Journal
doi.10.1093/journal/xxxxxx
Advance Access Publication Date: Day Month Year
Applications Note

Genetic and population analysis

U-PASS: unified power analysis and forensics for qualitative traits in genetic association studies

Zheng Gao [1,*], Jonathan Terhorst [1], Cristopher Van Hout [2], and Stilian Stoev [1]

[1] Department of Statistics, University of Michigan, Ann Arbor, MI 48109, USA and
[2] Regeneron Genetics Center, Tarrytown, NY 10591 USA.

* To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Summary: Despite efforts to calculate and compare statistical powers of association tests in genetic association studies, there has not been a model-invariant and test-independent analysis. In this work, we develop a unified framework for the analysis of the statistical limits of such tests, and derive the shared asymptotic power limits of common association tests for binary qualitative traits. The results allow us to quantify the fundamental statistical trade-off between risk allele frequency (RAF) and odds ratio (OR). We implement these ideas in a web-based application U-PASS (Unified Power analysis of ASsociation Studies) which calculates and visualizes the quantified OR-RAF diagram. The application also addresses the applicability of asymptotics-based power calculations in finite samples, and provides guidelines for single-SNP-based association tests. In addition to designing prospective studies, U-PASS enables researchers to retrospectively assess the statistical validity of previously reported associations.

Availability and implementation: U-PASS is available as a web-based R Shiny application at https://power.stat.lsa.umich.edu. Source code is available at https://github.com/Pill-GZ/U-PASS.

Contact: gaozheng@umich.edu

Supplementary information: Supplementary data are available in the application.

1 Introduction

The probability of detecting a true association between genetic and phenotype variations, known as statistical power, is influenced by a number of factors such as sample sizes, the statistical test used, the frequency of the risk variant, and magnitude of the effect on the trait. Power analysis, which determines the suitable factor combinations in order to achieve sufficient statistical power, plays an important role in determining study designs (Skol *et al*., 2003; Goodwin *et al*., 2005), and in interpreting published findings (Ioannidis, 2005).

There has been a number of widely used calculators for genome-wide association studies (GWAS). Early work by Sham (1998) enabled power analysis of likelihood ratio tests for associations between marker SNPs and quantitative or qualitative traits; the results were implemented in GPC (Purcell *et al*., 2003). Skol *et al*. (2003) studied the performance of two-sample t-tests, and extended the analysis to two-stage designs; the results were implemented in the CaTS caclulator, and later, in the GAS calculator for one-stage designs (Johnson *et al*., 2017). Independently, Menashe *et al*. (2008) implemented the calculations for one-stage designs in the PGA calculator. Recent works have also studied power of a number of SNP-set based tests targeting rare variants (Wang *et al*., 2014; Derkach *et al*., 2017). See Sham & Purcell (2014) for a review.

Despite these efforts, existing power calculators can be difficult to use in practice, for several reasons:

1. Lack of universality. Existing power analyses are tied to the underlying models and the statistical procedures used; power calculations for a certain model-method combination may not be valid if either the model or the method changes. In principle, power calculations based on likelihood ratio tests or t-tests cease to hold for studies running logistic regressions or chi-squared tests. Users are burdened with matching the appropriate tool to the specific type of analysis they wish to perform. This is complicated by the fact that the precise test and model assumptions are rarely made explicit in the existing calculators.

2. Inconsistent definitions of key quantities. While GWAS catalogs, e.g., NHGRI-EBI (MacArthur *et al*., 2016), require studies to report risk allele frequency (RAF) *in the control group*, all of the aforementioned power calculators assume the RAF input as the frequency *in the general population*. These quantities are not necessarily equal, and using one in place of the other may grossly distort power estimates.

3. Inaccuracies in finite samples. While existing tools rely on large-sample approximations in their power calculations, these approximations are not reliable in finite samples when genetic variants are rare. Existing calculators are silent about the applicability of asymptotics-based approximations, and how they should be corrected.

As a result, it is not only challenging to use the existing power calculation tools for planning genetic association studies correctly, but

© The Author 2019. Published by XXXXXXXXXXXX.

also difficult to systematically review the statistical validity of findings reported in the literature, since different models and tests must be handled differently, and with care.

In an effort to address these difficulties and deficiencies, we propose a unified framework for power analysis of single variant association studies. By abstracting abstract away the assumptions of disease models and testing procedures which may vary from study to study, we reduce the problem to the essential quantities that are invariant to nuisance parameters. These ideas are implemented in the software U-PASS, enabling model-invariant, test-independent power analysis, as well as systematic reviews of the statistical validity of reported findings.

We briefly summarize the important features and uses of the software below. Mathematical details and results from numerical experiments are collected in the Supplement.

2 Methods and features

2.1 A model-invariant parametrization

We identify the core parameters common to models of qualitative traits,

- Sample sizes, i.e., the number of Cases, n_1, and Controls, n_2,
- Conditional distribution of risk variant among Controls, i.e., risk allele frequency (RAF) in the Control group, denoted f.
- Odds ratio (OR) of genetic variants, denoted R.

An alternative hypothesis, e.g., a disease model, determines the set of quantities either implicitly or explicitly, and therefore determines statistical power for a given test procedure. Here, statistical power is calculated directly based the quartet (n_1, n_2, f, R), allowing us to perform power analysis valid for studies employing different models.

We make the important distinction between RAF in the Control group (f), versus RAF among all subjects in the study, and RAF in the general population. Throughout this work and the software, RAF refers to the risk allele frequency in the Control group, consistent with the reporting standard of the NHGRI-EBI Catalog.

2.2 A test-independent power analysis

While power calculations are necessarily tied to the statistical tests used, for many common association tests, statistical powers are asymptotically equivalent. We show in the Supplement that the likelihood ratio test, chi-square test, Welch's t-test, and LR test for logistic regressions have asymptotically the same power, as long as the core parameters assume the same values. Our model-invariant, test-independent analysis allows us to calculate power in a unified fashion. The formulas used for power calculations in terms of the core parameters are detailed in Section 1 of the Supplementary material.

Our software only requires users to specify the number of cases and controls. The common power limits are calculated as a function of RAF and OR, given numerically, and visualized as a heatmap in the OR-RAF diagram; see Figure 1 for a preview of the software graphical user interface.

2.3 Review and forensics of reported findings

This unified treatment allows us to examine results from different studies in the same diagram, even when they do not employ the same model or statistical test. This enables systematic reviews of reported findings for their statistical validity. In particular, a reported association predicted to have low power given the study's sample size – lying in the dark regions of the OR-RAF diagram – while not impossible, invites further scrutiny.[1]

[1] It should be noted that a reported association predicted to have high power is not automatically accurate, as winner's curse induced by multiple testing may inflate the OR and RAF estimates (Xiao & Boehnke, 2009).

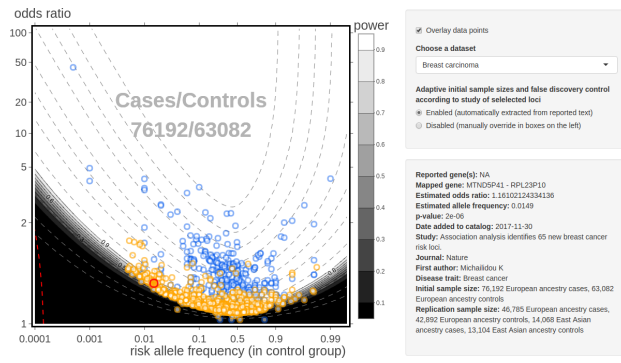

Fig. 1. Snapshot of the application's user interface, displaying reported associations from breast cancer studies in the NHGRI-EBI GWAS Catalog (circles). The findings are overlaid on the OR-RAF power diagram of association tests (greyscale heatmap). The initial sample sizes are dynamically adjusted, and automatically determined from texts of the article reporting the user selected loci (red circle). Information of the selected loci and the articles are also dynamically displayed; findings reported in the same article are highlighted (orange circles). We provide finite sample corrections by marking the rare variant region(s) where asymptotic approximations do not apply (red dashed lines, lower-left). The majority of the published findings we surveyed exhibited a striking level of concordance with our theoretical predictions, with most associations congregating just inside the detectable region.

Studies where reported associations show misalignment with the predicted powered curves may be further investigated for potential problems in the data curation process. We reached out to one study where gross misalignment was identified (Dominguez-Cruz *et al.*, 2014). The authors of the study confirmed that this was due to a data duration problem that occurred when their findings were added to the GWAS Catalog (Dominguez Cruz, personal communication).

The software provides options for users to load and overlay findings reported in the NHGRI-EBI GWAS Catalog, or upload data from other sources compliant with the Catalog's data format.

2.4 Rare variants and finite sample corrections

We address the quality of asymptotic approximations in our power analysis, as well as the applicability of single variant tests when rare genetic variants are present. Specifically, we provide a lower bound on the variant counts needed to calibrate Fisher's exact test. If variant counts fall below the threshold, exact tests, and by extension, single-SNP-based association tests, cannot be correctly calibrated to have the desired type I error rate without sacrificing all statistical power. In such cases, the asymptotic approximations do not apply. We mark this low-count, low-power region on the OR-RAF diagram. We refer readers to Section 2 in the Supplement for further details.

The software also provides options for users to specify the rare variant as 1) having less than a specified counts, or 2) occurring in less than a percentage of all subjects in the study, as is customary in the literature.

3 Implementation

U-PASS is implemented as an interactive web-based R Shiny application, hosted at https://power.stat.lsa.umich.edu, open to the public. It has been designed to be user-friendly and self-contained. Source code is freely available at https://github.com/Pill-GZ/U-PASS.

Funding

This work is partially supported by NSF Grant DMS-1830293, Algorithms for Threat Detection.

References

Derkach, A., Zhang, H., & Chatterjee, N. (2017). Power Analysis for Genetic Association Test (PAGEANT) provides insights to challenges for rare variant association studies. *Bioinformatics*, **34(9)**, 1506-1513.

Domínguez-Cruz, M. G., de Lourdes Muñoz, M., Totomoch-Serra, A., García-Escalante, M. G., BurgueÃ±o, J., Valadez-González, N., ... & Díaz-Badillo, Á. (2018). Pilot genome-wide association study identifying novel risk loci for type 2 diabetes in a Maya population. *Gene*, **677**, 324-331.

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, **17(6)**, 333.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine*, **2(8)**, e124.

Johnson, J. L., & Abecasis, G. R. (2017). GAS Power Calculator: web-based power calculator for genetic association studies. *bioRxiv*, 164343.

MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., ... & Pendlington, Z. M. (2016). The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic acids research*, **45(D1)**, D896-D901.

Menashe, I., Rosenberg, P. S., & Chen, B. E. (2008). PGA: power calculator for case-control genetic association analyses. *BMC genetics*, **9(1)**, 36.

Purcell, S., Cherny, S. S., & Sham, P. C. (2003). Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics*, **19(1)**, 149-150.

Sham, P. C. (1998). *Statistics in human genetics*. Wiley.

Sham, P. C., & Purcell, S. M. (2014). Statistical power and significance testing in large-scale genetic studies. *Nature Reviews Genetics*, **15(5)**, 335.

Skol, A. D., Scott, L. J., Abecasis, G. R., & Boehnke, M. (2006). Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nature genetics*, **38(2)**, 209.

Wang, G. T., Li, B., Lyn Santos-Cortez, R. P., Peng, B., & Leal, S. M. (2014). Power analysis and sample size estimation for sequence-based association studies. *Bioinformatics*, **30(16)**, 2377-2378.

Xiao, R., & Boehnke, M. (2009). Quantifying and correcting for the winner's curse in genetic association studies. *Genetic Epidemiology*, **33(5)**, 453-462.