Genomics 105 (2015) 1–4

Contents lists available at ScienceDirect

Genomics

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

Short Communication

Statistical power for identifying nucleotide markers associated with quantitative traits in genome-wide association analysis using a mixed model

Jimin Shin, Chaeyoung Lee *

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

ARTICLE INFO

Article history:
Received 12 March 2014
Accepted 5 November 2014
Available online 13 November 2014

Keywords:
Heritability
Mixed model
Simulation
Statistical power

ABSTRACT

Use of mixed models is in the spotlight as an emerging method for genome-wide association studies (GWASs). This study investigated the statistical power for identifying nucleotide variants associated with quantitative traits using the mixed model methodology. Quantitative traits were simulated through design of heritability, the number of causal variants (NCV), the number of polygenic variants, and genetic variance ratio of causal to polygenic variants (VRCTP). Statistical power estimates were influenced not only by individual factors of heritability, NCV, and VRCTP, but also by their interactions ($P < 0.05$). As the genetic variance ratio decreased, the difference in power between heritabilities of 0.3 and 0.5 increased with the use of 20 causal variants, but decreased when there were 100 causal variants ($P < 0.05$). The power empirically estimated from the simulation study would be applicable to the design of GWAS for quantitative traits with known genetic parameters by predicting the degree of false negative associations.

1. Introduction

Genome-wide association studies (GWASs) have become routine for the discovery of genetic factors. The statistical power in detecting a causal genetic variant for quantitative phenotypes of human and other species has been a great concern to geneticists. The statistical power in GWAS is typically influenced by effect size of the causal variant as well as sample size [1,2]. In identifying causal variants, the power decreases with moderate effect sizes (odds ratio (OR) = 1.1–1.5) and with low allele frequencies (<0.1) [3]. Of course, phenotyping and genotyping errors also reduce the power [4]. The statistical power of GWAS when testing genetic association with analytical models dealing only with a single variant may lead to a rather limited inference.

Mixed model methodology was recently introduced to GWAS to explain polygenic effects of quantitative phenotypes [5,6]. That is, associations with causal variants can be tested using the mixed model, which simultaneously reflects genetic relationships among individuals with genotypic information of variants across the human genome. Corresponding statistical powers have been suggested given some effect sizes [6–8]. However, unknown effect size when planning GWAS design limits prediction of the power. This study aimed to examine the statistical power of GWAS when employing the mixed model and to predict the power with heritability.

2. Results

2.1. Statistical power

The statistical powers estimated herein from a total of 1200 populations simulated with various designs ranged from 0.06 to 0.66. The estimates of statistical power were influenced by heritability, genetic variance ratio of causal genetic effects and polygenic effects, and the number of causal variants (Fig. 1, Table 1). Especially, the power estimates were greatly increased when there were a small number of causal variants ($P < 0.001$, Table 1). When the number of causal variants decreased from 100 to 20, the power was improved from 0.11–0.36 to 0.45–0.66 with the significance threshold of 10^{-5}. The statistical power also increased with a large genetic variance ratio of causal genetic effects and with larger heritability ($P < 0.05$). However, no difference in the statistical power was found by implementing different numbers of polygenic variants ($P > 0.05$).

The statistical power was further affected by interaction between heritability and the causal variant number ($P < 0.05$, Table 1). In addition, there were significant interaction effects among the three single factors: heritability, the number of causal variants, and the genetic variance ratio ($P < 0.05$, Table 1).

As the genetic variance ratio of causal to polygenic effects decreased, the difference in statistical power between the assigned heritabilities of

* 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.2014.11.001
0888-7543/

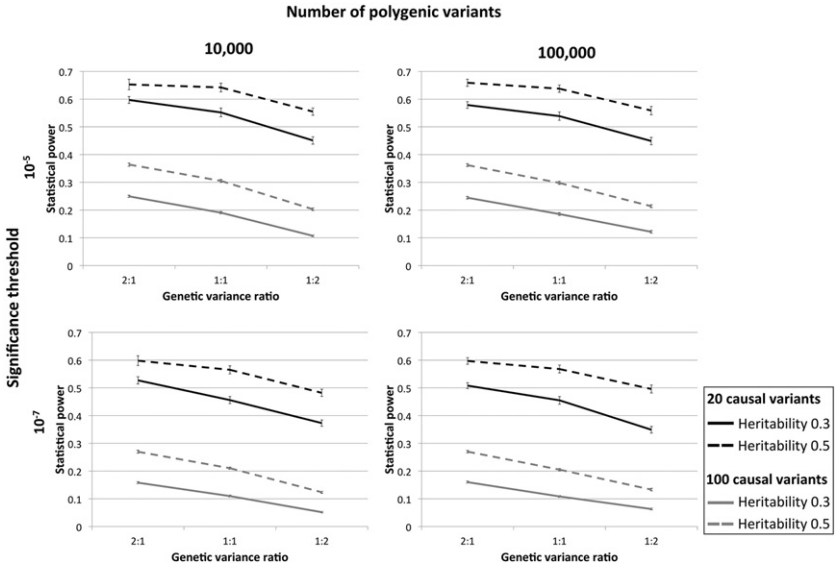

Fig. 1. Statistical power estimates using simulated quantitative traits. Each statistical power estimate is accompanied with a vertical bar representing its standard error, empirically obtained from 50 replicates. The power was estimated using the false positive significance threshold of 10^{-5} (above) or of 10^{-7} (below).

0.3 and 0.5 increased when there were 20 causal variants (Fig. 1). As the genetic variance ratio decreased, the difference in power decreased with the analysis of 100 causal variants.

2.2. Heritability

Heritability estimates considering both causal and polygenic variants did not differ from their input values used in the simulation ($P > 0.05$, Fig. 2). Heritability was found to be biased in some simulation designs when it was estimated using only the causal variants or the polygenic variants ($P < 0.05$). The bias especially increased for heritability when only polygenic variants were considered ($P < 0.05$). It also increased according to the genetic variance ratio ($P < 0.05$). For example, when the ratio of causal to polygenic variance was 2:1, the heritability of polygenic variants was overestimated more with 20 causal variants than with 100 causal variants ($P < 0.05$). However, the overestimation decreased with the ratio of 1:2.

2.3. Heritability of C-reactive protein

Preliminary analysis showed 112 causal variants ($P < 5.0 \times 10^{-4}$) and 9799 polygenic variants ($5.0 \times 10^{-4} < P < 0.05$) to be associated with CRP. They were used to estimate pairwise genetic relationship coefficients for heritability. The heritability estimates were 0.22 ± 0.02 for causal variants and 0.42 ± 0.01 for polygenic variants. The heritability estimated with the simultaneous use of causal and polygenic variants was 0.55 ± 0.01.

Table 1
Factors affecting statistical power of genetic association test in genome-wide association analysis using a mixed model.[a]

Factor	*d.f.*	SS	MS	F	P
CV	1	32.499	32.499	5707.616	<0.001
PV	1	0.077	0.077	1.196	0.274
H	1	2.563	2.563	450.161	<0.001
VR	2	3.470	1.735	304.668	<0.001
CV × H	1	0.071	0.071	12.550	<0.001
CV × H × VR	2	0.073	0.036	6.380	0.002

Abbreviations: CV, number of causal variants; PV, number of polygenic variants; H, heritability; VR, genetic variance ratio of causal to polygenic variants.

[a] Results from analysis of variance (ANOVA). Results for single factors are all presented, but only results of interactive factors demonstrating significance ($P < 0.05$) are presented in this table.

3. Discussion

This simulation study examined the influence of various factors on the statistical power when testing genetic associations with quantitative traits in a mixed model framework. The empirical statistical powers were estimated to range from 0.06 to 0.66. These values were relatively smaller than those obtained from the study of Segura et al. [9] in which they were estimated for an extensive range of significance thresholds (0–1). The smaller powers were observed herein because the current study employed stringent values for the threshold (10^{-5} and 10^{-7}), as used for multiple testing in most GWASs. GWAS may find these power estimates practical in spite of their low levels. The power estimates were affected by heritability, the causal variant number, and the variance ratio of causal genes to polygenes. These were assumed to be essential factors for determining the effect size of the causal variants. Thus, the results concurred with those of previous studies in which statistical powers increased with the increment of individual effect size of variants using the fixed model [10] and mixed model [6]. In fact, heritability, the causal variant number, and variance ratio simulated in the current study were all factors determining the effect size of individual variants. Heritability is a key factor of the effect size. It was generated by using different genetic variance ratios of causal variants to polygenic variants, and the portion of causal genetic variance divided by causal variant number indicated the expected effect size of individual variants.

The statistical power estimates were further influenced by the interaction of individual factors. For example, as the genetic variance ratio of causal to polygenic effects decreased, the difference in the statistical power by heritability increased with a small number (20) of causal variants, but decreased with a large number (100).

For only causal variants, heritability was underestimated in this simulation study. The underestimation was observed when there were 20 causal variants, but not with 100. The number of causal variants might be a critical factor for biasing the heritability. In fact, the bias observed with the smaller number of causal variants might have been caused by the ratio of causal to polygenic SNP numbers. A previous study showed that the increased ratio led to underestimation of the heritability [11].

Meanwhile, the current study showed heritability to be overestimated only for polygenic variants. This indicated that the heritability for polygenic variants partially reflected the genetic variability of the causal variants in linkage with them. Thus, the overestimation increased with a larger number (100,000) of polygenic variants.