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/

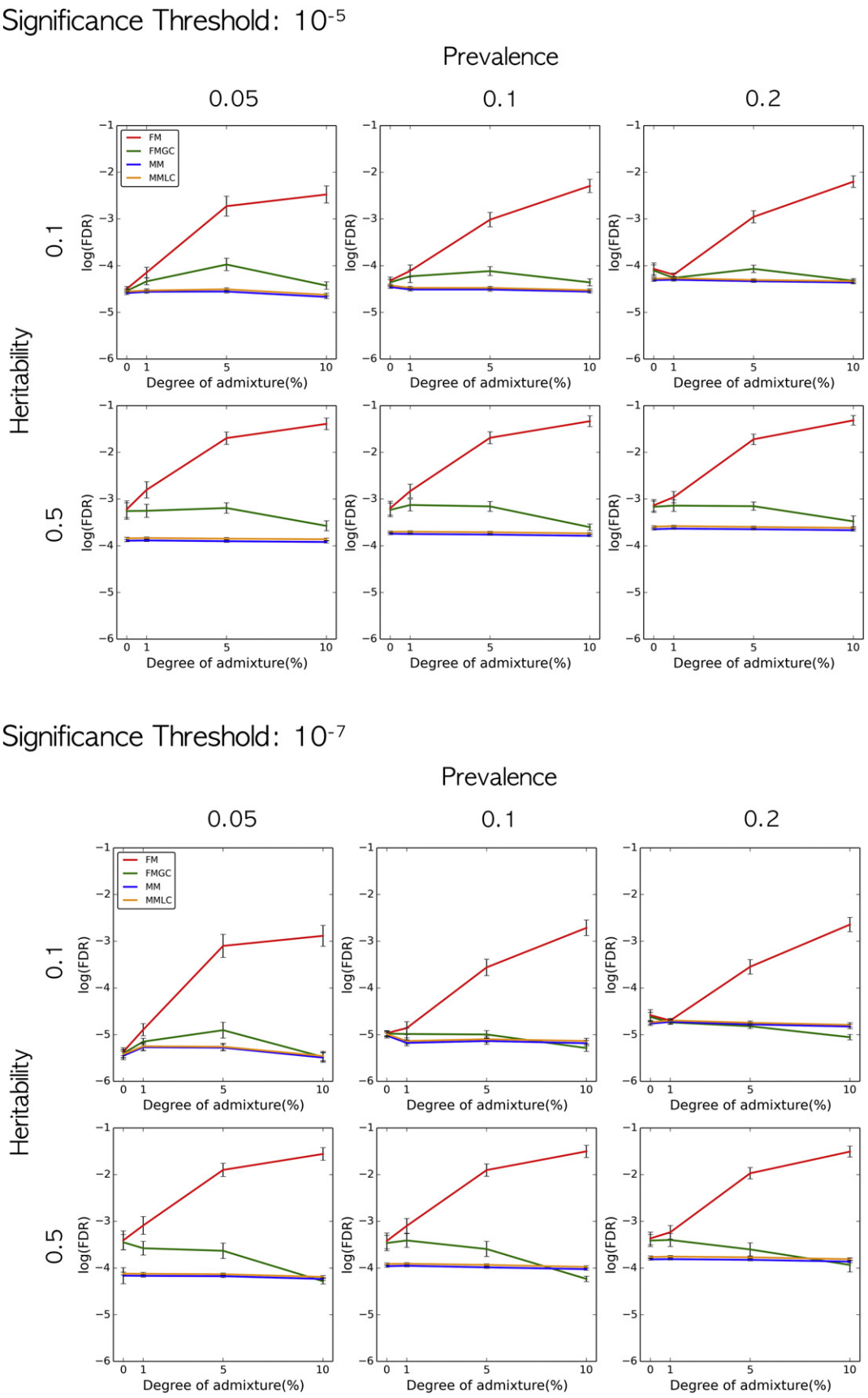

Fig. 1. False discovery rate (FDR) in the genome-wide association study (GWAS) for dichotomous traits using the fixed model with (FMGC) and without (FM) genomic control, and using the mixed model with (MMLC) and without (MM) reflecting loci linked to the candidate marker in the genetic relationship between 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.