Permutation methods for assessing significance in binary trait association mapping with structured samples

Joelle Mbatchou[1], Mark Abney[2], Mary Sara McPeek[1,2*]

1 Department of Statistics, The University of Chicago, Chicago, IL, USA
2 Department of Human Genetics, The University of Chicago, Chicago, IL, USA

* mcpeek@uchicago.edu

Abstract

In genetic association analysis of complex traits, permutation testing can be a valuable tool for assessing significance when the asymptotic distribution of the test statistic is not known. This commonly arises when the association test statistic is itself a function of multiple correlated statistics. Examples include omnibus tests that combine test statistics that perform well in different scenarios, as well as tests of gene-set, pathway or genome-wide significance. For genetic association testing in samples with population structure and/or related individuals, use of naive permutation can lead to inflated type 1 error. To address this in quantitative traits, the MVNpermute method was developed. However, for association mapping of a binary trait, the relationship between the mean and variance makes both naive permutation and the MVNpermute method invalid. We propose BRASS, a permutation method for binary trait association mapping in samples that have related individuals and/or population structure. BRASS allows for covariates and ascertainment and accommodates a wide range of test statistics. We use an estimating equation approach that can be viewed as a hybrid of logistic regression and linear mixed-effects model methods, and we use a combination of principal components and a genetic relatedness matrix to account for sample structure. We show in simulation studies that BRASS maintains correct control of type 1 error in a range of scenarios that include population structure, familial relatedness, ascertainment and phenotype model misspecification. We apply our method to a genome screen for elbow dysplasia (ED) in 82 breeds of domestic dog and to a genome screen for idiopathic epilepsy (IE) in the Irish Wolfhound breed of domestic dog. We find that a previously-identified region on chromosome 4 containing a number of candidate genes is significantly associated with IE.

Author summary

Permutation testing is commonly used when distributional assumptions cannot be made or do not apply, or when performing a multiple testing correction, e.g., to assess region-wide or genome-wide significance in association mapping studies. Naively permuting the data is only valid under the assumption of exchangeability, which, in the presence of sample structure and polygenicity, typcially does not hold. Linear mixed-model based approaches have been proposed for permutation-based tests with continuous traits that can also adjust for sample structure; however, these may not remain valid when applied to binary traits, as key features of binary data are not well accounted for. We propose BRASS, a permutation-based testing method for binary data

that incorporates important characteristics of binary data in the trait model, can accommodate relevant covariates and ascertainment, and adjusts for the presence of structure in the sample. We demonstrate the use of this approach in the context of correcting for multiple testing in two genome-wide association studies in domestic dog: one for elbow dysplasia and one for idiopathic epilepsy.

Introduction

To elucidate the genetic architecture of complext traits in either human populations or model organisms such as mouse, dogs or cattle, many studies have used genome-wide association (GWA) analyses. In these GWA studies, the primary objective has been to identify associations between a phenotype of interest and genetic markers, usually SNPs. This involves assessing the statistical significance of a given test statistic by deriving its null distribution or an asymptotic approximation to it. However, this is not always feasible as the distribution may be intractable; such a scenario can arise in region-based tests where association signals over multiple sites are combined (e.g. rare variant tests), or when the test statistic involves data-adaptive weights [1]. A further limitation can arise when, even if the distribution (or the asymptotic distribution) of the test statistic is known for single tests, significance needs to be assessed for the maximum of many correlated tests. This occurs in genome scans to establish a genome-wide significance threshold, where the linkage disequilibrium present between the markers induces correlation between the association tests [2].

To overcome these limitations, a common approach is to perform permutation testing so as to obtain replicates of the data under the null hypothesis from which an empirical distribution can be derived. A fundamental assumption permutation testing relies on is exchangeability of the subjects in the sample, which is usually satisfied in population-based samples. However, this assumption can be violated in the presence of genetic relatedness (e.g. population structure), as it will introduce correlation in the sample, including in the phenotype values through polygenic effects [3, 4]. Hence, naive application of permutation testing will usually not preserve the correlation structure and can result in inflated type 1 error rates [5] (though this can be avoided if, for example, all subjects in the sample are equally related [6]). We consider here the problem of permutation testing for a binary trait in the presence of polygenic effects in a sample with population structure, cryptic and/or family relatedness. While an exact permutation test may not be feasible, a permutation-based test that adjusts for the correlation structure is feasible and has been proposed for quantitative traits [4, 6]. The approach is based on a linear mixed model (LMM) and is asymptotically valid for multivariate-normal data. It incorporates genetic relatedness through the inclusion of random effects in the LMM and can also adjust for covariates. However, as it is primarily designed for quantitative traits that are multivariate-normal, there will be more model misspecification when applied to binary traits as the binary nature of the data is not be incorporated into the LMM.

We propose BRASS (for "binary trait resampling method adjusting for sample structure"), a permutation procedure for a binary trait which incorporates both covariates and the correlation structure present in the sample. In contrast to the LMM-based approach, it accommodates the binary nature of the trait through a quasi-likelihood framework that considers the effect of covariates on a logit scale in the mean structure as well as the relationship between the trait mean and its variance, both of which are important features of binary data. Hence, BRASS benefits from less model misspecification in the presence of important covariates relative to the LMM-based approach. Here, we show that the permutation-based replicates obtained with BRASS are able to maintain correct type 1 error control and compare it with several alternative

approaches. As the use of our method requires accounting for the phenotypic correlation structure, we go over modeling choices for the relatedness and how it can be incorporated in the quasi-likelihood framework used. We demonstrate the use of our method in a context of assessing genome-wide significance in domestic dog GWA studies. In animal studies, permutation testing is commonly used when there is not a consensus threshold for genome-wide significance [7–10]. We apply BRASS to perform association mapping in two GWA studies of domestic dogs, one for elbow dysplasia (ED [OMIA: 000330-9615]) and and one for idiopathic epilepsy (IE [OMIA: 000344-9615]).

Results

Overview of the resampling approaches proposed

BRASS method. We consider the problem of resampling for correlated binary data, where the correlation can arise from various sources such as population structure or related individuals. Our aim is to derive replicates of the trait under the null hypothesis of no association while accounting for the correlation that is present in the sample but whose structure is unknown. To model the response, we use a recently described quasi-likelihood framework for correlated binary traits [11, 12],

$$\boldsymbol{\mu} := \mathbb{E}(\mathbf{Y}|\mathbf{X},\mathbf{G}) = \text{logit}^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{G}\gamma), \text{ and} \tag{1}$$

$$\boldsymbol{\Omega} := \text{Var}(\mathbf{Y}|\mathbf{X},\mathbf{G}) = \boldsymbol{\Gamma}^{1/2}\,\boldsymbol{\Sigma}\,\boldsymbol{\Gamma}^{1/2}. \tag{2}$$

where $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$ denotes the phenotype vector for n subjects, $\mathbf{X}$ is the $n \times k$ matrix of k covariates (including an intercept term), $\mathbf{G} = (G_1, \ldots, G_n)^T$ is the vector of genotypes at the marker of interest, where G_i denotes the minor allele count for the i-th individual, $\boldsymbol{\beta}$ is a k-length vector representing the unknown effects of covariates, γ represents the unknown effect of the genetic marker, $\boldsymbol{\Gamma} = \text{diag}\{\mu_i(1-\mu_i)\}$, and $\boldsymbol{\Sigma} = \xi\,\boldsymbol{\Phi} + (1-\xi)\,\mathbf{I}_n$, where the unknown scalar parameter $\xi \in [0,1]$ allows for the incorporation of sample correlation as determined by the elements of the matrix $\boldsymbol{\Phi}$ which is assumed known.

The framework characterized by (1) and (2) allows to adjust for important covariates in the mean structure, such as biological or ancestry informative covariates, and the correlation between subjects is captured directly within the variance structure by the inclusion of $\boldsymbol{\Phi}$ in (2). Furthermore, by including the matrix $\boldsymbol{\Gamma}$ in (2), this framework allows for the dependence of the variance on the phenotypic mean, a key feature of binary data. Since we wish to simulate replicates of the binary trait under the null hypothesis of no association, $H_0 : \gamma = 0$, the unknown parameter $(\boldsymbol{\beta}, \xi)$ which specifies the mean and variance structure needs to be estimated from the data. Unlike with logistic mixed models (LogMM) which are a natural choice for correlated binary data but are computationally challenging to fit as they involve high dimensional integrals, obtaining parameter estimates is computationally efficient as it involves solving a system of estimating equations [11, 12].

Due to the lack of exchangeability in the data which renders an exact permutation test unfeasible, we instead perform an approximate permutation test based on second-order exchangeability. To do that, we first obtain a vector based on the assumed model fitted under the null that will have uncorrelated entries with same mean and variance, which can then be permuted. Using the residual vector from the null model which is mean-centered but has no closed-form expression, we want to approximate its second moment so as to map it to an orthonormal space in which its entries will be second-order exchangeable. We first obtain a closed-form expression for $\widehat{\boldsymbol{\beta}}$, the null estimate of $\boldsymbol{\beta}$, in terms of the data and, since the estimated mean $\widehat{\boldsymbol{\mu}}$ is a non-linear

function of $\widehat{\boldsymbol{\beta}}$, we use a linear approximation to obtain a closed-form expression for the residuals $(\mathbf{Y} - \widehat{\boldsymbol{\mu}})$ as a function of the data. As a result, we can approximate the covariance matrix of the residuals as,

$$\mathrm{Var}(\mathbf{Y} - \widehat{\boldsymbol{\mu}}) \approx \boldsymbol{\Omega}_0 - \boldsymbol{\Gamma}_0 \mathbf{X} \left(\mathbf{X}^T \boldsymbol{\Gamma}_0 \, \boldsymbol{\Omega}_0^{-1} \, \boldsymbol{\Gamma}_0 \mathbf{X}\right)^{-1} \mathbf{X}^T \boldsymbol{\Gamma}_0, \tag{3}$$

where $\boldsymbol{\Gamma}_0$ and $\boldsymbol{\Omega}_0$ are the matrices $\boldsymbol{\Gamma}$ and $\boldsymbol{\Omega}$ evaluated at the true (and unknown) value of $\boldsymbol{\beta}$ assuming that H_0 is true. The correlation present in $(\mathbf{Y} - \widehat{\boldsymbol{\mu}})$ arises from two sources: that introduced by $\boldsymbol{\Phi}$ through $\boldsymbol{\Omega}_0$ and that introduced from using the estimated mean $\widehat{\boldsymbol{\mu}}$ instead of the true unknown mean $\boldsymbol{\mu}$. We use a factorization $\mathbf{C}$ of $\boldsymbol{\Omega}_0$, with $\boldsymbol{\Omega}_0 = \mathbf{C}^T\mathbf{C}$, to remove the correlation due to $\boldsymbol{\Phi}$ and obtain,

$$\mathrm{Var}[\mathbf{C}^{-T}(\mathbf{Y} - \widehat{\boldsymbol{\mu}})] \approx \mathbf{I} - \mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T = \boldsymbol{\Psi}_0, \tag{4}$$

where $\mathbf{W} = \mathbf{C}^{-T}\boldsymbol{\Gamma}_0\mathbf{X}$. The matrix $\boldsymbol{\Psi}_0$ in (4) is symmetric and idempotent and thus can be expressed as $\boldsymbol{\Psi}_0 = \mathbf{V}\mathbf{V}^T$ where the columns of $\mathbf{V}$ contain the eigenvectors corresponding to the eigenvalue 1 with $\mathbf{V}^T\mathbf{V} = \mathbf{I}_{n-k}$. As a result, we use $\mathbf{V}^T$ to remove the remaining correlation that is driven by parameter estimation.

We thus obtain a vector with uncorrelated entries that have same mean and variance. In practice, a new trait replicate $\mathbf{Y}_\pi$ is generated as,

$$\mathbf{Y}_\pi = \widehat{\boldsymbol{\mu}} + \widehat{\mathbf{C}}^T\widehat{\mathbf{V}}\boldsymbol{\Pi}\widehat{\mathbf{V}}^T\widehat{\mathbf{C}}^{-T}(\mathbf{Y} - \widehat{\boldsymbol{\mu}}) \tag{5}$$

where we use the null estimate $(\widehat{\boldsymbol{\beta}}, \hat{\xi})$ to obtain the matrices $\widehat{\mathbf{C}}$ and $\widehat{\mathbf{V}}$ that replace $\mathbf{C}$ and $\mathbf{V}$, respectively, and $\boldsymbol{\Pi}$ is a permutation matrix. Under suitable regularity conditions, the quasi-likelihood estimates are consistent [13] so that the method we have proposed is asymptotically valid under second-order exchangeability. The quasi-likelihood model only involves the first two moments of the trait, and the corresponding mean and variance structure estimated under the null are preserved in the trait replicates. By estimating both of these structures, our method is applicable to general scenarios where there might be important covariates and the correlation structure present is unknown.

Additional resampling methods considered. As mentioned previously, the nature of the data suggests developing a resampling approach based on LogMMs. We propose to simulate trait replicates by sampling from a LogMM that is fitted under the null using penalized quasi-likelihood (PQL). This approach (referred to as LogMM-PQL) renders binary replicates that incorporate the mean and variance structure specified by the fitted logistic model. In addition, we also simulate replicates using MVNpermute [4, 6], a resampling method for correlated data with a quantitative phenotype, as well as replicates based on permuting residuals from a LMM (referred to as Naive). As BRASS, MVNpermute and Naive methods simulate quantitative replicates, we consider alternative approaches based on converting these replicates to binary (referred to as BRASS$_{mod}$, MVNpermute$_{mod}$ and Naive$_{mod}$, respectively, where the subscript indicates that the trait replicate has been converted to binary).

Further details of the methods considered are provided in the Methods section.

Simulations

Assessing the type 1 error. We begin by examining the performance of the proposed resampling methods in the context of correcting for multiple testing, with both population structure and pedigree structure present in the sample in addition to important covariates. We expect that by estimating the structure present in the sample in a way that incorporates the binary nature of the trait, our BRASS resampling method will lead to correct calibration of the type 1 error. As the PQL fitting algorithm used in LogMM-PQL is known to lead to underestimation of the variance component

parameter when it is high [14, 15], we expect this approach to result in inflated type 1 error when the sample contains a high amount of correlation. Naive and MVNpermute are expected to perform worse than BRASS and LogMM-PQL in the presence of covariates as they do not incorporate the dependence of the variance on the mean and hence, the covariates, and they model covariate effects on a linear scale. Furthermore, we expect the Naive method to perform worse with increasing sample correlation as it is unaccounted for when generating replicates.

We consider the setting in which a binary trait is measured in 46 equal-sized families in 2 sub-populations, and association is tested with each of $m = 100$ null markers (i.e. not associated with the binary trait) using the test statistic CARAT [12]. The aim is to sample trait replicates in order to determine the significance of the smallest p-value among the m tests. Correlation is introduced by the presence of population structure and family relatedness, and they are incorporated in the trait generative model through polygenic effects. We compare four resampling methods: (1) BRASS, our proposed approach based on a quasi-likelihood framework; (2) LogMM-PQL, which generates binary replicates by simulating from a fitted LogMM; (3) MVNpermute, a resampling method for correlated data with quantitative phenotypes; (4) Naive, which is based on permuting residuals from a LMM. Each of these four methods will lead to a p-value estimate corresponding to the significance of the smallest p-value among the m tests conducted with the observed trait, and the type 1 error for each method is estimated as the proportion of simulations where the estimated p-value is less than the nominal level.

All the resampling strategies involve fitting a prospective model that incorporates the correlation present in the sample. A common approach is to estimate the genetic correlation present by the use of an empirical genetic relatedness matrix (GRM) so as to model the structure using random effects [16]. Alternatively, the top principal components of a given GRM are used as covariates to model the structure present using fixed effects [17]. Here, we combine the two approaches and use both fixed and random effects to capture the sample correlation in the null model. We use the top PCs from an empirical GRM as fixed effects in the null model and then build a new GRM matrix based on the genotype information adjusting for the effects of these top PCs to capture the leftover structure as random effects (see Methods for details).

Robustness to model misspecification with all covariates included. As the true model for the trait is generally unknown in an association study, we simulate the data using both a logistic model as well as a liability threshold model (see Methods). The results using a logistic model are shown in Fig 1 and those using a liability threshold model are shown in Fig 2. Results for BRASS$_{mod}$, MVNpermute$_{mod}$ and Naive$_{mod}$ as well as additional results for all methods at the .05 level can be found in Figures S2-S5.

We obtain similar results for both models, where as expected we observe that permuting the residuals from a linear mixed model (Naive) leads in all settings to lack of control of the type 1 error. More precisely, when covariates explain most of the variation on the logit scale, the method is overly conservative and when instead polygenic effects explain most of the variability, the type 1 error rate is significantly inflated. Moreover, we also find that adjusting for sample correlation prior to permutation based on LMM residuals, as done in MVNpermute, still leads to a very conservative test when covariates highly influence the variability on the logit scale. This is most likely because a linear mixed model does not allow any dependence between the trait variance and the mean, which is influenced by the covariates. As covariates become less important, we do see better control of the type 1 error. Contrary to the former two approaches, BRASS maintains control of the type 1 error in all settings considered; we note that when covariates are more important relative to the polygenic effects, the method tends to be slightly more conservative. For LogMM-PQL, we observe an inflation in the type 1 error rate as the correlation present in the sample becomes more important. This approach

was expected to lead to such inflation in the empirical error rate as the PQL model fitting approach tends to underestimate the variance components and hence, would result in less correlation present in the simulated data compared to the one present in the original data. For Naive$_{mod}$, MVNpermute$_{mod}$ and BRASS$_{mod}$, control of the type 1 error is improved compared to the corresponding original methods when covariates strongly influence the phenotypic mean. However, they all lead to inflated error rates as the amount of correlation present in the sample increases.

We find that none of the resampling methods that simulate binary replicates consistently control the type 1 error rate, and MVNpermute and Naive lead to significantly deflated type 1 error in the presence of strong covariate effects relative to the polygenic effects. However, BRASS consistently results in well-calibrated type 1 error, regardless of the amount of population and family structure present in the sample relative to the covariate effects. This suggests that (1) incorporating the dependence between the trait mean and variance, as well the effects of covariates on a logit scale, lead to robustness of the method in the presence of important covariates; (2) converting the replicates from BRASS to binary, as done with BRASS$_{mod}$, does not well preserve the structure in the observed data that was captured in the quantitative replicates and performs equally as well as generating replicates based on a LMM with MVNpermute$_{mod}$.

Robustness to model misspecification with a missing important covariate. Next, we assess the robustness of the proposed methods when important covariates are omitted from the fitted model. As it is usually not known a priori which variables should be kept in the analysis, one would usually try including different combinations of covariates to finally determine the ones to include in the final model. It may occur that one of the covariates has a moderate effect on the trait and leads to a p-value close to the significance threshold (e.g. 0.05). Hence, a judgment call would be required for whether to keep the covariate in the model and one may decide to exclude it. It is thus of interest to see how the proposed methods would fare in such a scenario as the replicates generated would come from a more misspecified model.

To illustrate this situation, we simulated a binary trait using a logistic model with both covariates and polygenic effects present. However, among the covariates, we selected one of them to have an effect on the trait which corresponded to a p-value of about 0.05 when fitting a LMM with all the covariates included. In all resampling methods used, we excluded this covariate when fitting the null model. Since these methods all have misspecified first moment, the estimates of the covariate effects will have some amount of bias. We assessed the resulting type 1 error rate when correcting for multiple tests, where we used the same simulation design as previously mentioned to generate the genetic markers.

The simulation results are displayed in Fig 3 (results for BRASS$_{mod}$, MVNpermute$_{mod}$ and Naive$_{mod}$ and results for all methods at level .05 can be found in Figures S6 and S7). We see that BRASS retains good calibration of the type 1 error in all settings, with the empirical error rate being conservative when covariates explain most of the variability in the trait mean. In that same setting, MVNpermute and Naive lead to a highly conservative type 1 error rate, with barely any false rejections being made. The other methods which generate binary replicates lead to significantly inflated type 1 error rates as the amount of correlation present in the sample increases.

Robustness to model misspecification in the presence of ascertainment. We determine the robustness of the proposed methods when trait-based ascertainment has been applied to the sample. This is commonly used in case-control studies where individuals are included in the sample based on their disease affection status, such as if the prevalence of the trait in the population is too low to obtain sufficient power.

We simulated a binary trait under the same settings as for the logistic model with all

covariates included. To introduce ascertainment, we selected at random 500 cases to be retained in the sample, and an additional 500 controls so that the the sample had a 1:1 case-control ratio.

The results are shown in Fig 4 (results for BRASS$_{mod}$, MVNpermute$_{mod}$ and Naive$_{mod}$ and results for all methods at level .05 can be found in Figures S8 and S9). We again see that BRASS has well-controlled type 1 error rate in all settings considered, and is slightly conservative when covariates explain most of the variability in the trait mean. MVNpermute and Naive fail to control the type 1 error, as they give overly conservative rates when covariates have a major impact on the trait. The remaining methods all exhibit too liberal error rates as polygenic effects become more important in the trait model.

These results demonstrate the robustness of BRASS under various sources of model misspecification, including the use of a quasi-likelihood model that does not match to any true likelihood, the exclusion of covariates with moderate effects from the fitted null model and the presence of trait-based ascertainment.

Application to Domestic Dog Data

We now illustrate the use of our resampling method in the context of multiple testing correction in a GWAS of domestic dogs [18]. Unlike with humans, the genome-wide threshold for significance has not been well determined. This is one of the largest dog genotyping studies with 4,224 dogs genotyped at 185,805 SNPs and twelve clinical and morphological phenotypes recorded. We analyze two binary traits from this study (1) elbow dysplasia (ED), which is an ensemble of abnormalities that affect the articular surfaces of the elbow; (2) idiopathic epilepsy (IE), which is diagnosed when there is no identifiable cause for seizures in affected dogs.

The data for ED contained dogs of various breeds while for IE we had dogs from a single breed. Similarly to our simulations, we incorporated the structure present in the ED data as both fixed and random effects in our model. We used either the top 9 or 13 PCs to estimate a GRM based on genotype residuals removing the effects of these top PCs. Hence for the analysis of ED, we included the top PCs as covariates in the null model and used the estimated GRM in the variance structure. For IE, we used an empirical GRM based on the genotype information to capture the genetic relatedness. We used CARAT to perform single-SNP association tests. For genome-wide significance assessment, 100,000 trait replicates were generated under the null hypothesis of no association using both BRASS and MVNpermute$_{mod}$, the latter of which gave similar/better type 1 error calibration compared to the remaining resampling approaches, and genome-wide p-values were estimated using the empirical distribution of the test statistics based on the trait replicates. The parameter estimates from the fitted models of both BRASS and MVNpermute$_{mod}$ (excluding effects for the PCs) are reported in Table 1 for both traits analyzed.

The genomic control inflation factors for ED with 9 and 13 PCs as well as for IE are $\lambda_{\mathrm{GC}} = 0.968, 0.98$ and 0.979, respectively. The genome-wide significance threshold is estimated at nominal level 0.05 from the empirical distribution of the top association signal using the trait replicates. Manhattan plots of the p-values of the single-SNP tests for the observed data are presented in Fig 5 for both ED and IE phenotypes, along with the estimated genome-wide significance thresholds. In the analysis of ED using the top 9 PCs, the strongest association signal is found on chromosome 26 for SNP rs9000666, and it fails to reach the genome-wide significance threshold of 2.1×10^{-7} estimated using BRASS and 1.8×10^{-7} using MVNpermute$_{mod}$ (Table 2). Similar results are obtained using the top 13 PCs (not included). The genome-wide thresholds of BRASS and MVNpermute$_{mod}$ are fairly similar; these two approaches are expected to differ the least when covariates contribute a minor portion of the variability relative to polygenic

effects. Hence, as patterns of relatedness within breeds are much stronger than those across breeds (Fig S11), which are captured in the PC covariates, this suggests that more of the phenotypic variability is attributable to within-breed genetic similarities compared to covariate effects. For IE, we find a 13Mb region on chromosome 4 (position 7.53-20.8 Mb) that reaches the genome-wide significance estimated with BRASS, which is 1.7×10^{-6} (Table 3). The candidate genes in this region are: *DISC1* [NCBI Gene: 488970], which has been associated with various neuronal abnormalities in humans [19, 20]; *MYPN* [NCBI Gene: 489013], which has been associated with dilated cardiomyopathy [MIM: 615248] in humans [21, 22]; *ARV1* [NCBI Gene: 488975], which has been associated with epileptic encephalopathy [MIM: 617020] in humans [23]; and *ACTA1* [NCBI Gene: 488984], which has been associated with scapulohumeroperoneal myopathy [MIM: 616852], a neurological disease, in humans [24] . The significance threshold established with MVNpermute$_{mod}$ is much more stringent than that of BRASS, at 2.0×10^{-7}, and a shorter 1 Mb region on chromosome 4 remains above this threshold. Contrary to the analysis with ED, here we obtain quite different results between BRASS and MVNpermute$_{mod}$. This could be due to covariates having a strong influence on the trait variability relative to the genetic relatedness patterns present within the Irish wolfhound breed.

Computation time

In order to generate replicates from BRASS, the null prospective model of CARAT needs to be fitted, which includes specifying covariates and a GRM $\mathbf{\Phi}$. If not provided, obtaining the eigendecomposition of $\mathbf{\Phi}$ is the main computationally intensive step involved in fitting the null model[12]. However, this step (or one of equivalent computational complexity) is also performed when using methods based on a linear or logistic mixed model. Once parameter null estimates are obtained, the transformed residuals $\hat{\boldsymbol{\zeta}}$ as well as the linear map $\widehat{\mathbf{C}}^T\widehat{\mathbf{V}}$ in (5) need to be computed; however, once obtained, they can be re-used when generating trait replicates. As the eigendecomposition of $\mathbf{\Phi}$, and hence $\widehat{\mathbf{\Sigma}}$, is obtained prior to fitting the null model, we use it to compute $\widehat{\mathbf{C}}$. Therefore, the most computationally intensive step is to obtain the eigenvectors of $\widehat{\mathbf{\Psi}}$ (the estimate for $\mathbf{\Psi}_0$ in (4)), which involves a time complexity of $O(n^3)$. By using the singular value decomposition of the matrix $\mathbf{W}$ in (4), which involves a time complexity of $O(nk^2)$ (assuming the number of covariates $k < n$), we are able to obtain the eigenvectors of $\widehat{\mathbf{\Psi}}$ without having to construct it and compute its eigendecomposition. Past this step, the additional cost for generating L replicates is $O(n^2L)$; this computation can easily be parallelized to generate sets of traits replicates independently.

BRASS is implemented in a freely downloadable software package at `http://galton.uchicago.edu/~mcpeek/software/index.html`. We report run times for BRASS in real and simulated data. Using a single processor on a machine with 6 core Intel Xeon 3.50 GHz CPUs and 32 GB RAM, it takes 1 s to fit the null model and generate 1,000 trait replicates for the dog data with ED and 538 dogs and 0.16 s for IE in 202 dogs. On simulated data with 1,000, 5,000 and 10,000 individuals, it takes 1.6 s , 1 min and 4.6 min (275 s), respectively, to generate 1,000 trait replicates. As the use of BRASS will vary based on the context (e.g. multiple testing correction, region-based association testing), the computational burden involved in comparing the statistic between the observed trait and its replicates will vary based on what statistic is being considered. We note that since the trait replicates are generated under the null hypothesis of no association, they can be re-used when performing separate analyses on a large number of markers.

Discussion

A key component of case-control studies lies in assessing significance of association results performed between markers and the binary trait. In some instances, permutation tests are required in which, assuming exchangeability, the trait labels are shuffled to generate new replicates under the null hypothesis of no association. In the presence of population structure, cryptic and/or family relatedness, which are all common sources of confounding in genetic association studies, exact permutation tests are usually not feasible as they disregard the patterns of genetic relatedness among the subjects, and thus fail to retain the correlation structure present in the data. We have introduced BRASS, a novel permutation-based resampling procedure for generating trait replicates in samples with population structure, cryptic and/or family relatedness. BRASS allows for covariates and ascertainment, and it accommodates a wide range of test statistics. Using a quasi-likelihood model that incorporates the correlation present in the data, BRASS relies on obtaining an invertible transformation of phenotypic residuals to achieve approximate (asymptotic) second-order exchangeability. After permutation, new trait replicates are obtained by inverting the transformation, and the assumed correlation in the phenotype values as well as the genotype structure present in the original data are maintained. Compared to MVNpermute, the model assumed by BRASS incorporates the effects of covariates on a logit scale and also accounts for the dependence of the trait variance on the mean, both of which reflect key aspects of binary data. Consequently, BRASS benefits from less model misspecification when the covariates effects are substantial.

We demonstrate the validity of our approach in simulation studies. More specifically, we find that BRASS maintains control of the type 1 error under varying amounts of population structure, familial relatedness, ascertainment and sources of trait model misspecification. We compare BRASS to two resampling approaches based on a linear mixed model, where in the first the raw residuals are permuted and in the second, MVNpermute, permutation is applied after removing the correlation in the residuals. We find that both of the LMM-based approaches fail to maintain the type 1 error when covariate effects are substantial, suggesting that only accounting for the genetic relatedness patterns present without incorporating the key features associated with the binary nature of the trait (i.e. the dependence of the phenotypic variance on the mean) is not sufficient to obtain replicates that correctly estimate the null distribution of the statistic. We also find that when polygenic effects are important, accounting for the genetic relatedness in the sample prior to permutation as done in BRASS and MVNpermute leads to better control of the type 1 error compared to ignoring it, as done when permuting the raw residuals from a LMM. As the replicates simulated from all three approaches are quantitative, yet the original trait is binary, we also considered converting the simulated quantitative replicates to binary. We find that this additional step results in an improvement of the type 1 error rate calibration for the LMM-based methods when the polygenic component has a low impact on the trait distribution relative to covariates, but does not control the type 1 error when more structure is present in the sample.

Finally, we evaluated the approach based on sampling trait replicates from a fitted LogMM using PQL. We find that this approach leads to worse control of the type 1 error rate the more the polygenic component of the trait is important, a result likely explained by the fitting algorithm used in LogMM-PQL. Indeed, while a logistic mixed model is a natural choice for binary data that contains correlation, fitting such a model is computationally intensive and often requires some approximation of the high-dimensional integral involved. LogMM-PQL uses PQL which is known to give biased estimates in highly correlated data; we could have considered fitting algorithms with higher accuracy (e.g. using Laplace approximation[25] or Gauss-Hermite

quadrature[26]) but these also lead to increased computational burden, which may not scale well for large GWA studies.

We applied BRASS in the context of multiple testing correction in association mapping studies of ED and IE in domestic dogs. In the analysis of IE which involved dogs from a single breed, we found a 13 Mb region on chromosome 4 that reached the genome-wide significance threshold estimated using the replicates from BRASS, and has been previously associated with IE [18]. On the other hand, in the ED data set which contained multiple breeds, we did not find any region that reached the estimated genome-wide significance threshold. In the latter analysis, we chose to include PCs as covariates and these mostly captured genetic differences across breeds. Alternatively, as we had breed labeling information provided in the data set, we could have included that as an ancestry informative covariate instead of the PCs. However, in the presence of mislabeling in the breed information, such an approach would result in inaccurate inference on the breed structure present. In addition, as we had dogs of mixed breed present in the data set (akin to admixture), it would not be possible to differentiate between these dogs (i.e. they could come from different breed combinations but would be treated as originating from the same breed combination). Notably, the genome-wide p-value threshold estimated using replicates from BRASS was an order of magnitude different for the ED trait in a sample of 82 breeds than for the IE trait in a single breed. This is expected given the extremely different population structure between the two samples.

BRASS assumes that an estimate of the trait covariance matrix is available in order to capture polygenic effects, which can arise through various sources including population structure, cryptic and family relatedness. Here, we chose to use both fixed effects and random effects to capture the structure that is present in the data. This modeling choice was taken as a means to combine two common approaches to correct for population structure, which are either the use of PCs as covariates or the use of random effects in the trait model. Our approach does not attempt to distinguish whether distant (recent) sources of genetic relatedness are captured in the fixed (random) components of the model; we simply partition the relatedness into random and fixed effects and include both in our model. To the extent that population membership is discrete, using PCs that reflect that structure as covariates would provide more confidence that population structure has been well corrected for in the analysis. A recently proposed method, PC-Air [27], obtains PCs from a derived subset of mutually unrelated subjects that are representative of the ancestral diversity present in the sample, so as to ensure that the top PCs will only capture distant genetic relatedness (i.e. population structure). However, we did not investigate the use of such an approach in our analyses.

An alternative resampling strategy not considered here is to obtain replicates for the genotypes by using gene dropping. In this approach, which is applicable to samples that contain families, the genotypes for the pedigree founders are dropped down the pedigree assuming Mendelian segregation. Hence, the correlation present between the trait values is preserved. The resulting simulated replicate can be used to get an empirical distribution of the test statistic considered under the null hypothesis of no association. A major caveat of gene dropping is that along with the presence of families in the sample, both the complete pedigree structure and the founders' genotypes have to be known in order to simulate transmissions down the pedigree; this is infeasible when cryptic relatedness is present in the sample or when the genealogy in earlier generations is unknown. Another resampling strategy we did not explore is based on permuting the genotypes; since an exact permutation of the genotypes would not preserve the sample structure, we could propose a model based on the first two moments of the genotypes and obtain a linear map that projects them onto an orthonormal space, similarly to the approach in BRASS. This would could be useful when looking at a smaller region in the

genome as it involves the joint distribution of multiple markers; however, it would also ignore any higher order dependencies in this distribution beyond the first two moments. By leaving the genotypes undisturbed, BRASS ensures that the dependence structure of the genotypes is preserved in the simulated replicates.

We chose here to consider the application of BRASS in the context of performing a scan over a genetic region with multiple tests being performed and significance being assessed after correcting for the simultaneous tests. More generally, BRASS can be applied to situations where either the null distribution or the asymptotic null distribution of the statistic is intractable. Example of such situations arise in the analysis of rare variants, where the statistic is obtained from combining signals over multiple sites or when the statistic includes weights that depend on the phenotype. Depending on the application, one can consider a different approach to capture the structure present in the sample, as the model used by BRASS can accommodate either covariates or random effects, which are represented by the inclusion of an additional matrix besides the identity in the trait covariance structure, or both. We have shown here one example of how the structure can be incorporated in the model.

Acknowledgments

This work was supported by NIH grant R01 HG001645 (to M.S.M.). This study makes use of data generated from a previous study.[18] The data are available from the Dryad online repository at https://datadryad.org/resource/doi:10.5061/dryad.266k4.

Methods

Software implementing BRASS will be available at `http://galton.uchicago.edu/~mcpeek/software/index.html`.

Derivation of trait replicates

We derive a permutation-based method to generate replicates for correlated data with a binary response based on the quasi-likelihood framework in (1) and (2) that incorporates the correlation present in addition to potential covariates. A permutation test based on the phenotype involves generating replicates under the null hypothesis of no association ($H_0 : \gamma = 0$). Under this null, the form of the quasi-likelihood framework proposed in (1) and (2) contains unknown parameter $(\boldsymbol{\beta}, \xi)$, which needs to be estimated from the data. The null estimate $(\widehat{\boldsymbol{\beta}}, \hat{\xi})$ is obtained by iteratively solving the following system of estimating equations [12] under the constraint $\gamma = 0$,

$$\boldsymbol{D}_{\boldsymbol{\beta}}^T \boldsymbol{\Omega}^{-1}(\mathbf{Y} - \boldsymbol{\mu}) = \mathbf{0}, \tag{6}$$

$$(\mathbf{Y} - \boldsymbol{\mu})^T \boldsymbol{\Gamma}^{-1/2} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Phi} - \mathbf{I}) \boldsymbol{\Sigma}^{-1} \boldsymbol{\Gamma}^{-1/2}(\mathbf{Y} - \boldsymbol{\mu}) = \mathrm{trace}(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Phi} - \mathbf{I})). \tag{7}$$

where $\boldsymbol{D}_{\boldsymbol{\beta}} = \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}}$, is a Jacobian of the conditional mean of the trait with respect to $\boldsymbol{\beta}$. In our model with a logit link function, $\boldsymbol{D}_{\boldsymbol{\beta}} = \boldsymbol{\Gamma}\mathbf{X}$. Under the constraint $\gamma = 0$, let $U(\boldsymbol{\beta}) = \boldsymbol{D}_{\boldsymbol{\beta}}^T \boldsymbol{\Omega}^{-1}(\mathbf{Y} - \boldsymbol{\mu})$, which corresponds to the quasi-score function for $\boldsymbol{\beta}$ with fixed ξ. As this quasi-score function involves both the unknown vector $\boldsymbol{\beta}$ as well as the unknown scalar ξ through $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$, respectively, we first consider the case where ξ is known. Assuming that H_0 is true, let $\boldsymbol{\beta}_0$ denote the true value of $\boldsymbol{\beta}$, and let $\boldsymbol{\mu}_0, \boldsymbol{\Gamma}_0$, $\mathbf{D}_0$ and $\boldsymbol{\Omega}_0$ correspond to $\boldsymbol{\mu}, \boldsymbol{\Gamma}, \mathbf{D}_\beta$ and $\boldsymbol{\Omega}$, respectively, evaluated at $\boldsymbol{\beta}_0$ and $\gamma = 0$. Similarly, let $\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Gamma}}$, $\widehat{\mathbf{D}}$ and $\widehat{\boldsymbol{\Omega}}$ be the same quantities evaluated at $\widehat{\boldsymbol{\beta}}$ and $\gamma = 0$.

We take an approach that is an extension of previous work [4] on deriving permutation-based replicates for a quantitative response in the presence of correlation. More precisely, we aim to obtain a vector with entries that are uncorrelated and have the same mean and variance under the null hypothesis so that permutation will be valid under the second-order exchangeability. In order to remove the mean effect, we first consider the residual vector $(\mathbf{Y} - \widehat{\boldsymbol{\mu}})$, which asymptotically has zero mean under H_0. We then aim to derive a closed-form expression of its second moment, so that we can determine a linear transformation that would result in a vector with uncorrelated entries. The mean $\widehat{\boldsymbol{\mu}}$ depends on the fixed effects estimate $\widehat{\boldsymbol{\beta}}$, which we can express in terms of the data by applying a Taylor series expansion to $U(\boldsymbol{\beta})$ around $\boldsymbol{\beta}_0$, and evaluated at $\widehat{\boldsymbol{\beta}}$,

$$U(\widehat{\boldsymbol{\beta}}) \approx U(\boldsymbol{\beta}_0) + \left.\frac{\partial U(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \cdot (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0), \tag{8}$$

Using the fact that $U(\widehat{\boldsymbol{\beta}}) = \mathbf{0}$, and replacing the Jacobian in (8) by its expectation (similar to Fisher scoring), we get,

$$\begin{aligned} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 &\approx - \left\{ \mathbb{E}\left[\frac{\partial U(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right]\Bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \right\}^{-1} U_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0), \\ &= (\mathbf{D}_0^T\, \boldsymbol{\Omega}_0^{-1}\, \mathbf{D}_0)^{-1} \mathbf{D}_0^T\, \boldsymbol{\Omega}_0^{-1} (\mathbf{Y} - \boldsymbol{\mu}_0), \end{aligned} \tag{9}$$

As $\widehat{\boldsymbol{\mu}}$ is a non-linear function of $\widehat{\boldsymbol{\beta}}$, a Taylor series expansion is performed on $\boldsymbol{\mu}$ around $\boldsymbol{\beta}_0$, and evaluated at $\widehat{\boldsymbol{\beta}}$,

$$\widehat{\boldsymbol{\mu}} \approx \boldsymbol{\mu}_0 + \mathbf{D}_0\, (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0), \tag{10}$$

Combining (9) and (10), we get for the residual vector $(\mathbf{Y} - \widehat{\boldsymbol{\mu}})$,

$$\begin{aligned} \mathbf{Y} - \widehat{\boldsymbol{\mu}} &= \mathbf{Y} - \boldsymbol{\mu}_0 - (\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0), \\ &\approx \left[\mathbf{I} - \mathbf{D}_0\, (\mathbf{D}_0^T\, \boldsymbol{\Omega}_0^{-1}\, \mathbf{D}_0)^{-1} \mathbf{D}_0^T\, \boldsymbol{\Omega}_0^{-1}\right]\, (\mathbf{Y} - \boldsymbol{\mu}_0), \end{aligned} \tag{11}$$

Hence, we obtain the approximation in (3) for the covariance matrix of the residuals, and after removing the correlation due to the structure in the data by using the factorization of $\boldsymbol{\Omega}_0$, we obtain a vector whose covariance matrix is specified in (4). The matrix $\boldsymbol{\Psi}_0$ is symmetric and idempotent, and assuming that the covariate matrix $\mathbf{X}$ is of full rank k, we have that $\text{rank}(\boldsymbol{\Psi}_0) = n - k$. By the spectral theorem, $\boldsymbol{\Psi}_0 = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$ where $\mathbf{U} = (\mathbf{V}\ \mathbf{Z})$ is an orthogonal matrix with the columns of $\mathbf{V}$ being the $(n-k)$ eigenvectors corresponding to the eigenvalue 1 and the columns of $\mathbf{Z}$ being the remaining k eigenvectors corresponding to the eigenvalue 0. We have $\boldsymbol{\Psi} = \mathbf{V}\mathbf{V}^T$ and $\mathbf{V}^T\mathbf{V} = \mathbf{I}_{n-k}$.

The linear transformation $\hat{\boldsymbol{\zeta}} = \mathbf{V}^T\mathbf{C}^{-T}(\mathbf{Y} - \widehat{\boldsymbol{\mu}})$ has for covariance matrix,

$$\text{Var}(\hat{\boldsymbol{\zeta}}) \approx \mathbf{V}^T\boldsymbol{\Psi}\mathbf{V} = \mathbf{I}_{n-k}, \tag{12}$$

Hence, we obtain a transformation of the residuals with approximately uncorrelated entries, and on which permutation is asymptotically valid assuming second-order exchangeability. With $\boldsymbol{\Pi}$ denoting a random permutation matrix, we generate a new trait replicate $\mathbf{Y}_\pi$ as,

$$\mathbf{Y}_\pi = \widehat{\boldsymbol{\mu}} + \mathbf{C}^T\mathbf{V}\boldsymbol{\Pi}\mathbf{V}^T\mathbf{C}^{-T}(\mathbf{Y} - \widehat{\boldsymbol{\mu}}) \tag{13}$$

In (13), $\mathbf{D}_0$ (the matrix used to get $\mathbf{V}$), and $\mathbf{C}$, both depend on the unknown vector $\boldsymbol{\beta}_0$ and so we replace these by $\widehat{\mathbf{D}}$ (to get $\widehat{\mathbf{V}}$) and $\widehat{\mathbf{C}}$ respectively, where

$\widehat{\boldsymbol{\Omega}} = \widehat{\boldsymbol{\Gamma}}^{1/2}\,\boldsymbol{\Sigma}\,\widehat{\boldsymbol{\Gamma}}^{1/2} = \widehat{\mathbf{C}}^T\widehat{\mathbf{C}}$. We note that $\widehat{\mathbf{C}}$ can be obtained from a factorization of $\boldsymbol{\Sigma}$ (e.g. using cholesky or eigendecomposition), as $\widehat{\boldsymbol{\Gamma}}$ is just a diagonal matrix. This derivation relies on the assumption that the parameter ξ in $\boldsymbol{\Sigma}$ is known, yet in practice, ξ is also estimated from the data under H_0. In this case, we compute $\widehat{\mathbf{C}}$ based on the factorization of $\widehat{\boldsymbol{\Sigma}} = \hat{\xi}\,\boldsymbol{\Phi} + (1-\hat{\xi})\,\mathbf{I}$, where $\hat{\xi}$ represents the null estimate of ξ. Hence, a trait replicate is obtained as,

$$\mathbf{Y}_\pi = \widehat{\boldsymbol{\mu}} + \widehat{\mathbf{C}}^T\widehat{\mathbf{V}}\boldsymbol{\Pi}\widehat{\mathbf{V}}^T\widehat{\mathbf{C}}^{-T}(\mathbf{Y} - \widehat{\boldsymbol{\mu}}) \tag{14}$$

Under the identity permutation, meaning that $\boldsymbol{\Pi} = \mathbf{I}$, we are able to recover the original response vector. Examining the form of (14), there are three main steps used to obtain a transformation of the residuals with second-order exchangeable entries. The first is to center the response by using the estimated phenotypic mean. The second is to remove the correlation present due to polygenic effects, which is done by the pre-multiplication by $\widehat{\mathbf{C}}^{-T}$. The last step is to remove the correlation that is generated from using parameter estimates instead of the true values when centering, and is represented by the pre-multiplication by $\widehat{\mathbf{V}}_1^T$. We expect this last step to have only a minor impact on the method for large sample sizes and low number of parameters. The replicates generated from this approach are quantitative, and so the binary nature of the response is not preserved during resampling. However, the first two moments assumed in the quasi-likelihood model and estimated from the data are approximately preserved. Thus, we obtain an asymptotically valid permutation-based test (under second-order exchangeability) using this approach.

Other resampling methods proposed

In addition to BRASS, we propose three alternate approaches to simulate trait replicates: (1) Fitting a logistic mixed model using penalized-quasi-likelihood (PQL), where the model form under the null is $\text{logit}(\boldsymbol{\mu}) = \mathbf{Z}\boldsymbol{\alpha} + \mathbf{u}$, with $\mathbf{Z}$ being the covariate matrix and $\mathbf{u} \sim \text{MVN}(0, \sigma^2\boldsymbol{\Phi})$, and then sampling trait replicates from the model using the estimates for $(\boldsymbol{\alpha}, \sigma^2)$; (2) Fitting a linear mixed model (LMM) whose form under the null is $\mathbf{Y} = \mathbf{Z}\boldsymbol{\alpha} + \mathbf{e}$, with $\mathbf{Z}$ being the covariate matrix and $\mathbf{e} \sim \text{MVN}(0, \sigma_1^2\boldsymbol{\Phi} + \sigma_2^2\mathbf{I})$, permuting the residuals $(\mathbf{Y} - \mathbf{Z}\widehat{\boldsymbol{\alpha}})$, adding back $\mathbf{Z}\widehat{\boldsymbol{\alpha}}$ and converting the resulting vector to binary using the procedure described below ; (3) Obtaining replicates from MVNpermute [4, 6], which is uses a similar approach to BRASS but for quantitative phenotypes, and converting them to binary using the procedure below. These approaches are referred to as LogMM-PQL, Naive$_{mod}$ and MVNpermute$_{mod}$, respectively, where the subscript indicates that the trait replicate has been converted to binary. The methods are compared in Table S1.

To convert a quantitative replicate $\mathbf{Y}_\pi$ to binary, a threshold is set and the values of $\mathbf{Y}_\pi$ above and below that threshold are converted to 1 and 0, respectively, where the threshold is chosen so that the mean of the resulting binary replicate matches with that of $\mathbf{Y}$. Thus, this conversion enforces that the original trait $\mathbf{Y}$ and the binary replicate $\mathbf{Y}_\pi$ will have the same sample moments (equal to the proportion of cases). We use GMMAT [28] to fit the null model for LogMM-PQL and use GEMMA [29] to fit the null for both Naive$_{mod}$ and MVNpermute$_{mod}$.

We also consider the original Naive and MVNpermute methods which do not include the last step of converting the quantitative replicate to binary. MVNpermute is asymptotically exact for multivariate normal data, as the entries of the transformed residuals being uncorrelated implies independence under the normal assumption, and hence exchangeability. In contrast, our method BRASS is approximate as it relies on second-order exchangeability rather than full exchangeability; the framework we use only considers the first two moments of the response. However, the covariate effects

being modeled on a logit scale, rather than a linear scale as in MVNpermute, is more appropriate for binary data as the mean is constrained within $(0, 1)$. More importantly, the covariance matrix assumed in our method allows for dependence on the mean function, which is an inherent feature of binary data; this is absent in MVNpermute. Thus, we expect these two methods to differ the most when a major portion of the trait variability is attributable to covariate rather than polygenic effects.

Adaptive resampling procedure. Given the computationally intensive nature of permutation-based approaches, we use an adaptive procedure when simulating replicates for multiple testing correction at level α, where our chosen stopping criteria are checked initially with 1000 simulated replicates and then in increments of 5000 starting at 5000 replicates. More specifically, we continue generating replicates until either (1) N_{max}, the maximum number of replicates allowed, has been reached (where $N_{max} \geq 1000$); (2) A hypothesis test for $H_0 : p = \alpha$ against $H_a : p < \alpha$ is rejected at significance level 0.01, where the estimate for the p-value p is the proportion of replicates with test statistic at least as extreme as the one observed. An exact test for H_0 is used when the number of replicates is 1000; otherwise, a z-test is performed. If N_{max} has been reached, the p-value estimated from the N_{max} replicates is compared to α.

Incorporating the sample structure in the null model

In order to get trait replicates, all seven resampling methods involve fitting a prospective model under the null that requires an estimate for the GRM $\mathbf{\Phi}$ to capture the genetic relatedness in the sample. Our approach is to model genetic relatedness as both fixed and random effects in the null model, where the fixed component is represented by the inclusion of covariates that represent major axes of genetic variation, and the random component is represented by the inclusion of a GRM that reflects the leftover structure. We use a previously proposed method, PC-Relate [30], where we first estimate the GRM $\mathbf{\Phi}$ using population estimates for the minor allele frequencies (MAF),

$$\widehat{\mathbf{\Phi}} = \frac{1}{L} \sum_{l=1}^{L} \frac{(\mathbf{G}_l - 2\hat{p}_l \mathbf{1}_n)(\mathbf{G}_l - 2\hat{p}_l \mathbf{1}_n)^T}{2\hat{p}_l(1 - \hat{p}_l)}, \tag{15}$$

where L is the number of markers, $\mathbf{G}_l$ is the column vector of minor allele counts for marker l, and $\hat{p}_l = 0.5 \cdot \overline{G}_l$, where $\overline{G}_l = n^{-1}\mathbf{1}_n^T \mathbf{G}_l$ with $\mathbf{1}_n$ being a n-length vector of ones, is the sample average estimator of the population MAF for marker l. We obtain the top D PCs from $\widehat{\mathbf{\Phi}}$, and these PCs are then included as ancestry informative covariates in the null model. We then build a GRM estimate $\tilde{\mathbf{\Phi}}$ to capture the remaining genetic similarities among subjects that are not reflected by the top PCs. The entries of $\tilde{\mathbf{\Phi}}$ are,

$$\tilde{\mathbf{\Phi}}_{ij} = \frac{\sum_{l=1}^{L} (G_{il} - 2\widetilde{p}_{il})(G_{jl} - 2\widetilde{p}_{jl})}{2 \sum_{l=1}^{L} \left[\widetilde{p}_{il}(1 - \widetilde{p}_{il})\widetilde{p}_{jl}(1 - \widetilde{p}_{jl})\right]^{1/2}}, \tag{16}$$

where G_{il} and $\widetilde{p}_{il}$ are the minor allele count and a predicted subject-specific MAF, respectively, for the i-th individual at marker l. For each marker l, the subject-specific MAF estimate is obtained as half of the fitted values from a linear regression of $\mathbf{G}_l$ on the top D PCs. So in the numerator of (16), the centered genotype values represent the residuals from this linear regression, effectively removing the effects captured through the top D PCs. The resulting $\tilde{\mathbf{\Phi}}$ is used as the GRM estimate in all seven resampling methods.

Simulation studies

We perform simulation studies to evaluate the effectiveness of permutation-based trait replicates in the context of correcting for multiple testing, with both population structure and pedigree structure present in the sample. We simulate non-causal markers that are tested for association with a binary trait, and estimate the type 1 error rate for the top signal amongst them using the empirical distribution from trait replicates generated under the null hypothesis of no association. Genotype, trait and covariates are generated under multiple simulation settings. As in real applications it is usually not known a priori which variables should be included as covariates in the null model for the phenotype, we also consider the case of misspecification in the fitted null model due to the exclusion of a relevant covariate. In addition, we investigate the impact of ascertainment in the sample as case-control studies often involve phenotype-based ascertainment which introduces misspecification in the model if not properly accounted for. We simulate $M_c = 10,000$ causal SNPs which are used to generate polygenic effects in the trait generative model as well as to correct for correlation in the sample when simulating trait replicates from the seven methods.

We consider a setting in which a binary trait is tested for association with each of m markers and the aim is to assess the significance of the smallest p-value out of the m association tests. This context would arise in testing significance of a genomic region or in assessing genome-wide significance. Owing to the computational intensive nature of the simulations, the trait vector is simulated 200 times, and 150 independent marker panels, each consisting of $m = 100$ non-causal markers, are simulated for each trait replicate. Retaining only the top association signal in each panel, this results in 150 association test statistics which overall amounts to 20,000 replicates used for type 1 error estimation. To assess significance in each data simulation, up to 10,000 trait replicates are generated under the null using each of the seven methods. For each marker panel, the trait replicates are individually tested for association with the non-causal markers, and the smallest p-value for each replicate is compared with the one from the original data simulation. We use CARAT [12] for single-SNP association testing; it is a retrospective association test for binary traits in structured samples.

Simulation studies: Genotypes

The Balding-Nichols model [31] is used to draw allele frequencies for 2 sub-populations with the fixation index F set to 0.01 which is representative of the population structure seen in humans within continents [32]. For each marker s, given the ancestral allele frequency p_s drawn from a uniform distribution on $[0.2, 0.8]$ (and independently across markers), the allele frequency in sub-population $k = 1, 2$ is drawn independently from a Beta distribution with parameters $p_s(1-F)/F$ and $(1-p_s)(1-F)/F$. An equal number of pedigrees (each with structure given in Fig S1) is assigned to the 2 sub-populations with all founders within a given pedigree being assigned to the same sub-population. For each marker, genotypes for a pedigree's founders are simulated as independent Bernoulli draws using the corresponding sub-population allele frequency, and gene dropping is used to determine genotypes for other pedigree members. The M_c causal markers are simulated only once, and they are used to compute the GRM estimate $\tilde{\mathbf{\Phi}}$ in (16), which is then reused in all simulations. In addition, independent non-causal markers are simulated for each simulation replicate using the same Balding-Nichols model, where the ancestral allele frequencies are drawn from a uniform distribution on $[0.2, 0.8]$, and these are the markers tested for association.

Simulation studies: Trait and Covariates

We simulate 46 pedigrees of the same configuration (Fig S1), each consisting of 22 individuals. For each individual, three covariates are simulated: sex, age and an i.i.d. standard normal covariate. Sex is determined by the pedigree configuration and age is drawn uniformly and independently within 1.5 years of $73, 75, 46, 43, 40, 46, 40, 43, 47, 51, 18, 21, 15, 15, 12, 9, 13, 17, 24, 21, 18$, and 14 years respectively for individuals labeled $1-22$ in the pedigree. All the covariates (besides sex) are re-generated for each simulation replicate.

We consider two generative models for the trait. First, a logistic model is used in which Y_i is given by,

$$Y_i|\mathbf{X}_i, \mathbf{W}_i, \boldsymbol{\alpha} \sim \text{Bernoulli}(p_i), \text{ independently, with } \text{logit}(p_i) = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{W}_i\boldsymbol{\alpha}, \tag{17}$$

where $\mathbf{X}_i$ is the covariate row vector for the i-th individual; $\boldsymbol{\beta}$ are the fixed effects for the covariates; $\mathbf{W}_i$ is a row vector representing the SNP information for the i-th individual corresponding to the M_c causal markers standardized to have mean 0 and variance 1; $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_{M_c})^T \sim MVN(0, \sigma^2\mathbf{I})$ represent the random effects of the M_c causal markers with polygenic variance $\sigma_a^2 = \sigma^2 M_c$. The values for $\boldsymbol{\beta}$ and σ_a^2 are chosen such that: (1a) If all covariates are included in the fitted model, they each explain an equal amount of the variation on the logit scale due to covariates; (1b) If the standard normal covariate is excluded, its effect size is set to correspond to an average p-value of 0.05 in a LMM Wald test (determine by simulations), and age and sex equally explain the remaining portion of the variability due to covariates; (2) Considering the total variance on the logit scale due to covariate effects and polygenic random effects, the fraction of this variance due to covariates is fixed at 20%, 40%, 60% or 80%; (3) Bernoulli error explains on average about 20% of the total phenotypic variability; (4) The prevalence is approximately 30%. When ascertainment is included in the sample, we simulate data under the model in (17) and select 500 cases and 500 controls at random to be retained in the sample.

We also consider generating the trait using a liability threshold model,

$$Y_i = \mathbb{1}_{\{\mathbf{L}_i > 0\}}, \text{ with } \mathbf{L}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{W}_i\boldsymbol{\alpha} + \epsilon_i, \tag{18}$$

where $\mathbf{L}_i$ is the latent liability for individual i. The values of $\boldsymbol{\beta}$ and σ_a^2 are chosen using the same conditions as for the logistic model, where conditions (1a), (1b), (2) and (4) are enforced on the liability scale instead of the logit scale, and condition (3) is used for the random error ϵ_i in (18), rather than for the Bernoulli error.

Larger values of σ_a^2 will correspond to more severe structure in the sample (from both population structure and family relatedness). We chose $D = 1$ for the number of top PCs of the GRM $\widehat{\boldsymbol{\Phi}}$ in (15) to include as covariates when fitting the null models in all seven resampling methods and when performing association testing with CARAT. We note that the pedigree information is only needed to simulate the data; it is not used when fitting the null models as the realized rather than the expected identity by descent (IBD) sharing across the genome is used to model correlation in the trait values.

Application to Domestic Dog Data from Hayward *et al.*

The dogs in the sample from [18] represent over 150 breeds, with an additional 170 from mixed breeds and 350 village dogs (i.e. dogs from free-breeding human-commensal populations). Data was obtained from the Dryad online repository [33]. For the ED trait, the data contain 113 cases and 633 controls among 82 breeds, and for IE, there are 34 cases and 168 controls from the Irish Wolfhound breed. After quality control for the markers [18], there are 150,418 and 98,350 autosomal markers (MAF $> 1\%$) for ED and

IE respectively. For ED, we exclude dogs that come from breeds not well represented in the data (fewer than 10 members). We also exclude dogs that are not most closely related to dogs of the same breed using hierarchical clustering with UPGMA [34, 35] (Table S2 and Fig S10). This results in 93 cases and 445 controls amongst 10 breeds (Table S3). We compute the GRM estimate $\widehat{\mathbf{\Phi}}$ in (15) from 138,192 markers with MAF $> 5\%$ (Fig S11), and determine the number of top PCs that are sufficient to capture the breed structure. Since there are 10 breeds in the data, it would be sensical for the top 9 PCs to capture genetic similarities due to breed membership, which is corroborated by plotting the top 10 PCs and the eigen spectrum (Figs S12 and S13). Since in some instances it may not be possible to have prior population membership information (e.g. breed membership), we use either the top 9 or the top 13 PCs to estimate subject-specific allele frequencies, from which we compute the GRM estimate $\tilde{\mathbf{\Phi}}$ in (16) that will represent within-breed genetic similarities. These top PCs are also included as covariates in the analysis of ED. For both traits, sex is included as a covariate.

Modification of CARAT

In order to assess the variance under the null of no association, CARAT uses a retrospective model for the tested marker $\mathbf{G}$ (see equation (14) in [12]). When we include the top PCs as covariates in our analysis, the GRM $\tilde{\mathbf{\Phi}}$ only reflects the leftover genetic correlation that is not captured by the top PCs. To incorporate all the genetic correlation, we consider the following retrospective model,

$$\mathbb{E}(\mathbf{G}|\mathbf{Y},\mathbf{Z}) = \mathbf{Z}\boldsymbol{\alpha} \text{ and } \mathrm{Var}(\mathbf{G}|\mathbf{Y},\mathbf{Z}) = \sigma_G^2\tilde{\mathbf{\Phi}}, \tag{19}$$

where $\mathbf{Z}$ is a matrix containing in its columns the top D PCs along with an intercept, and $\boldsymbol{\alpha}$ is an unknown vector for their effects. The model in (19) allows for different MAFs conditional on each subject's genetic ancestry captured in the top PCs. Hence, we modify CARAT and use $\widetilde{\sigma}_G^2 = \mathbf{G}^T\mathbf{P}\mathbf{G}/(n-D-1)$, with $\mathbf{P} = \tilde{\mathbf{\Phi}}^{-1} - \tilde{\mathbf{\Phi}}^{-1}\mathbf{Z}(\mathbf{Z}^T\tilde{\mathbf{\Phi}}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^T\tilde{\mathbf{\Phi}}^{-1}$, as an estimate of the variance parameter.

References

1. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: Study designs and statistical tests. American Journal of Human Genetics. 2014;95(1):5–23. doi:10.1016/j.ajhg.2014.06.009.

2. Dudbridge F, Gusnanto A. Estimation of significance thresholds for genomewide association scans. Genetic Epidemiology. 2008;32(3):227–234. doi:10.1002/gepi.20297.

3. Astle W, Balding DJ. Population structure and cryptic relatedness in genetic association studies. Statistical Science. 2009; p. 451–471.

4. Abney M, Ober C, McPeek MS. Quantitative-Trait Homozygosity and Association Mapping and Empirical Genomewide Significance in Large, Complex Pedigrees: Fasting Serum-Insulin Level in the Hutterites. The American Journal of Human Genetics. 2002;70(4):920–934. doi:10.1086/339705.

5. Churchill GA, Doerge RW. Naive application of permutation testing leads to inflated type I error rates. Genetics. 2008;178(1):609–610. doi:10.1534/genetics.107.074609.

6. Abney M. Permutation testing in the presence of polygenic variation. Genetic Epidemiology. 2015;39(4):249–258. doi:10.1002/gepi.21893.

7. Bianchi M, Dahlgren S, Massey J, Dietschi E, Kierczak M, Lund-Ziener M, et al. A multi-breed genome-wide association analysis for canine Hypothyroidism identifies a shared major risk locus on CFA12. PLoS ONE. 2015;10(8):e0134720. doi:10.1371/journal.pone.0134720.

8. Melin M, Rivera P, Arendt M, Elvers I, Murén E, Gustafson U, et al. Genome-Wide Analysis Identifies Germ-Line Risk Factors Associated with Canine Mammary Tumours. PLoS Genetics. 2016;12(5):e1006029. doi:10.1371/journal.pgen.1006029.

9. Safra N, Bassuk AG, Ferguson PJ, Aguilar M, Coulson RL, Thomas N, et al. Genome-Wide Association Mapping in Dogs Enables Identification of the Homeobox Gene, NKX2-8, as a Genetic Component of Neural Tube Defects in Humans. PLoS Genetics. 2013;9(7):e1003646. doi:10.1371/journal.pgen.1003646.

10. Tengvall K, Kierczak M, Bergvall K, Olsson M, Frankowiack M, Farias FHG, et al. Genome-Wide Analysis in German Shepherd Dogs Reveals Association of a Locus on CFA 27 with Atopic Dermatitis. PLoS Genetics. 2013;9(5):e1003475. doi:10.1371/journal.pgen.1003475.

11. Zhong S, Jiang D, McPeek MS. CERAMIC: Case-Control Association Testing in Samples with Related Individuals, Based on Retrospective Mixed Model Analysis with Adjustment for Covariates. PLoS Genetics. 2016;12(10):e1006329. doi:10.1371/journal.pgen.1006329.

12. Jiang D, Zhong S, McPeek MS. Retrospective Binary-Trait Association Test Elucidates Genetic Architecture of Crohn Disease. American Journal of Human Genetics. 2016;98(2):243–255. doi:10.1016/j.ajhg.2015.12.012.

13. Heyde CC. Quasi-likelihood and its application: a general approach to optimal parameter estimation. Springer Science & Business Media; 2008. Available from: `http://link.springer.com/10.1007/b98823`.

14. Rodriguez G, Goldman N. Improved estimation procedures for multilevel models with binary response: a case-study. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2001;164(2):339–355. doi:10.1111/1467-985X.00206.

15. Jang W, Lim J. A numerical study of PQL estimation biases in generalized linear mixed models under heterogeneity of random effects. Communications in Statistics: Simulation and Computation. 2009;38(4):692–702. doi:10.1080/03610910802627055.

16. Thornton TA. Statistical methods for genome-wide and sequencing association studies of complex traits in related samples. Current Protocols in Human Genetics. 2015;2015(January):1.28.1–1.28.9. doi:10.1002/0471142905.hg0128s84.

17. Price AL, j patterson N, m plenge R, e weinblatt M, a shadick N. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006;38(8):904–909. doi:10.1038/ng1847.

18. Hayward JJ, Castelhano MG, Oliveira KC, Corey E, Balkman C, Baxter TL, et al. Complex disease and phenotype mapping in the domestic dog. Nature Communications. 2016;7. doi:10.1038/ncomms10460.

19. Hodgkinson CA, Goldman D, Jaeger J, Persaud S, Kane JM, Lipsky RH, et al. Disrupted in schizophrenia 1 (DISC1): association with schizophrenia, schizoaffective disorder, and bipolar disorder. Am J Hum Genet. 2004;75(5):862–872. doi:10.1086/425586.

20. Iqbal Z, Vandeweyer G, Van der voet M, Waryah AM, Zahoor MY, Besseling JA, et al. Homozygous and heterozygous disruptions of ANK3: At the crossroads of neurodevelopmental and psychiatric disorders. Human Molecular Genetics. 2013;22(10):1960–1970. doi:10.1093/hmg/ddt043.

21. Duboscq-Bidot L, Xu P, Charron P, Neyroud N, Dilanian G, Millaire A, et al. Mutations in the Z-band protein myopalladin gene and idiopathic dilated cardiomyopathy. Cardiovascular Research. 2008;77(1):118–125. doi:10.1093/cvr/cvm015.

22. Meyer T, Ruppert V, Ackermann S, Richter A, Perrot A, Sperling SR, et al. Novel mutations in the sarcomeric protein myopalladin in patients with dilated cardiomyopathy. European Journal of Human Genetics. 2013;21(3):294–300. doi:10.1038/ejhg.2012.173.

23. Palmer EE, Jarrett KE, Sachdev RK, Zahrani FA, Hashem MO, Ibrahim N, et al. Neuronal deficiency of ARV1 causes an autosomal recessive epileptic encephalopathy. Human Molecular Genetics. 2016;25(14):3042–3054. doi:10.1093/hmg/ddw157.

24. Zukosky K, Meilleur K, Traynor BJ, Dastgir J, Medne L, Devoto M, et al. Association of a novel ACTA1 mutation with a dominant progressive scapuloperoneal myopathy in an extended family. JAMA Neurology. 2015;72(6):689–698. doi:10.1001/jamaneurol.2015.37.

25. Raudenbush SW, Yang ML, Yosef M. Maximum Likelihood for Generalized Linear Models with Nested Random Effects via High-Order, Multivariate Laplace Approximation. Journal of Computational and Graphical Statistics. 2000;9(1):141–157. doi:10.1080/10618600.2000.10474870.

26. Rabe-hesketh S, Skrondal A, Pickles A. Reliable estimation of generalized linear mixed models using adaptive quadrature. Stata Journal. 2002;2(1):1–21.

27. Conomos MP, Miller MB, Thornton TA. Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. Genetic Epidemiology. 2015;39(4):276–293. doi:10.1002/gepi.21896.

28. Chen H, Wang C, Conomos MP, Stilp AM, Li Z, Sofer T, et al. Control for Population Structure and Relatedness for Binary Traits in Genetic Association Studies via Logistic Mixed Models. American Journal of Human Genetics. 2016;98(4):653–666. doi:10.1016/j.ajhg.2016.02.012.

29. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nature Genetics. 2012;44(7):821–824. doi:10.1038/ng.2310.

30. Conomos MP, Reiner AP, Weir BS, Thornton TA. Model-free Estimation of Recent Genetic Relatedness. American Journal of Human Genetics. 2016;98(1):127–148. doi:10.1016/j.ajhg.2015.11.022.

31. Balding DJ, Nichols RA. A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. Genetica. 1995;96(1-2):3–12. doi:10.1007/BF01441146.

32. Marchini J, Cardon LR, Phillips MS, Donnelly P. The effects of human population structure on large genetic association studies. Nature Genetics. 2004;36(5):512–517. doi:10.1038/ng1337.

33. Hayward JJ, Castelhano MG, Oliveira KC, Corey E, Balkman C, Baxter TL, et al.. Data from: Complex disease and phenotype mapping in the domestic dog; 2016. Available from: `https://doi.org/10.5061/dryad.266k4`.

34. Murtagh F. Complexities of hierarchic clustering algorithms: state of the art. Computational Statistics Quarterly. 1984;1(2):101–113.

35. Sokal RR. A statistical method for evaluating systematic relationship. University of Kansas science bulletin. 1958;28:1409–1438.

Figure Titles and Legends

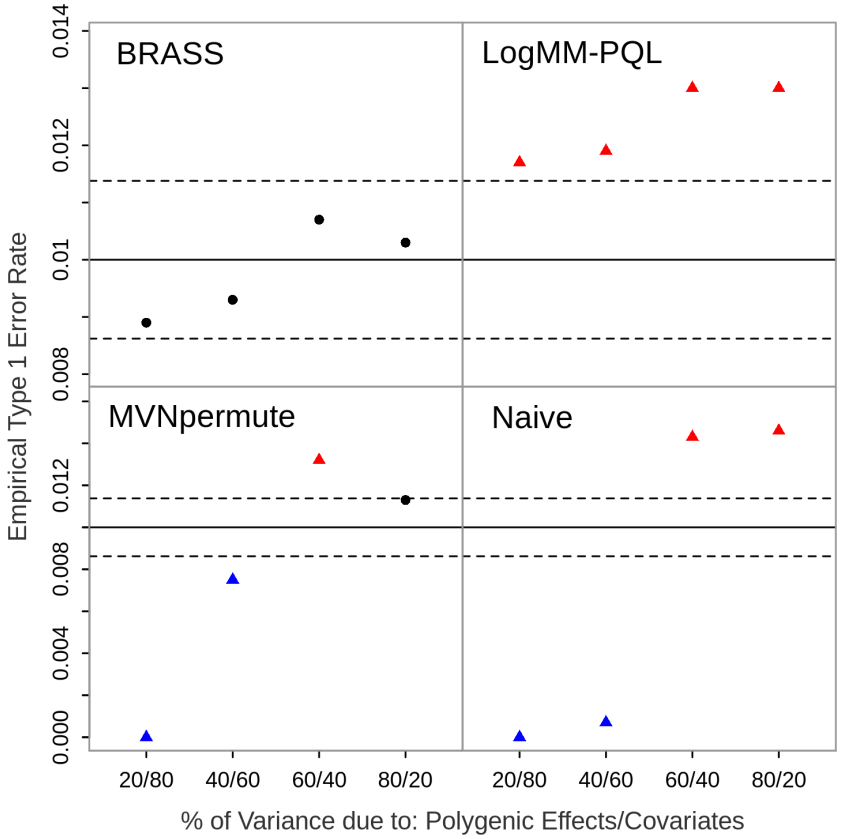

Fig 1. Empirical Type 1 Error Rates with Logistic Model Including all Covariates at Nominal Level 0.01.
The error rate is based on 20,000 simulated replicates. The solid horizontal line represents the nominal level and the dashed lines represent rejection bounds outside of which the z-test comparing the estimated type 1 error to the nominal level is rejected at level .05. Estimates inside the rejections bounds are represented by circles and those outside the bounds are represented by triangles. Red, black and blue symbols represent liberal, well-controlled and conservative type 1 error rates, respectively. The proportion of variability on the logit scale attributable to polygenic effects vs. covariates is varied from 20 to 80% in increments of 20%.

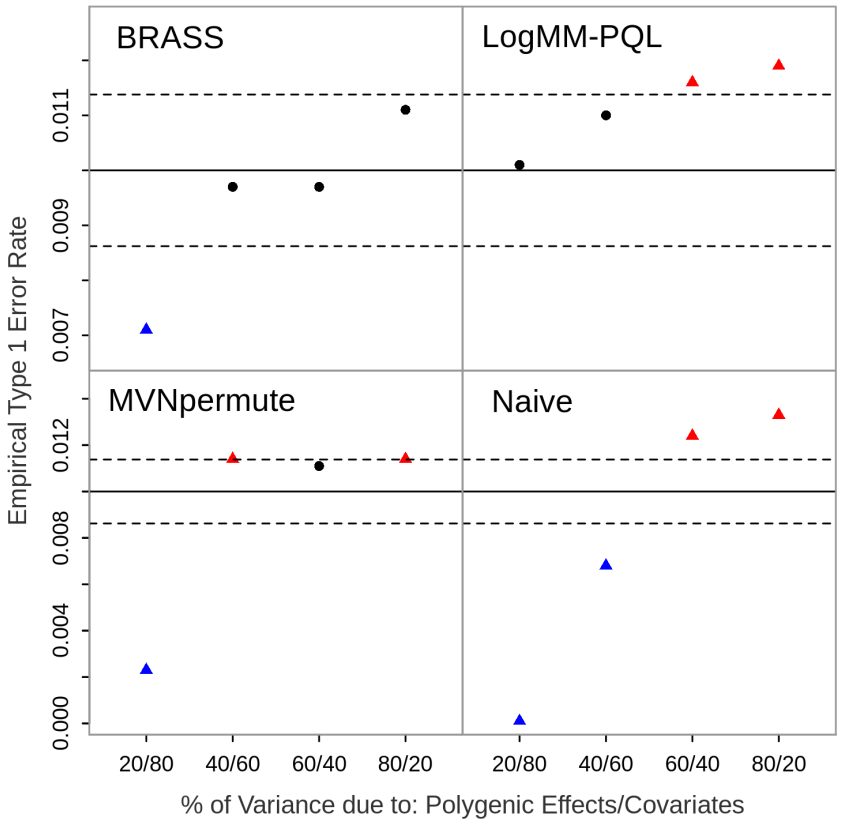

Fig 2. Empirical Type 1 Error Rates with Liability Threshold Model at Nominal Level 0.01.
The error rate is based on 20,000 simulated replicates. The solid horizontal line represents the nominal level and the dashed lines represent rejection bounds outside of which the z-test comparing the estimated type 1 error to the nominal level is rejected at level .05. Estimates inside the rejections bounds are represented by circles and those outside the bounds are represented by triangles. Red, black and blue symbols represent liberal, well-controlled and conservative type 1 error rates, respectively. The proportion of variability on the liability scale attributable to polygenic effects vs. covariates is varied from 20 to 80% in increments of 20%.

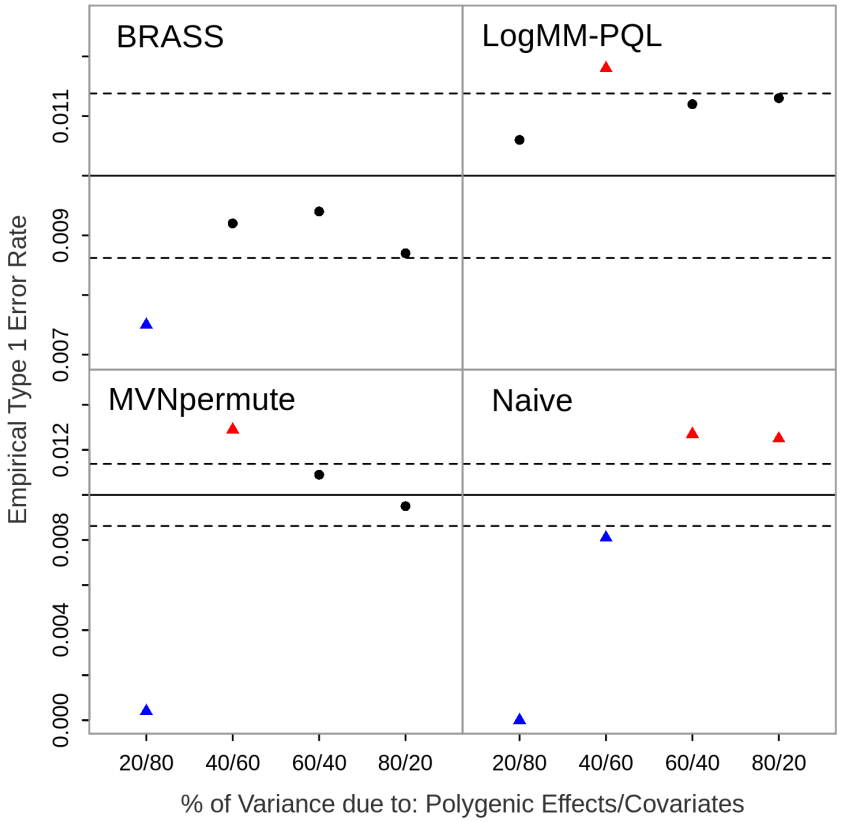

Fig 3. Empirical Type 1 Error Rates Omitting a Relevant Covariate at Nominal Level 0.01.
The error rate is based on 20,000 simulated replicates. The solid horizontal line represents the nominal level and the dashed lines represent rejection bounds outside of which the z-test comparing the estimated type 1 error to the nominal level is rejected at level .05. Estimates inside the rejections bounds are represented by circles and those outside the bounds are represented by triangles. Red, black and blue symbols represent liberal, well-controlled and conservative type 1 error rates, respectively. The proportion of variability on the logit scale attributable to polygenic effects vs. covariates is varied from 20 to 80% in increments of 20%. The effect of the omitted covariate on the trait correspond to a Wald test p-value of .05 using a linear mixed model.

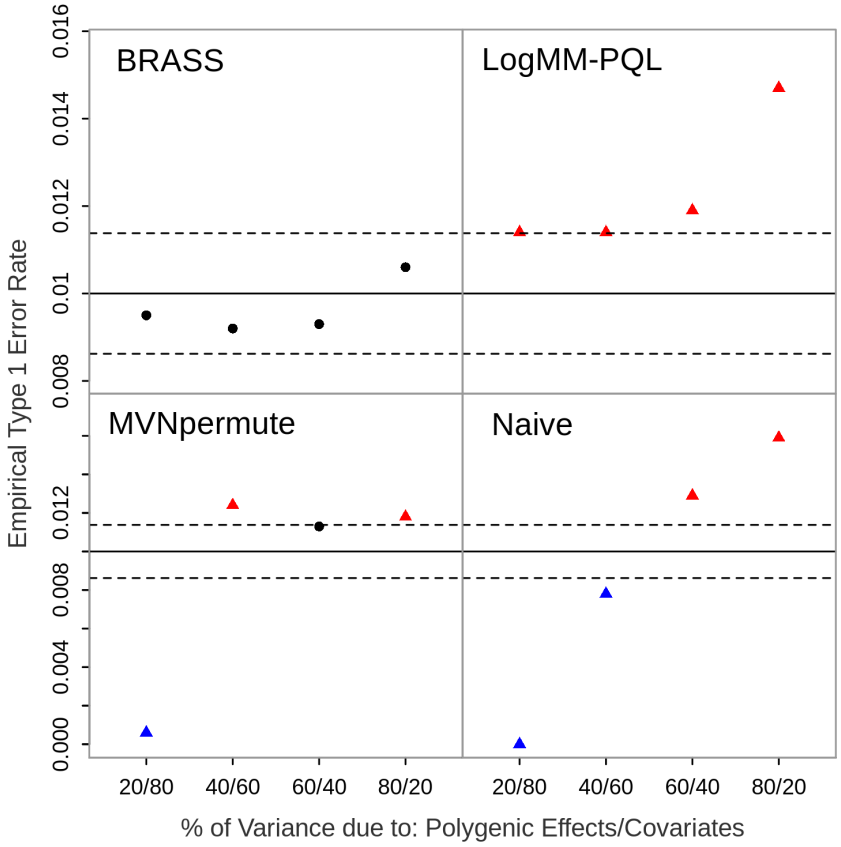

Fig 4. Empirical Type 1 Error Rates with Ascertainment Present at Nominal Level 0.01.
The error rate is based on 20,000 simulated replicates. The solid horizontal line represents the nominal level and the dashed lines represent rejection bounds outside of which the z-test comparing the estimated type 1 error to the nominal level is rejected at level .05. Estimates inside the rejections bounds are represented by circles and those outside the bounds are represented by triangles. Red, black and blue symbols represent liberal, well-controlled and conservative type 1 error rates, respectively. The proportion of variability on the logit scale attributable to polygenic effects vs. covariates is varied from 20 to 80% in increments of 20%.

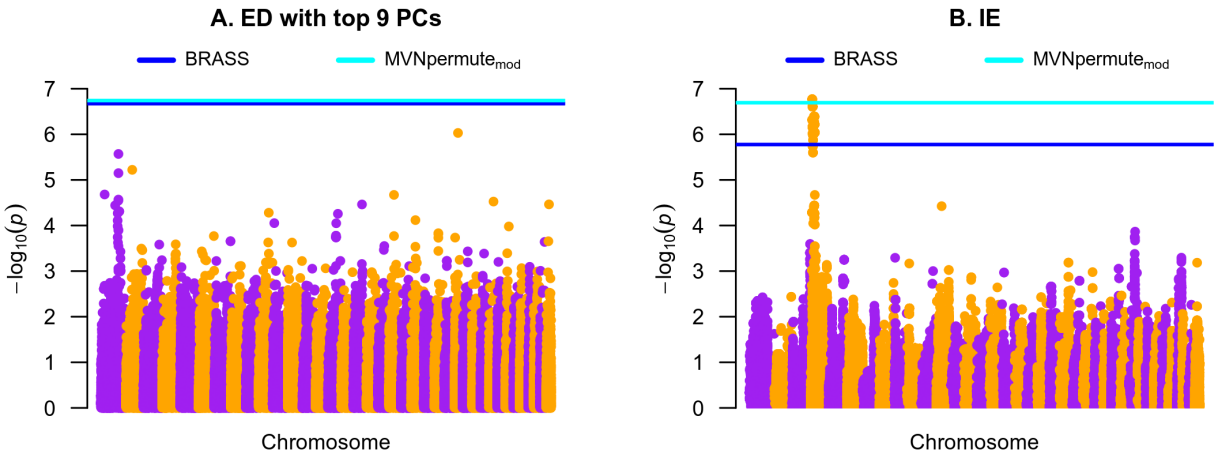

Fig 5. Genome screen results in the domestic dog data
Manhattan plots of the single-SNP association p-values using CARAT (on $-\log_{10}$ scale) for elbow dysplasia (ED) and idiopathic epilepsy (IE), with genomic position on the x-axis. The horizontal lines represent estimated genome-wide thresholds at nominal level 0.05 using 100,000 trait replicates: (A) 2.1×10^{-7} (BRASS) and 1.8×10^{-7} (MVNpermute$_{mod}$); (B) 1.67×10^{-6} (BRASS) and 2.03×10^{-7} (MVNpermute$_{mod}$). The Manhattan plot for the analysis of ED using the top 13 PCs (Fig S14) is similar to that in (A).

Tables

Table 1. Parameter null estimates in the domestic dog data for elbow dysplasia (ED) and idiopathic epilepsy (IE)

Parameter	Estimate	SE
VC ξ	0.712	–
Intercept	−1.2	0.38
Sex	−0.44	0.24

(a) ED: Quasi-likelihood model

Parameter	Estimate	SE
Additive variance σ_a^2	0.046	–
Total variance σ_T^2	0.130	–
Heritability σ_a^2/σ_T^2	0.38	0.16
Intercept	0.27	0.05
Sex	−0.06	0.03

(b) ED: Linear mixed model

Parameter	Estimate	SE
VC ξ	0.09	–
Intercept	0.08	0.60
Sex	−1.11	0.39

(c) IE: Quasi-likelihood model

Parameter	Estimate	SE
Additive variance σ_a^2	0.01	–
Total variance σ_T^2	0.13	–
Heritability σ_a^2/σ_T^2	0.09	0.11
Intercept	0.42	0.09
Sex	−0.16	0.05

(d) IE: Linear mixed model

The estimates are obtained using: (a) a quasi-likelihood model for ED fitted with CARAT; (b) a linear mixed model for ED fitted with GEMMA; (c) a quasi-likelihood model for IE fitted with CARAT; (d) a linear mixed model for IE fitted with GEMMA. For ED, the results are shown for the analysis including top 9 PCs as covariates (corresponding null estimates are omitted); the parameter null estimates using the top 13 PCs are included in Table S14. Sex is coded as 1=male, 2 = female.

Table 2. Top association signal for elbow dysplasia

					Genome-wide p-value	
SNP	Chr	Position	MAF	Single SNP p-value	BRASS	MVNpermute$_{mod}$
rs9000666	26	16554631	0.017	9.3×10^{-7}	0.148	0.178

The results are shown for the analysis including the top 9 PCs; the results with top 13 PCs are similar (not included).

Table 3. Top 3 association signals for idiopathic epilepsy in domestic dog data

						Genome-wide p-value	
SNP[a]	Chr	Position (Mb)	MAF	Nearest Genes[b]	P-value	BRASS	MVNpermute$_{mod}$
{rs9032722, rs24053634, rs24080710, rs24054032, rs24068259}	4	7.59-8.39	0.016	*ARV1, DISC1, FAM89A*	1.7×10^{-7}	0.0103	0.0454
{rs24158651, rs24077813, rs24089517, and 12 others}	4	9.5-10.1	0.16	*ACTA1, CCSAP, NUP133*	2.5×10^{-7}	0.0131	0.0560
{rs24072714, rs24111345, rs24058877, and 55 others}	4	16.5-19.4	0.18	*CTNNA3, DNAJC12, LRRTM3, MYPN*	4.1×10^{-7}	0.01830	0.07557

A 13Mb region on chromosome 4 (position 7.53-20.8 Mb) attained genome-wide significance with BRASS$_{mod}$ (the 52 remaining SNPs in that region are omitted from the table).

[a]SNPs within the brackets resulted in the same single-SNP test p-value.

[b]Top 3 nearest genes to SNPs in the brackets (using CanFam3.1). For the least significant SNP set, at least one of the SNPs was located within the listed genes. NCBI Gene ID for genes not mentioned in the text: $FAM89A$ (NCBI Gene: 488974), $CCSAP$ (NCBI Gene: 479209), $NUP133$ (NCBI Gene: 479208), $CTNNA3$ (NCBI Gene: 489008), $DNAJC12$ (NCBI Gene: 479224), $LRRTM3$ (NCBI Gene: 489009) .