approach adapted for genome-wide association studies, Nat. Genet. 42 (2010) 355–360.
[9] A. Korte, B.J. Vilhjálmsson, V. Segura, A. Platt, Q. Long, M. Nordborg, A mixed-model approach for genome-wide association studies of correlated traits in structured populations, Nat. Genet. 44 (2012) 1066–1071.
[10] V. Segura, B.J. Vilhjálmsson, A. Platt, A. Korte, Ü. Seren, Q. Long, M. Nordborg, An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations, Nat. Genet. 44 (2012) 825–830.
[11] G. Tucker, A.L. Price, B.A. Berger, Improving the power of GWAS and avoiding confounding from population stratification with PC-Select, Genetics 197 (2014) 1045–1049.
[12] J. Yang, N.A. Zaitlen, M.E. Goddard, P.M. Visscher, A.L. Price, Advantages and pitfalls in the application of mixed-model association methods, Nat. Genet. 46 (2014) 100–106.
[13] J. Listgarten, C. Lippert, C.M. Kadie, R.I. Davidson, E. Eskin, D. Heckerman, Improved linear mixed models for genome-wide association studies, Nat. Methods 9 (2012) 525–526.
[14] H. Ryoo, C. Lee, Underestimation of heritability using a mixed model with a polygenic covariance structure in a genome-wide association study for complex traits, Eur. J. Hum. Genet. 22 (2014) 851–854.
[15] T.V. Perneger, What's wrong with Bonferroni adjustments, BMJ 316 (1998) 1236.
[16] C. Lippert, J. Listgarten, Y. Liu, C.M. Kadie, R.I. Davidson, D. Heckerman, FaST linear mixed models for genome-wide association studies, Nat. Methods 8 (2011) 833–835.
[17] Y.S. Cho, M.J. Go, Y.J. Kim, J.Y. Heo, J.H. Oh, H.J. Ban, D. Yoon, M.H. Lee, D.J. Kim, M. Park, S.H. Cha, J.W. Kim, B.G. Han, H. Min, Y. Ahn, M.S. Park, H.R. Han, H.Y. Jang, E.Y. Cho, J.E. Lee, N.H. Cho, C. Shin, T. Park, J.W. Park, J.K. Lee, L. Cardon, G. Clarke, M.I. McCarthy, J.Y. Lee, J.K. Lee, B. Oh, H.L. Kim, A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits, Nat. Genet. 41 (2009) 527–534.
[18] J. Shin, C. Lee, Statistical power for identifying nucleotide markers associated with quantitative traits in genome-wide association analysis using a mixed model, Genomics 105 (2015) 1–4.
[19] 1000 Genomes Project Consortium, G.R. Abecasis, A. Auton, L.D. Brooks, M.A. DePristo, R.M. Durbin, R.E. Handsaker, H.M. Kang, G.T. Marth, G.A. McVean, An integrated map of genetic variation from 1,092 human genomes, Nature 491 (2012) 56–65.
[20] P. Armitage, Tests for linear trends in proportions and frequencies, Biometrics 11 (1955) 375–386.
[21] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M.A. Ferreira, D. Bender, J. Maller, P. Sklar, P.I. de Bakker, M.J. Daly, P.C. Sham, PLINK: a tool set for whole-genome association and population-based linkage analyses, Am. J. Hum. Genet. 81 (2007) 559–575.
[22] J. Yang, S.H. Lee, M.E. Goddard, P.M. Visscher, GCTA: a tool for genome-wide complex trait analysis, Am. J. Hum. Genet. 88 (2011) 76–82.