The paltry power of priors versus populations

Jianan Zhan[1], Dan E. Arking[2], Joel S. Bader[1,3*]

1 Department of Biomedical Engineering, School of Medicine, Johns Hopkins University, Baltimore, MD 21205
2 Institute of Genetic Medicine, School of Medicine, Johns Hopkins University, Baltimore, MD 21205
3 Department of Biomedical Engineering, School of Engineering, Johns Hopkins University, Baltimore, MD 21218

* joel.bader@jhu.edu

Abstract

Biological experiments often involve hypothesis testing at the scale of thousands to millions of tests. Alleviating the multiple testing burden has been a goal of many methods designed to boost test power by focusing tests on the alternative hypotheses most likely to be true. Very often, these methods either explicitly or implicitly make use of prior probabilities that bias significance for favored sets thought to be enriched for significant finding. Nevertheless, most genomics experiments, and in particular genome-wide association studies (GWAS), still use traditional univariate tests rather than more sophisticated approaches. Here we use GWAS to demonstrate why unbiased tests remain in favor. We calculate test power assuming perfect knowledge of a prior distribution and then derive the population size increase required to provided the same boost without a prior. We show that population size is exponentially more important than prior, providing a rigorous explanation for the observed avoidance of prior-based methods.

Author summary

Biological experiments often test thousands to millions of hypotheses. Gene-based tests for human RNA-Seq data, for example, involve approximately 20,000; genome-wide association studies (GWAS) involve about 1 million effective tests. The conventional approach is to perform individual tests and then apply a Bonferroni correction to account for multiple testing. This approach implies a single-test p-value of 2.5×10^{-6} for RNA-Seq experiments, and a p-value of 5×10^{-8} for GWAS, to control the false-positive rate at a conventional value of 0.05. Many methods have been proposed to alleviate the multiple-testing burden by incorporating a prior probability that boosts the significance for a subset of candidate genes or variants. At the extreme limit, only the candidate set is tested, corresponding to a decreased multiple testing burden. Despite decades of methods development, prior-based tests have not been generally used. Here we compare the power increase possible with a prior with the increase possible with a much simpler strategy of increasing a study size. We show that increasing the population size is exponentially more valuable than increasing the strength of prior, even when the true prior is known exactly. These results provide a rigorous explanation for the continued use of simple, robust methods rather than more sophisticated approaches.

Introduction

Genomics experiments involve testing thousands to millions of hypotheses. In functional genomics and proteomics, each gene or protein usually corresponds to a single test, with 20,000 or more tests required for an RNA-Seq or proteomics experiment. In human genetics, the number of independent tests accounting for linkage disequilibrium in a single ethnicity is usually assumed to be about 1 million. To maintain a family-wise error rate (FWER) controlled at 0.05, a long-standing approach has been to apply a Bonferroni correction, requiring a single-test p-value of 0.05 divided by the number of hypotheses tested. This multiple-testing correction from this stringent approach is seen as a burden for identifying genome-wide significant findings.

A current direction of GWAS is to incorporate prior knowledge about functional effects of SNPs, in order to increase the power to detect SNPs with true associations or to identify which SNP in an linkage disequilibrium (LD) region is most likely to be the causal variant [1–4]. A representative approach incorporated 450 different annotations into GWAS analysis of 18 human traits; the number of loci with high-confidence associations was increased by around 5% [5]. Despite the intuitive value of incorporating pre-existing biological knowledge, it remains unclear whether this roughly 5% increase in genome-wide significant findings is the best that could be obtained, and additionally whether the increase comes at the cost of false negatives for true positives that lack similar annotations.

Other groups, including our own, have developed methods that incorporate priors based on patterns learned from the data [6–8]. These patterns may include multiple independent effects found within a single genes, or patterns of pleiotropic variants that contribute to a shared subset of traits in a multi-phenotype data set. While these methods have value in providing a clearer view of genetic architecture than available through univariate tests, the number of new significant findings has been small [4, 9].

Still other methods introduce prior distributions for model parameters, or equivalently regularizations, which implicitly define a prior favoring candidate variants with the largest observed effects. These methods have usually not been used in practice for GWAS because the computational expense has not been justified by improved results.

In this paper, we use theoretical models and derivations to investigate into the dependency of power on population size and incorporating priors. We consider two types of priors: hard prior, which is an idealized prior that only a fraction of total hypotheses are tested; soft prior, for which all hypotheses are divided into two classes, and a higher prior value is given to the favored class which is believed to be enriched with true associations. For hard prior, we proved analytically that the dependence of power on population size is linear, whereas the dependence on prior strength is logarithmic, which indicates the importance of having larger population size over bigger prior strength when doing association tests. For soft prior, we provide numerically exact results showing that the power gains for the favored class are large only for limited circumstances; the gains for the favored class imply power loss for the non-favored class, and the average power gain considering both classes is only 5-10%. These gains require exact knowledge of the true priors; in practice, gains with estimated priors should be smaller.

Methods

Hypothesis testing

We consider tests of association between a feature of the data, x, and an observed phenotype or response variable, y, assumed to be scalars for simplicity. For a population of size N, these are aggregated into vectors $\mathbf{x}$ and $\mathbf{y}$. An association test compares a null model M_0, to an alternative, M_1, which for a linear model takes the form

$$\begin{aligned} M_0 &: y \sim \mathrm{Norm}(\mu_0, \sigma_0^2); \\ M_1 &: y \sim \mathrm{Norm}(\mu_1 + \beta x, \sigma_1^2). \end{aligned}$$

One such M_1 exists for each possible feature to be tested. With A total possible alternatives to be tested, these could be denoted $\{M_a\}$, $a \in \{1, 2, \ldots, A\}$. We consider one such alternative at a time and for simplicity denote it M_1. Model parameters are $\Theta_0 = \{\mu_0, \sigma_0^2\}$ for the null model and $\Theta_1 = \{\mu_1, \sigma_1^2, \beta\}$ for the alternative model. These models correspond to a null hypothesis H_0 and alternative hypothesis H_1,

$$\begin{aligned} H_0 &: \beta = 0; \\ H_1 &: \beta \neq 0. \end{aligned}$$

For nested models, the hypothesis test is usually performed by a likelihood ratio test or its equivalent. Denote the maximum likelihood parameters as $\hat{\Theta}_0$ and $\hat{\Theta}_1$, and assume independence of the model and data. A test statistic τ is defined as

$$\begin{aligned} \tau &= 2 \ln \frac{\Pr(\mathbf{y}|\mathbf{x}, \hat{\Theta}_1) \Pr(M_1)}{\Pr(\mathbf{y}|\mathbf{x}, \hat{\Theta}_0) \Pr(M_0)} \\ &= q^2 + 2 \ln[\Pr(M_1)/\Pr(M_0)]. \end{aligned} \tag{1}$$

According to Wilks' Theorem, under the null hypothesis, q^2 is a random variable distributed as χ_1^2, or more generally as a χ_d^2 random variable where the null model is nested inside an alternative model with d additional parameters [10]. Under the alternative hypothesis, q^2 is distributed as a non-central χ^2 with non-centrality parameter q_1^2,

$$q_1^2 = NR^2/(1 - R^2), \tag{2}$$

where R^2 is the fraction of variance explained by the alternative hypothesis, and $1 - R^2$ is the residual fraction of variance.

For a conventional test, the prior $\Pr(M)$ is identical for the null and each alternative; it does not contribute to the test statistic. To control the type I error (false-positive rate) at family-wise error rate FWER α, the Bonferroni method requires a single-test p-value of α/A for A total tests. Define the quantile of the uniform normal distribution corresponding to a two-tailed test at this stringency z_I. More formally, if $\Phi(z)$ is the cumulative lower tail probability distribution for standard normal random variable z, then $\Phi(-z_I) = \alpha/2A$. For true effect q_1, the power is $\Phi(|q_1| - z_I)$, or equivalently

$$(z_I - z_{II})^2 = \frac{NR^2}{1 - R^2}. \tag{3}$$

This key expression relates the type I error (false-positive rate), the type II error (false-negative rate or complement of power), the population size N, and the effect size R^2.

Hard prior

A hard prior is an idealized prior in which only hypotheses corresponding to a faction $1/S$ of the total are tested. Larger S corresponds to a stronger prior. For 20,000 gene-based tests, testing 10% of the total corresponds to $S = 10$, and testing 20 genes corresponds to $S = 1000$. Realistically, priors stronger than $S = 100$, corresponding to 200 genes tested, are unlikely.

The effect of a hard prior is to reduce the multiple-testing burden. To maintain FWER α, each two-tailed test is performed at stringency $S\alpha/2A$ rather than $\alpha/2A$. This reduces the quantile z_I required for significance and increases the power to detect an association with a smaller effect R^2. Equivalently, Eq. 3 can be solved for R^2 to calculate the critical effect size to achieve desired power at stated type I error,

$$R^2 = \frac{(z_I - z_{II})^2}{N + (z_I - z_{II})^2}. \tag{4}$$

The effect of a hard prior on z_I may also be estimated analytically. A steepest descents approximation relates the quantile $z > 0$ to its upper-tail area ϵ,

$$\begin{aligned}
\epsilon &= (2\pi)^{-1/2} \int_z^\infty du e^{-u^2/2} \\
&= (2\pi)^{-1/2} e^{-z^2/2} \int_z^\infty du e^{-(u+z)(u-z)/2} \\
&\approx (2\pi)^{-1/2} e^{-z^2/2} \int_z^\infty du e^{-z(u-z)} \\
&= \frac{1}{\sqrt{2\pi}z} e^{-z^2/2}.
\end{aligned}$$

Equivalently,

$$z^2 \approx -2\ln[\sqrt{2\pi}z\epsilon].$$

In terms of the quantile z_I for prior strength S and a two-tailed test, we have approximately

$$z_I^2 \approx -2\ln[\sqrt{2\pi}z_I S\alpha/A]. \tag{5}$$

Define ζ as the value of z_I for no prior, $S = 1$, with $\Phi(-\zeta) = \alpha/2A$ and

$$\zeta^2 \approx -2\ln(\sqrt{2\pi}\zeta\alpha/A). \tag{6}$$

For GWAS with a p-value threshold of 5×10^{-8}, $\zeta = 5.45$ and $\zeta^2 = 29.7$. Because the dependence of Eq. 5 on $\ln z$ is weak, we replace $\ln z$ with $\ln \zeta$,

$$z_I^2 \approx -2\ln[\sqrt{2\pi}\zeta S\alpha/A] \approx \zeta^2 - 2\ln S = \zeta^2(1 - 2\zeta^{-2}\ln S).$$

Keeping terms of order $1/\zeta$,

$$\begin{aligned}
z_I &\approx \zeta(1 - \zeta^{-2}\ln S) \\
z_I - z_{II} &\approx \zeta - z_{II} - \zeta^{-1}\ln S \\
(z_I - z_{II})^2 &\approx (\zeta - z_{II})^2 - \frac{2(\zeta - z_{II})}{\zeta}\ln S \\
&= (\zeta - z_{II})^2 \left[1 - \frac{2}{\zeta(\zeta - z_{II})}\ln S\right].
\end{aligned}$$

According to Eq. 3, the critical effect size depends only on the ratio $(z_I - z_{II})^2/N$. Consider two scenarios with equal critical effect size, one with population size N_1 and

prior strength S_1, and the second with population size N_2 and prior strength S_2. For these to have equal critical effect size,

$$(\zeta - z_{II})^2 \left[1 - \frac{2}{\zeta(\zeta - z_{II})} \ln S_1\right] / N_1 \approx (\zeta - z_{II})^2 \left[1 - \frac{2}{\zeta(\zeta - z_{II})} \ln S_2\right] / N_2.$$

Cancelling constant terms $\zeta - z_{II}$ and noting that $2\zeta^{-1}(\zeta - z_{II}) \ln S$ is small,

$$\begin{aligned} \frac{N_1}{N_2} &\approx \left[1 - \frac{2}{\zeta(\zeta - z_{II})} \ln S_1\right] / \left[1 - \frac{2}{\zeta(\zeta - z_{II})} \ln S_2\right] \\ &\approx \left[1 - \frac{2}{\zeta(\zeta - z_{II})} \ln S_1\right] \times \left[1 + \frac{2}{\zeta(\zeta - z_{II})} \ln S_2\right] \\ &\approx 1 + \frac{2}{\zeta(\zeta - z_{II})} \ln \frac{S_2}{S_1}. \end{aligned}$$

The dependence on population size is linear, whereas the dependence on prior strength is logarithmic. Equivalently, population size is exponentially more important that prior strength. Again for GWAS with z_{II} selected for 80% power, $\zeta(\zeta - z_{II})/2 = 17.15$, and only a small fractional population increase is required to obtain the equivalent power increase for a strong prior. An extremely strong prior with $S_2 = 1000$, with effectively only 20 genes selected for testing, can be matched by a population increase of about 40%.

Contours of N and S with equal critical effect size can be estimated by returning to the approximate result

$$NR^2/(1 - R^2) \approx (\zeta - z_{II})^2[1 - \frac{2}{\zeta(\zeta - z_{II})} \ln S].$$

Noting that for small ϵ, $1 + \epsilon \ln S \approx S^{\epsilon}$, contours are given by

$$NS^{2/\zeta(\zeta - z_{II})} \approx (\zeta - z_{II})^2 R^2/(1 - R^2). \quad (7)$$

On a log-log plot of $\log S$ versus $\log N$, these contours would have steep negative slope equal to $-\zeta(\zeta - z_{II})/2$.

Soft prior

Soft priors are incorporated into association analysis such that sequence variants like loss-of-function and missense variants, which are more likely to affect protein function and therefore more likely to be causative, are given higher prior belief to have true signals before data was analyzed. For simulation, this is done by first dividing all possible associations into two classes, a favored class F and a non-favored class NF, and assuming true associations are enriched in the favored class and depleted in the non-favored class. For Eq. 1, if we normalize $\Pr(M_0) = 1$, the expression could be simplified into the following form:

$$\tau = q^2 + 2 \ln [\Pr(M_1)] \quad (8)$$

Denote the model for variants in the favored class as M_F and in the non-favored class as M_{NF}, if priors for the two classes $Pr(M_F)$ and $Pr(M_{NF})$ are known exactly, the probability distribution for the test statistic will now depend on the classes, with their individual test statistics being:

$$\tau_F = q^2 + 2 \log [\Pr(M_F)]$$

$$\tau_{NF} = q^2 + 2\log\left[\Pr(M_{NF})\right]$$

Based on assumptions given above, we are able to simulate the change of power in detecting real effects after incorporating the two classes. For simulation, we first fixed the association effective size corresponding to 50% power to detect a true association in a GWAS study at genome-wide significance ($p = 5 \times 10^{-8}$) assuming 1 million effective tests. Using the same notation as above, this is equivalent to solving for first the critical value τ_c such that

$$\Phi(-\tau_c) = 5 \times 10^{-8}/2 = 2.5 \times 10^{-8}$$

then solve for true effect q_c such that $\Phi(|q_c| - \tau_c) = 0.5$, which gives us $|q_c| = \tau_c$. To avoid calculation on both tails of the standard normal distribution, the expression could be simplified by introducing $\Psi(z^2; q^2)$ to denote $\Pr(t > z^2)$ where t follows a $1df$ non-central χ^2 distribution with non-centrality parameter q^2. The relationship between Φ and Ψ thus satisfies

$$2 \times \Phi(z - |q|) = \Psi(z^2; q^2)$$

for $z < |q|$ and

$$2 \times \Phi(|q| - z) = \Psi(z^2; q^2)$$

for $z \geq |q|$.

Power to detect SNPs with true associations for two classes combined could then be calculated as a function of two variables:

1. S, power strength, which is defined as inverse of the fraction of variants in the favored class;
2. $[\Pr(M_F)/\Pr(M_{NF})]$ the relative priors of the two classes.

The exact steps of simulation are as following:

1. For a pair of values for fraction of variants in the favored class 1/S and prior enrichment fold-enrichment $[\Pr(M_F)/\Pr(M_{NF})]$, the critical threshold for genome-wide significance τ_c' could be solved using the following equation:

$$(1 - 1/S)\Pr(q^2 > \tau_c') + (1/S)\Pr(q^2 + 2\ln\left[\frac{\Pr(M_F)}{\Pr(M_{NF})}\right] > \tau_c') = 5 \times 10^{-8}$$

   With q^2 following an $1df$ χ^2 distribution, with the denotations defined above, this equation simplifies to:

$$(1 - 1/S)\Psi(\tau_c'; 0) + (1/S)\Psi(\tau_c' - 2\ln\left[\frac{\Pr(M_F)}{\Pr(M_{NF})}\right]; 0) = 5 \times 10^{-8} \quad (9)$$

2. Now with the critical threshold τ_c' and the non-centrality parameter q_c calculated above, power for the favored class could be calculated as:

$$power(M_F) = \Psi(\tau_c' - 2\ln\left[\frac{\Pr(M_F)}{\Pr(M_{NF})}\right]; q_c^2) \quad (10)$$

   And the power for the non-favored class could calculated as:

$$power(M_{NF}) = \Psi(\tau_c'; q_c^2) \quad (11)$$

3. The average power for true associations could be calculated as:

$$\begin{aligned} power(Avg) = \quad & \frac{\Pr(M_F)}{\Pr(M_F) + (S-1)\Pr(M_{NF})} \times power(M_F) \\ & + \frac{(S-1)\Pr(M_{NF})}{\Pr(M_F) + (S-1)\Pr(M_{NF})} \times power(M_{NF}) \end{aligned} \quad (12)$$

4. Population size fraction increase to achieve the same average power could then be calculated as $N_1/N_2 = q'_c/q_c$, where q'_c correspond to the same average prior without using any prior. Here N_2 = population size to achieve the specific power and 0.05 FWER using a prior, and N_1=population size to achieve the same power and 0.05 FWER without a prior. This is similar to the exploration between S and N in the hard prior case.

Results

Hard prior

Fig. 1 shows contours for critical R^2 for $p = 5 \times 10^{-8}$ at power = 0.8 as a function of prior strength and population size. As could be observed, the color corresponding to critical R^2 changes rapidly as population size changes, and doesn't change much as a function of prior strength. This indicates that power is much more sensitive to population size compared to using a hard prior, namely restricting tests to a subset of variants. On the log-log scale, given a fixed value of R^2, the prior strength and population size exhibits a clear linear pattern, which leads to the derivation on the Hard prior part in the Methods section, and yielding a slope of 17.15 at 80% power. Fig. 2 shows the analytical solution of the relationship between population size and prior strength. The left panel denotes the population increase ratio versus prior strength increase to achieve the same increase in power. If we have a prior strength of 100 for example, which correspond to testing 1/100 of all variants, we could get the same power increase by increasing the population from N to fN where f is the factor increase. Reading from the figure, this would be a factor about 1.35, which correspond to a 33% increase in the cohort size. At the very extreme, a prior strength of 1000, which corresponds to testing SNPs in only 20 genes, will only do as well as increasing the cohort size by 70%. Panel on the right shows the linear relationship between population size exponent and prior strength for a fixed effect size, which quantifies the linear relationship as described in Eq. 7. Both figures provide numerically exact result of the relationship between population size and prior strength, with the conclusion that population size is a much more important factor to gain power than incorporating priors.

Soft prior

Power for the two classes

Fig. 3 shows power for the two classes at different fold-enrichment for favored class versus prior strength. For power of the favored class, when prior strength is small, corresponding to a large fraction of favored class, power is not sensitive to fold enrichment. For example, when $S = 2$, corresponding to the favored class consisting 50% of the total variants, power boost for the favored class is at around 2% to 4% regardless of fold enrichment. This is because, a large favored class fraction is essentially equivalent to a less well-defined subset of variants, which usually fails to provide much valuable information regarding prior beliefs. Therefore, giving the big favored class a higher prior value only results in the decreased power for the non-favored class, as is shown in the right panel of Fig. 3.

As S increases, which corresponds to decreasing the fraction of favored class, power becomes more sensitive to fold enrichment, and assigning it with a bigger prior enhances its power. Specifically, a prior 2 to 5 folds as big as the non-favored class prior gives a 5% to 10% power boost. This power increases even more as the fold-enrichment enhances, as long as the fraction is fixed at the same level. This is because, as the favored class gets smaller, the subset of variants becomes more informative, thus giving

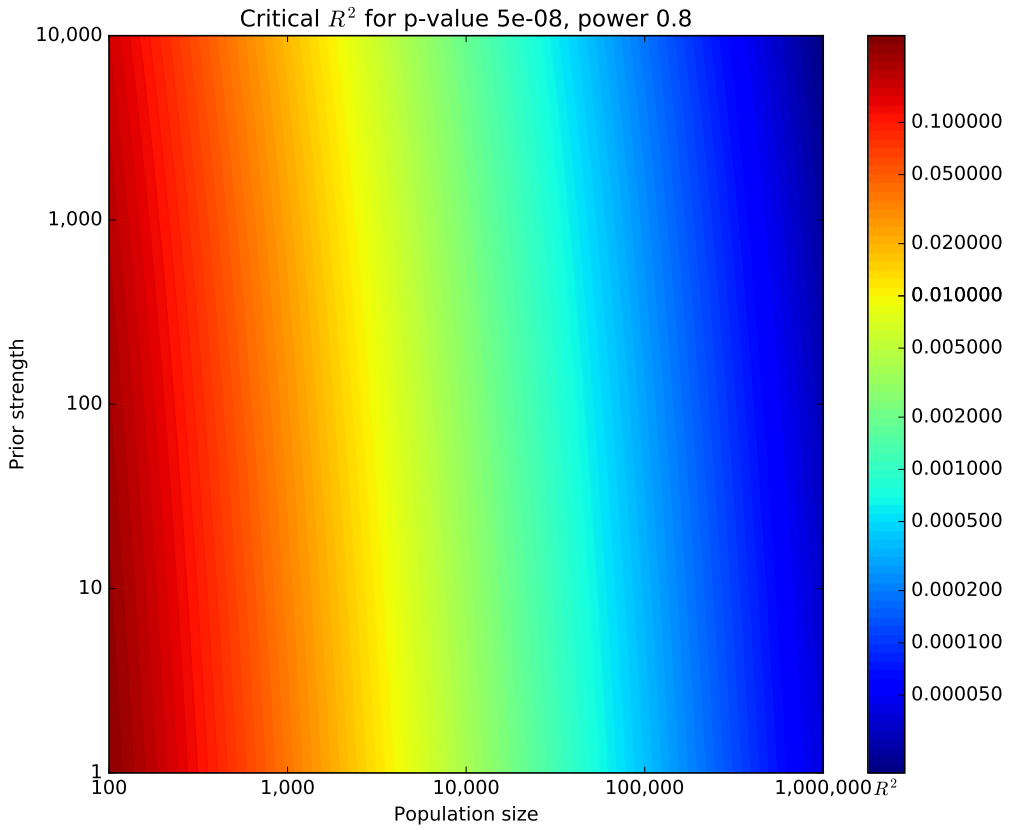

Fig 1. Contour plot for critical R^2 for $p = 5 \times 10^{-8}$ at power $= 0.8$ as a function of prior strength and population size. The color corresponding to critical value R^2 changes rapidly as population size changes, and doesn't change much as a function of prior strength, indicating that power is much more sensitive to population size compared to restricting tests to a subset of variants. Relationship between prior strength and population size appears to follow a linear relationship on the log-log scale, and the slope of the linear relationship given a fixed value of R^2 indicates the relative importance of the two factors.

the class higher prior greatly boosts its power. From the calculation perspective, as S becomes bigger, τ_c' term in Eq. 9 is mostly determined by the non-favored class and remains mostly unchanged; thus little change to the fold enrichment would result in big difference of the power of favored class.

Also note that, since the overall power is fixed at 50%, gain in power for the favored class implies a loss of power for the non-favored class, which corresponds to the "no-free-lunch theorem". Power for the non-favored class remains around 50% when fraction of favored class is low or fold enrichment is low, as in these two situations the impact of the favored class is small; its power decreases, yielding more power for the favored class, when both fold enrichment is high and fraction of favored class is high.

Average power for combining both classes

Average power for combining both classes is shown in Fig. 4. For same fold-enrichment and class fraction, the average power is relatively smaller than power for the favored class alone, and higher than the non-favored class alone. This is because gain for power of the favored class leads to loss for the non-favored class, as discussed above, thus a combination of the two will result a value of power in the middle.

Shape of the contour is determined by weights of the two classes: for small prior strength and large fold enrichment, shape of contour is more similar to the favored class in Fig. 3; for large prior strength and small fold enrichment, shape is similar to the non-favored class. This explains the curve which tilts up towards large prior strength, as when fraction for favored class is small, the power is mostly determined by the un-favored class. Fig. 5 shows the population size fraction increase in order to achieve the same power increase. As could be observed, the maximum population increase is 1.3 fold to obtain the maximum power gain fulfilled through incorporating a prior, further strengthening the conclusion from the hard prior part that, population size is of a more crucial factor compared to prior incorporation as for association involving large number of hypothesis testings.

Discussion and Conclusion

Despite the efforts on developing methods that incorporate priors into association hypothesis tests, traditional unbiased univariate tests combined with Bonferroni correction to control for FWER remains the rule of thumb method to test for associations. In this paper, we exploited the relationship of power to detect true associations on increasing study size and incorporating priors. Two scenarios were considered in this study: hard prior, for which only a fraction of all variants are tested to lower the burden coming from multiple testing; soft prior, for which a fraction of variants are given a higher belief a prior to doing the association analysis. For hard prior, the dependence of heritability on population size and prior strength was analytically derived, and it was proved that the dependence on population size is linear, whereas the dependence on prior strength is logarithmic. Soft prior was able to boost power with very specific requirement on class fraction and fold-enrichment, and even so, its maximum boost of power could be achieved by increasing population size by approximately 30%. For both scenarios, it was concluded that increasing population size is a better strategy to boost power compared to incorporating priors. With recent developments in high throughput biology, immense amount of data is being generated, making improving power through increasing population size possible; in the meantime, a lack of prior-based methods with extraordinary performance on association tests has been observed, which further strengthen the favor of population size from an practical perspective. These results give valuable insights into what strategy should be taken

towards future directions for establishing associations between biological variants and traits.

The usefulness of having larger sample size is not restricted to association tests in the biological field. Huge success has been achieved in the application of machine learning and deep learning into image analysis, nature language processing fields and etc. While big credit has been given to the design and implementation of sophisticated learning structures like convolution neural networks, recurrent neural networks and etc., from results in this project, the role that the tremendous amount of data for training these networks might have been well under-estimated.

References

1. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential Etiologic and Functional Implications of Genome-Wide Association Loci for Human Diseases. Proceedings of the National Academy of Sciences. 2009;106(23):9352–9367.

2. Schork AJ, Thompson WK, Pham P, et al. All SNPs Are Not Created Equal: Genome-Wide Association Studies Reveal a Consistent Pattern of Enrichment among Functionally Annotated SNPs. PLOS Genetics. 2013;9(4):e1003449.

3. Petersen A, Spratt J, Tintle NL. Incorporation prior knowledge to increase the power of genome-wide association studies. Methods in Molecular Biology. 2013;1019:519–541.

4. Sveinbjornsson G, Alberechtsen A, Zink F, Gudjonsson SA, Oddson A, et al. Weighting sequence variants based on their annotation increases power of whole-genome association studies. Nature Genetcis. 2016;48(3):314–317.

5. Pickrell JK. Joint Analysis of Functional Genomic Data and Genome-Wide Association Studies of 18 Human Traits. American Journal of Human Genetics. 2014;94(4):559–573.

6. Chanda P, Huang H, Arking DE, Bader JS. Fast Association Tests for Genes with FAST. PLoS ONE. 2013;8(7):e68585.

7. Huang H, Chanda P, Alonso A, Bader JS, Arking DE. Gene-Based Tests of Association. PLoS Genet. 2011;7(7):e1002177.

8. Zhan J, Arking DE, Bader JS. Discovering patterns of pleiotropy in genome-wide association studies. bioRxiv. 2018;doi:10.1101/273540.

9. Meyer VH, Francesco PC, Oliver S, Ewan B. LiMMBo: a simple, scalable approach for linear mixed models in high-dimensional genetic association studeis. bioRxiv. 2018;doi:10.1101/255497.

10. Wilks SS. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypothesis. The Annals of Methmatical Statistics. 1938;9:60–62.

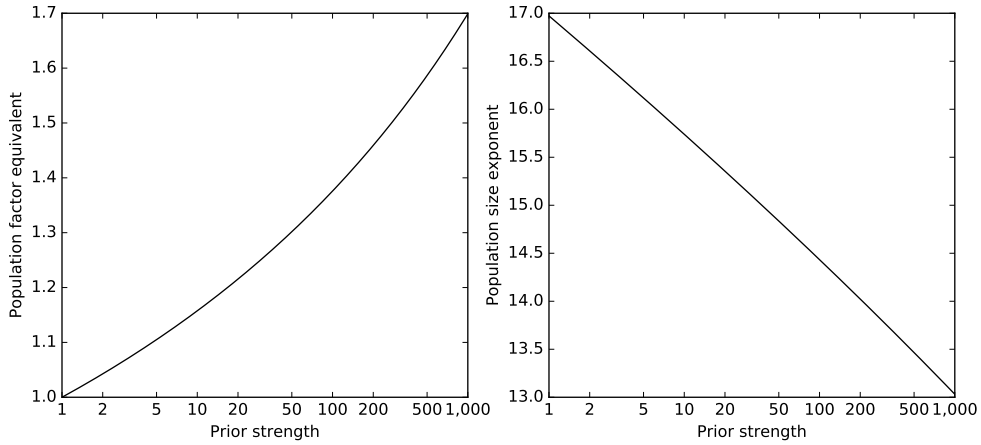

Fig 2. Population size and prior strength to achieve same power. X-axis denotes the prior strength S; y-axis of the left panel denotes population fraction increase; y-axis of the right panel denotes population size exponent for a fixed effect value. The left panel shows the population size and prior strength to achieve the same power; a prior strength at 1000 will have equivalent power if the cohort size increases by 70%. The right panel shows the relation between population size exponent and prior strength given a fixed effect size. Both figures shows numerically the linear relationship between population size and logarithmic of prior strength.

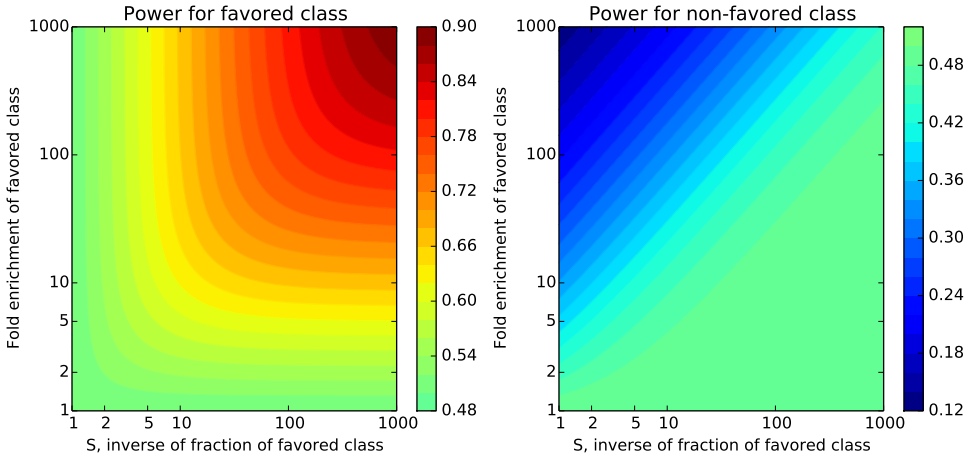

Fig 3. Contour plot of power for the favored class and non-favored class at $p = 5 \times 10^{-8}$ threshold for 50% power. The left panel shows power for the favored class, and the right panel shows power for the non-favored class. X-axis denotes prior strength S, which is equal to inverse of fraction of SNPs in the favored class; larger S value denotes smaller group of favored class and a more focused subset of variants with higher prior. Y-axis denotes fold enrichment of the favored class $\Pr(M_F)/\Pr(M_{NF})$. For power for the favored class, it could be observed that when prior strength is small, the power is insensitive to prior fold enrichment; this is because large fraction for the favored class is essentially equivalent to a less well-defined subset of variants, thus effect of prior enrichment becomes less obvious. When the favored class is more well-defined, corresponding to larges S values and smaller fractions of favored class, the effect of incorporating prior becomes more obvious; this is reflected by the power gain at large S values, and the power increases with higher fold-enrichment. Power of the non-favored class remains at around 50% when fraction of favored class is small or fold-enrichment is small, because small fold enrichment or small fraction of favored subset is unlikely to make big impact to the non-favored class; when the favored class is given a large prior and consists of large proportion of the total variants, the non-favored class begins to loose power.

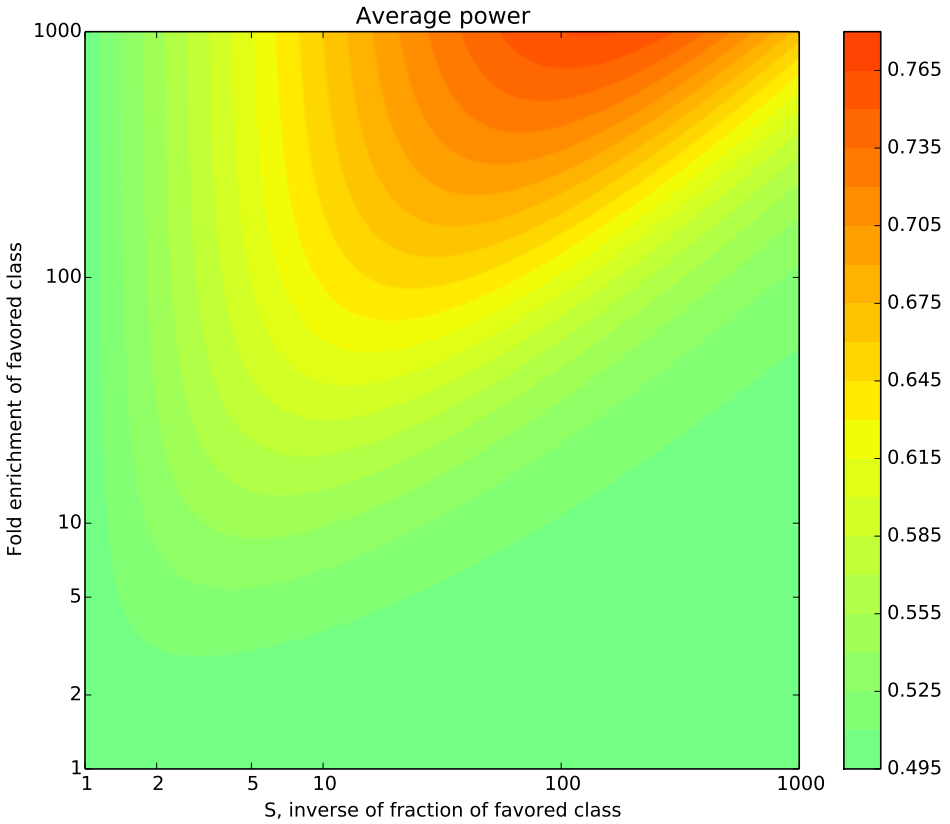

Fig 4. Average power for both classes combined at $p = 5 \times 10^{-8}$ threshold for 50% power. Axes are defined in the same way as for Fig.3. Average power is smaller than power for the favored class and larger than power for the non-favored class given the same prior strength and fold enrichment, as a balance between gain of power for the favored class and loss of power for the non-favored class. Shape of the contour is determined by weights of the two different classes: for small prior strength and large fold enrichment, shape of contour is more similar to the favored class in Fig. 3; for large prior strength and small fold enrichment, shape is similar to the non-favored class.

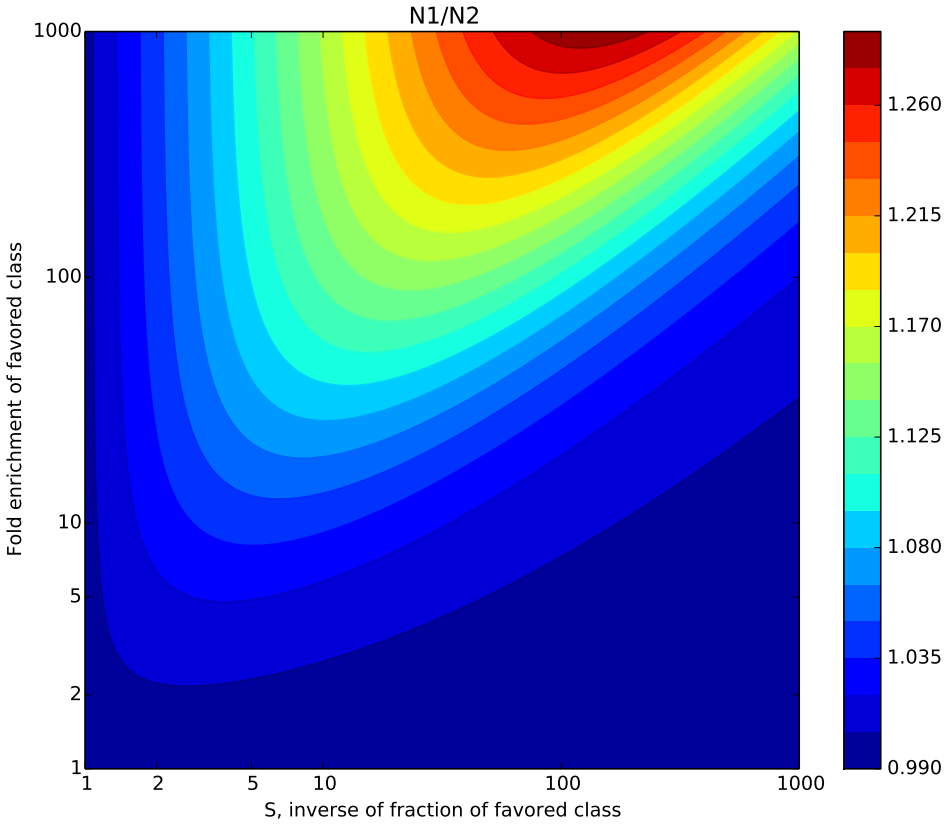

Fig 5. Population size fraction increase to achieve the same average power for at 5×10^{-8}. Axes are defined in the same way as for Fig. 3. Values are N_1/N_2 where N_1 is the population size without prior, and N_2 is population size to achieve the same power using prior. The maximum population size fraction increase is 1.3 fold to obtain the maximum power gain fulfilled through incorporating a prior, further prove the point that population size is of a more crucial factor for association testing.