

Genetic association models are robust to common population kinship estimation biases

Zhuoran Hou<sup>1</sup>, Alejandro Ochoa<sup>1,2,\*</sup>

<sup>4</sup> <sup>1</sup> Department of Biostatistics and Bioinformatics, Duke University, Durham, NC 27705, USA

<sup>5</sup> Duke Center for Statistical Genetics and Genomics, Duke University, Durham, NC 27705, USA

<sup>6</sup> \* Corresponding author: alejandro.ochoa@duke.edu

## Abstract

Common genetic association models for structured populations, including Principal Component Analysis (PCA) and Linear Mixed-effects Models (LMM), model the correlation structure between individuals using population kinship matrices, also known as Genetic Relatedness Matrices or “GRMs”. However, the most common kinship estimators can have severe biases that were only recently determined. Here we characterize the effect of these kinship biases on genetic association. We employ a large simulated admixed family and genotypes from the 1000 Genomes Project, both with simulated traits, to evaluate key kinship estimators. Remarkably, we find practically invariant association statistics for kinship matrices of different bias types (matching all other features). We then prove using statistical theory and linear algebra that LMM association tests are invariant to these kinship biases, and PCA approximately so. Our proof shows that the intercept and relatedness effect coefficients compensate for the kinship bias, an argument that extends to generalized linear models. As a corollary, association testing is also invariant to changing the reference ancestral population of the kinship matrix. Lastly, we observed that all kinship estimators, except for popkin ROM, can give improper non-positive semidefinite matrices, which can be problematic although some LMMs handle them surprisingly well, and condition numbers can be used to choose kinship estimators. Overall, we find that existing association studies are robust to kinship estimation bias, and our calculations may help improve association methods by taking advantage of this unexpected robustness, as well as help determine the effects of kinship bias in related problems.

27       **Abbreviations:** PCA: principal component analysis; PCs: principal components; LMM: linear  
28 mixed-effects model; MOR: mean of ratios; ROM: ratio of means; WG: Weir-Goudet (kinship  
29 estimator); MRCA: Most Recent Common Ancestor; SRMSD<sub>p</sub>: p-value Signed Root Mean Square  
30 Deviation; AUC<sub>PR</sub>: Area Under the Precision Recall Curve; GCTA: Genome-wide Complex Trait  
31 Analysis (software); PSD: positive semidefinite.

## 32     1 Introduction

33     The goal of genetic association is to detect loci that are related to a specific trait, either causally  
34 or by proximity to causal loci. When applied to structured populations with admixed individuals,  
35 multiethnic cohorts, or close relatives, controlling for relatedness is crucial to avoid spurious associa-  
36 tions and loss of power (Devlin and Roeder, 1999; Voight and Pritchard, 2005; Astle and Balding,  
37 2009; Yao and Ochoa, 2022). The most popular association models for structured populations are  
38 Linear Mixed-effects Models (LMM) and Principal Component Analysis (PCA), which are closely  
39 related except LMM is capable of modeling high-dimensional structures whereas PCA is strictly a  
40 low-dimensional model (Astle and Balding, 2009; Hoffman, 2013; Yao and Ochoa, 2022).

41     Various association models, including both PCA and LMM, parameterize relatedness using  
42 kinship matrices, also known as Genetic Relatedness Matrices or “GRMs”. Kinship coefficients  
43 are well suited for this task since they model the covariance structure of genotypes (Malécot, 1948;  
44 Jacquard, 1970). Kinship is often encountered in family studies, where they reflect recent relatedness  
45 and can be calculated from pedigrees (Wright, 1922; Emik and Terrill, 1949; García-Cortés, 2015).  
46 However, as kinship is defined as a probability of identity by descent, it may also capture ancient  
47 population relatedness (Malécot, 1948; Astle and Balding, 2009), and common non-parametric  
48 kinship estimators from genotypes indeed include population structure in their estimates (Ochoa and  
49 Storey, 2021). In LMMs, the kinship matrix is an explicit parameter determining the random effect  
50 covariance structure (Xie et al., 1998; Yu et al., 2006; Aulchenko et al., 2007; Astle and Balding,  
51 2009; Kang et al., 2008; Kang et al., 2010; Zhou and Stephens, 2012; Yang et al., 2014; Loh et al.,  
52 2015; Sul et al., 2018). In PCA, the principal components (PCs) are in practice the eigenvectors  
53 of an empirical genetic covariance matrix that is equivalent to the most common kinship estimator

54 (Price et al., 2006; Astle and Balding, 2009; Hoffman, 2013; Yao and Ochoa, 2022).

55 Although several kinship estimators have been used with LMMs in the past, work from the  
56 last 15 years has converged on what we call the “standard” kinship estimator, which is the same  
57 estimator used in PCA and other related models (Price et al., 2006; Astle and Balding, 2009;  
58 Rakovski and Stram, 2009; Thornton and McPeek, 2010; Yang et al., 2010; Yang et al., 2011; Zhou  
59 and Stephens, 2012; Speed et al., 2012; Yang et al., 2014; Speed and Balding, 2015; Loh et al.,  
60 2015; Wang et al., 2017; Sul et al., 2018). The impetus of our work is the recent characterization  
61 of a complex bias for this standard estimator, which varies for every pair of individuals (Weir  
62 and Goudet, 2017; Ochoa and Storey, 2021). These recent works also produced two new kinship  
63 estimators, which we are interested in characterizing in the context of association. The Weir-Goudet  
64 (WG) estimator constitutes a key improvement in that it has a uniformly downward bias (Weir and  
65 Goudet, 2017; Ochoa and Storey, 2021). Lastly, the popkin estimator is the only unbiased estimator  
66 under arbitrary relatedness (Ochoa and Storey, 2021). To the best of our knowledge, the new WG  
67 and popkin estimators have not been used in association studies before, but represent potential  
68 improvements over the use of the standard estimator for association.

69 One potential confounder when comparing the above kinship estimators is that the standard  
70 estimator upweights rare variants in a formulation previously called “mean-of-ratios” (MOR), whereas  
71 WG and popkin do not, instead following a “ratio-of-means” (ROM) estimation strategy (Bhatia  
72 et al., 2013; Ochoa and Storey, 2021). Recent work also formulated a ROM version of the standard  
73 estimator, which has a more predictable bias than the widely used MOR version (Ochoa and Storey,  
74 2021). Following a locus weight formulation that allows the standard estimator to weigh loci in both  
75 ways (Wang et al., 2017), here we generalize the popkin and WG estimators to have both MOR  
76 and ROM versions, to test estimators without confounding by locus weighing strategy.

77 In this work, we originally hypothesized that kinship estimation bias would affect association  
78 testing. We perform evaluations using an admixed family simulation (Yao and Ochoa, 2022) as  
79 well as real genotypes from the 1000 Genomes project (Consortium, 2010; 1000 Genomes Project  
80 Consortium et al., 2012; Fairley et al., 2020), in both cases with simulated traits, to characterize type  
81 I error control and power using robust statistics. Surprisingly, we find that both LMM and PCA

82 association statistics are largely invariant to kinship estimation bias. We theoretically characterize  
 83 the conditions under which these kinship biases result in invariant association statistics, which  
 84 encompass changing ancestral population in the kinship matrix too. As we discover that most  
 85 kinship estimates are non-positive semidefinite (non-PSD), breaking a key modeling assumption, we  
 86 perform additional empirical validations and discover that some LMMs can handle these improper  
 87 covariance matrices surprisingly well. Overall, we find that long-used association approaches are  
 88 unaffected by the most common kinship estimation biases, and develop theory that may help improve  
 89 association and related approaches such as heritability estimation.

## 90 2 Methods

### 91 2.1 Genetic model

92 The following genetic model justifies the use of kinship matrices in association studies, and is the  
 93 basis of all kinship estimation bias calculations that our theoretical work depends upon.

94 Suppose there are  $m$  biallelic loci and  $n$  diploid individuals. The genotype  $x_{ij} \in \{0, 1, 2\}$  at a  
 95 locus  $i$  of individual  $j$  is encoded as the number of reference alleles, for a preselected but otherwise  
 96 arbitrary reference allele per locus. Genotypes are treated as random variables structured according  
 97 to relatedness. If  $T$  is the ancestral population on which allele frequencies are conditioned,  $\varphi_{jk}^T$  is  
 98 the kinship coefficient of two individuals  $j$  and  $k$ , and  $p_i^T$  is the ancestral allele frequency at locus  
 99  $i$ , then under the kinship model (Malécot, 1948; Wright, 1949; Jacquard, 1970; Astle and Balding,  
 100 2009; Ochoa and Storey, 2021) the expectation and covariance are given by

$$E[\mathbf{x}_i|T] = 2p_i^T \mathbf{1}, \quad \text{Cov}(\mathbf{x}_i|T) = 4p_i^T(1 - p_i^T) \boldsymbol{\Phi}^T,$$

101 where  $\mathbf{x}_i = (x_{ij})$  is the length- $n$  column vector of genotypes at locus  $i$ ,  $\boldsymbol{\Phi}^T = (\varphi_{jk}^T)$  is the  $n \times n$   
 102 kinship matrix, and  $\mathbf{1}$  is a length- $n$  column vector of ones. Both  $\boldsymbol{\Phi}^T$  and  $p_i^T$  are parameters that  
 103 depend on the choice of ancestral population, for which the Most Recent Common Ancestor (MRCA)  
 104 population is the most sensible choice (Ochoa and Storey, 2021). However, one of the results of this  
 105 work is proof that the choice of ancestral population does not affect association testing.

106 **2.2 Kinship estimators**

107 Each subsection below corresponds to a kinship estimator bias type: Popkin is unbiased, while  
108 Standard and WG have different bias functions (defined shortly). Each estimator bias type has two  
109 locus weight types called *ratio-of-means* (ROM) and *mean-of-ratios* (MOR), a terminology that  
110 follows previous convention for these and related estimators (Bhatia et al., 2013; Ochoa and Storey,  
111 2021). Only ROM estimators have closed-form limits. Below  $\hat{p}_i^T = \frac{1}{2n}\mathbf{x}_i^\top \mathbf{1}$  is the standard ancestral  
112 allele frequency estimator, where the  $\top$  superscript denotes matrix transposition (do not confuse  
113 with ancestral population superscript  $T$ ), and  $\hat{\Phi}^{T,\text{name}} = (\hat{\varphi}_{jk}^{T,\text{name}})$  relates the scalar and matrix  
114 formulas of each named kinship estimator. In our evaluations, all loci were used to estimate kinship  
115 and to test for association, as is common practice.

116 **2.2.1 Popkin estimator**

117 The popkin (population kinship) estimator (Ochoa and Storey, 2021), generalized here to include  
118 locus weights  $w_i$ , is given by

119

$$\hat{\varphi}_{jk}^{T,\text{popkin}} = 1 - \frac{A_{jk}}{\hat{A}_{\min}}, \quad A_{jk} = \frac{1}{m} \sum_{i=1}^m w_i((x_{ij} - 1)(x_{ik} - 1) - 1), \quad (1)$$

120 where in this work  $\hat{A}_{\min} = \min_{j \neq k} A_{jk}$ , and  $w_i$  must be positive but need not add to 1. We consider  
121 two broad forms for this estimator. The original ROM estimator has  $w_i = 1$  and has an unbiased  
122 almost sure limit as the number of loci  $m$  go to infinity,

$$\hat{\Phi}^{T,\text{popkin-ROM}} \xrightarrow[m \rightarrow \infty]{\text{a.s.}} \Phi^T,$$

123 under the assumption that the true minimum kinship is zero. The MOR version, introduced here,  
124 upweights rare variants by using  $w_i = (\hat{p}_i^T(1 - \hat{p}_i^T))^{-1}$ ; although it has no closed-form limit, it is  
125 approximately unbiased as well (Appendix A) and it is connected to the most common estimator,  
126 Standard MOR (Appendix B). The use of locus weights here is inspired by previous calculations  
127 relating the standard ROM and MOR estimators (Wang et al., 2017).

<sup>128</sup> **2.2.2 Standard estimator**

<sup>129</sup> The ROM and MOR versions of the standard kinship estimator are, respectively,

$$\hat{\varphi}_{jk}^{T,\text{std-ROM}} = \frac{\sum_{i=1}^m (x_{ij} - 2\hat{p}_i^T)(x_{ik} - 2\hat{p}_i^T)}{\sum_{i=1}^m 4\hat{p}_i^T(1 - \hat{p}_i^T)}, \quad (2)$$

$$\hat{\varphi}_{jk}^{T,\text{std-MOR}} = \frac{1}{m} \sum_{i=1}^m \frac{(x_{ij} - 2\hat{p}_i^T)(x_{ik} - 2\hat{p}_i^T)}{4\hat{p}_i^T(1 - \hat{p}_i^T)}. \quad (3)$$

<sup>130</sup> The ROM estimator has a biased limit, which is a function of the true kinship matrix (Ochoa and

<sup>131</sup> Storey, 2021):

$$\hat{\Phi}^{T,\text{std-ROM}} \xrightarrow[m \rightarrow \infty]{\text{a.s.}} F^{\text{std}}(\Phi^T) = \frac{1}{1 - \bar{\varphi}^T} \left( \Phi^T + \bar{\varphi}^T \mathbf{J} - \boldsymbol{\varphi}^T \mathbf{1}^\top - \mathbf{1} (\boldsymbol{\varphi}^T)^\top \right), \quad (4)$$

<sup>133</sup> where  $\mathbf{J} = \mathbf{1}\mathbf{1}^\top$  is the  $n \times n$  matrix of ones,  $\boldsymbol{\varphi}^T = \frac{1}{n} \Phi^T \mathbf{1}$  is a length- $n$  vector of per-row mean

<sup>134</sup> kinship values, and  $\bar{\varphi}^T = \frac{1}{n^2} \mathbf{1}^\top \Phi^T \mathbf{1}$  is the scalar overall mean kinship. The MOR estimator does

<sup>135</sup> not have closed-form limit, but it is well approximated by Eq. (4) in practice, especially when loci

<sup>136</sup> with small minor allele frequencies are excluded prior to calculating this estimate. In Appendix B

<sup>137</sup> we prove that, when there are no missing genotypes, the two standard estimators are functions of

<sup>138</sup> the corresponding popkin estimators, given by the bias function  $F^{\text{std}}$ :

$$\hat{\Phi}^{T,\text{std-ROM}} = F^{\text{std}}(\hat{\Phi}^{T,\text{popkin-ROM}}),$$

$$\hat{\Phi}^{T,\text{std-MOR}} = F^{\text{std}}(\hat{\Phi}^{T,\text{popkin-MOR}}).$$

<sup>139</sup> **2.2.3 Weir-Goudet estimator**

<sup>140</sup> The ROM version of the Weir-Goudet (WG) kinship estimator is given by (Weir and Goudet, 2017;

<sup>141</sup> Ochoa and Storey, 2021)

$$\hat{\varphi}_{jk}^{T,\text{WG-ROM}} = 1 - \frac{A_{jk}}{\hat{A}_{\text{avg}}}, \quad \hat{A}_{\text{avg}} = \frac{2}{n(n-1)} \sum_{j=2}^n \sum_{k=1}^{j-1} A_{jk}, \quad (5)$$

143 where  $A_{jk}$  is as in Eq. (1). Its biased limit is also a function of the true kinship matrix:

$$\hat{\Phi}^{T,\text{WG-ROM}} \xrightarrow[m \rightarrow \infty]{\text{a.s.}} F^{\text{WG}}(\Phi^T) = \frac{1}{1 - \tilde{\varphi}^T} (\Phi^T - \tilde{\varphi}^T \mathbf{J}), \quad (6)$$

145 where  $\tilde{\varphi}^T$  is the mean kinship excluding the matrix diagonal:

$$\tilde{\varphi}^T = \frac{2}{n(n-1)} \sum_{j=2}^n \sum_{k=1}^{j-1} \varphi_{jk}^T. \quad (7)$$

147 In Appendix C we prove that

$$0 \leq \tilde{\varphi}^T \leq \bar{\varphi}^T \leq 1,$$

148 and equalities are achieved if and only if all kinship values are equal. Since the WG-ROM estimator

149 closely resembles the popkin estimator in Eq. (1), it is clear that they are related by the bias function

150  $F^{\text{WG}}$ , while WG-MOR is introduced here and defined by the below formula:

$$\begin{aligned} \hat{\Phi}^{T,\text{WG-ROM}} &= F^{\text{WG}}(\hat{\Phi}^{T,\text{popkin-ROM}}), \\ \hat{\Phi}^{T,\text{WG-MOR}} &= F^{\text{WG}}(\hat{\Phi}^{T,\text{popkin-MOR}}). \end{aligned}$$

### 151 2.3 Association models

152 LMM and PCA are closely related association models (Astle and Balding, 2009; Hoffman, 2013;

153 Yao and Ochoa, 2022):

$$\text{LMM: } \mathbf{y} = \mathbf{1}\alpha + \mathbf{x}_i\beta_i + \mathbf{s} + \boldsymbol{\epsilon}, \quad (8)$$

$$\mathbf{s} \sim \text{Normal}(\mathbf{0}, 2\sigma^2 \Phi^T), \quad (9)$$

$$\text{PCA: } \mathbf{y} = \mathbf{1}\alpha + \mathbf{x}_i\beta_i + \mathbf{U}_r\boldsymbol{\gamma}_r + \boldsymbol{\epsilon}, \quad (10)$$

$$\Phi^T = \mathbf{U}\Lambda\mathbf{U}^\top, \quad (11)$$

154 where  $\mathbf{y}$  is a length- $n$  vector of continuous trait values,  $\alpha$  is the intercept coefficient,  $\beta_i$  is the genetic

155 effect (association) coefficient of locus  $i$ ,  $\mathbf{s}$  is the (genetic) random effect,  $\sigma^2$  is the random effect

variance factor,  $\mathbf{U}_r$  is the  $n \times r$  matrix of the  $r$  eigenvectors (PCs) with the largest eigenvalues of  $\Phi^T$ ,  $\gamma_r$  is a length- $r$  vector of coefficients for each eigenvector,  $\epsilon \sim \text{Normal}(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I})$  are random independent residuals, and  $\mathbf{I}$  is the  $n \times n$  identity matrix. Furthermore, Eq. (11) is the complete eigendecomposition of  $\Phi^T$ , where  $\mathbf{U}$  is the  $n \times n$  matrix of eigenvectors, and  $\Lambda$  is the  $n \times n$  diagonal matrix of eigenvalues. As  $\mathbf{s}$  and  $\mathbf{U}_r$  play analogous roles in modeling the effect of relatedness in LMM and PCA, respectively, we refer to them jointly as relatedness effects, and  $\sigma^2$  and  $\gamma_r$  as their coefficients.

## 2.4 Simulations

### 2.4.1 Admixed family genotype simulation

An admixed family is simulated following previous work (Yao and Ochoa, 2022), except here only  $K = 3$  ancestries are simulated and  $FST = 0.3$  for the admixed individuals, which more closely resembles Hispanics and African Americans. Briefly, our admixture model first simulates  $n = 1000$  founder individuals with  $m = 100,000$  loci, which was purposefully reduced compared to previous work to increase the difference between estimated kinship matrices (which will be noisier) and their limits. Random ancestral allele frequencies  $p_i^T$ , subpopulation allele frequencies  $p_i^{S_u}$ , individual-specific allele frequencies  $\pi_{ij}$ , and genotypes  $x_{ij}$  are drawn from this hierarchical model:

$$\begin{aligned}
 p_i^T &\sim \text{Uniform}(0.01, 0.5), \\
 p_i^{S_u} | p_i^T &\sim \text{Beta}\left(p_i^T \left(\frac{1}{f_{S_u}^T} - 1\right), (1 - p_i^T) \left(\frac{1}{f_{S_u}^T} - 1\right)\right), \\
 \pi_{ij} &= \sum_{u=1}^K q_{ju} p_i^{S_u}, \\
 x_{ij} | \pi_{ij} &\sim \text{Binomial}(2, \pi_{ij}),
 \end{aligned}$$

where this Beta is the Balding-Nichols distribution (Balding and Nichols, 1995) with mean  $p_i^T$  and variance  $p_i^T (1 - p_i^T) f_{S_u}^T$ . This is implemented in the R package `bnpd`.

We also include family structure in the simulation. 20 generations are generated iteratively. Individuals in the first generation ( $n = 1000$ ) are ordered by 1D geography, randomly assigned

176 sex, and treated as locally unrelated. From the next generation, individuals are paired iteratively:  
 177 randomly choosing males from the pool and pairing them with the nearest available female with  
 178 local kinship  $< 1/4^3$  (to preserve the admixture structure) until there are no available males or  
 179 females. Family sizes are drawn randomly ensuring every family has at least one child. Children  
 180 are reordered by the average coordinates of their parents, their sex are assigned randomly, and their  
 181 alleles are drawn from parents independently per locus. The simulation is implemented in the R  
 182 package `simfam`.

183 **2.4.2 Trait simulation algorithm**

184 Given an  $m \times n$  genotype matrix  $\mathbf{X} = (\mathbf{x}_i^\top)$ , traits are simulated from

$$\mathbf{y} = \mathbf{1}\alpha + \mathbf{X}^\top \boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \text{Normal}(\mathbf{0}, (1 - h^2)\mathbf{I}).$$

185 Given a desired heritability  $h^2$  (0.8 or 0.3 in this work) and the number of causal loci  $m_1$  (here  
 186 chosen using  $m_1 = \text{round}(nh^2/8)$ , which empirically balances power as sample size and heritability  
 187 are varied), the goal is to choose causal coefficients  $\boldsymbol{\beta}$  and the intercept  $\alpha$  that result in zero mean and  
 188 the desired trait heritability. Here, we use the “fixed effect sizes” trait simulation model described  
 189 in (Yao and Ochoa, 2022). Briefly, first  $m_1$  causal loci are randomly selected, and for these steps  
 190 only  $\mathbf{X}$  is subset to these loci and reindexed. For known  $p_i^T$ , causal coefficients are constructed as:

$$\beta_i = \sqrt{\frac{h^2}{2m_1 v_i^T}},$$

191 where  $v_i^T = p_i^T (1 - p_i^T)$ ; for unknown  $p_i^T$  (real genotypes), the unbiased estimate  $\hat{v}_i^T = \hat{p}_i^T (1 - \hat{p}_i^T) / (1 - \bar{\varphi}^T)$   
 192 is used, where  $\bar{\varphi}^T$  is the mean kinship estimated from `popkin`. Coefficients are made negative  
 193 randomly with probability 0.5. For known  $p_i^T$ , we obtain the desired zero trait mean with  $\alpha =$   
 194  $-2(\mathbf{p}^T)^\top \boldsymbol{\beta}$ , where here  $\mathbf{p}^T = (p_i^T)$  contains causal loci only. When  $p_i^T$  are unknown, to avoid  
 195 covariance distortions, the intercept coefficient is constructed as

$$\alpha = -2\hat{p}^T \mathbf{1}_{m_1}^\top \boldsymbol{\beta}, \quad \hat{p}^T = \frac{1}{m_1} \mathbf{1}_{m_1}^\top \hat{\mathbf{p}}^T,$$

196 where  $\mathbf{1}_{m_1}$  is a length- $m_1$  column vector of ones. Genotypes were resimulated from the admixed  
197 family model for each heritability value and every replicate.

## 198 2.5 Real genotype data processing

199 To evaluate different kinship estimators on a real dataset, we use the high-coverage NYGC version  
200 of the 1000 Genomes Project (Fairley et al., 2020), which is processed as before (Yao and Ochoa,  
201 2022). Briefly, using `plink2` (Chang et al., 2015) we keep only autosomal biallelic SNP loci with  
202 filter “PASS”, pruned for linkage disequilibrium with parameters “`--indep-pairwise 1000kb 0.3`”  
203 to remove loci that have a greater than 0.3 squared correlation coefficient with other loci within  
204 1000kb, and lastly remove loci with minor allele frequencies < 0.01. The resulting data have  
205  $m = 1,111,266$  loci and  $n = 2,504$  individuals.

## 206 2.6 Evaluation of performance

207  $AUC_{PR}$  and  $SRMSD_p$  are used to evaluate approaches as before (Yao and Ochoa, 2022). Briefly,  
208  $SRMSD_p$  (Signed Root Mean Square Deviation) measures the difference between the observed null  
209 p-value quantiles and the expected uniform quantiles:

$$SRMSD_p = \text{sgn}(u_{\text{median}} - p_{\text{median}}) \sqrt{\frac{1}{m_0} \sum_{i=1}^{m_0} (u_i - p_{(i)})^2},$$

210 where  $m_0 = m - m_1$  is the number of null (non-causal) loci,  $i$  indexes null loci only,  $p_{(i)}$  is the  $i$ th  
211 ordered null p-value,  $u_i = (i - 0.5)/m_0$  is its expectation,  $p_{\text{median}}$  is the median observed null p-value,  
212  $u_{\text{median}} = \frac{1}{2}$  is its expectation, and  $\text{sgn}$  is the sign function (1 if  $u_{\text{median}} \geq p_{\text{median}}$ , -1 otherwise).  
213  $SRMSD_p = 0$  corresponds to calibrated p-values,  $SRMSD_p > 0$  indicate anti-conservative p-values,  
214 and  $SRMSD_p < 0$  are conservative p-values.

215  $AUC_{PR}$  (Area Under the Precision and Recall Curve) is a binary classification measure that  
216 reflects calibrated power (Yao and Ochoa, 2022), which is calculated from the total numbers of true

217 positives (TP), false positives (FP) and false negatives (FN) at some threshold or parameter  $t$ :

$$\text{Precision}(t) = \frac{\text{TP}(t)}{\text{TP}(t) + \text{FP}(t)},$$
$$\text{Recall}(t) = \frac{\text{TP}(t)}{\text{TP}(t) + \text{FN}(t)},$$

218 followed by calculating the area under the curve traced as  $t$  varies recall from zero to one. Higher  
219  $\text{AUC}_{\text{PR}}$  is better, with best performance at  $\text{AUC}_{\text{PR}} = 1$  for a perfect classifier, while worst perfor-  
220 mance at  $\text{AUC}_{\text{PR}} = \frac{m_1}{m}$  (overall proportion of causal loci) is for random classifiers.

## 221 2.7 Software

222 Popkin kinship estimates are calculated with the `popkin` R package. Standard MOR kinship esti-  
223 mates are calculated with GCTA (version 1.93.2beta). All other kinship estimators and limits are  
224 calculated using the `popkinsuppl` R package. PCs are calculated with the `eigen` function of R.

225 GCTA, which implements Eqs. (8) and (9), is used to run all LMM associations (Yang et al.,  
226 2011; Yang et al., 2014). We pass  $2\Phi^T$  for all kinship matrices tested (the same scale as its own  
227 kinship estimate). `plink2`, which implements Eqs. (10) and (11), performs the PCA associations  
228 (Chang et al., 2015). We use  $r = K - 1 = 2$  PCs for the admixed family simulations, and  $r = 10$   
229 PCs for 1000 Genomes.

## 230 3 Results

### 231 3.1 Empirical analysis using admixed family simulation

232 To quantify the effect of kinship estimation bias, we simulate genotypes and traits, and calculate  
233 association p-values using a factorial design that tests all kinship matrix (3 bias types, times two lo-  
234 cus weight types and one limit) and association model (PCA and LMM) combinations. We simulate  
235 an admixed population with  $K = 3$  ancestries, who serve as founders for a 20-generation random  
236 pedigree. This high-dimensional admixed family scenario yields a large difference in performance  
237 between PCA and LMM (Yao and Ochoa, 2022).

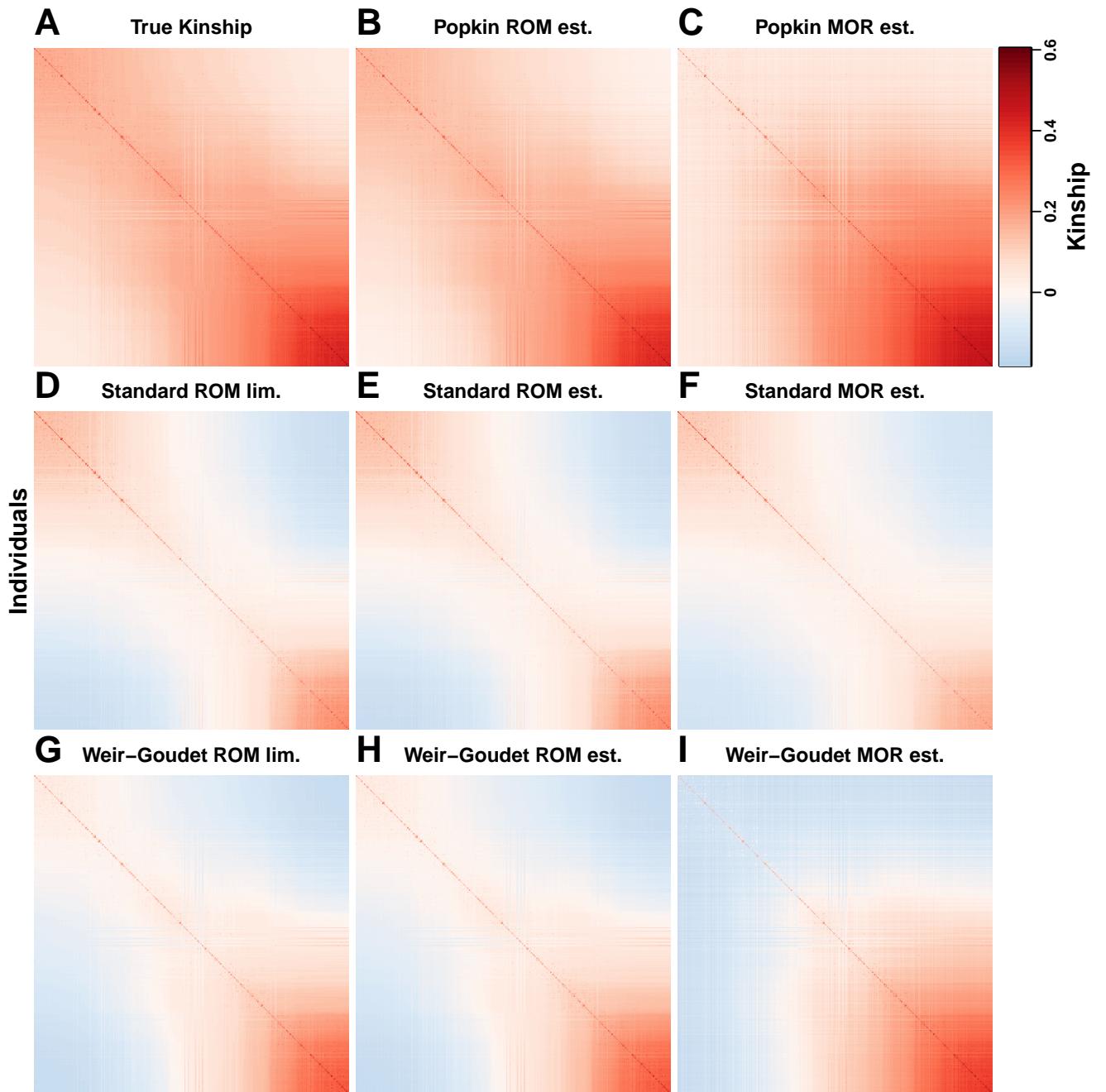


Figure 1: **Kinship estimates and limits on the admixed family simulation.** Each panel shows a kinship matrix as a heatmap, with each of the  $n = 1000$  individuals along both x and y axes, color represents kinship: positive values in red, negative in blue. Diagonal contains inbreeding values. Each estimator bias type (Popkin, Standard, and Weir-Goudet; rows) has three matrices (columns): two locus weight types (ROM (ratio of means) and MOR (mean of ratios)) and limit of ROM.

238 Kinship estimates and limits for this simulation are shown in Fig. 1. The true kinship matrix  
 239 shows the family relatedness as high values concentrated near the diagonal and the ancestry-driven  
 240 population structure as the broad patterns off-diagonal. Only Popkin ROM is unbiased, while pop-  
 241 kin MOR has a slight upward bias that varies across the matrix (Fig. S1A). In contrast, the Standard  
 242 and Weir-Goudet (WG) estimates have large downward biases overall, resulting in abundant nega-  
 243 tive values; Standard biases vary for every pair of individuals (as described in Eq. (4)), while WG  
 244 has a uniform bias (following Eq. (6)). The difference is most noticeable near the diagonal: the  
 245 true kinship matrix has monotonically-increasing values, WG has smaller values but which are still  
 246 monotonically increasing, and Standard estimates follow a U-shaped pattern (decreasing at first,  
 247 then increasing again in Fig. 1D-F).

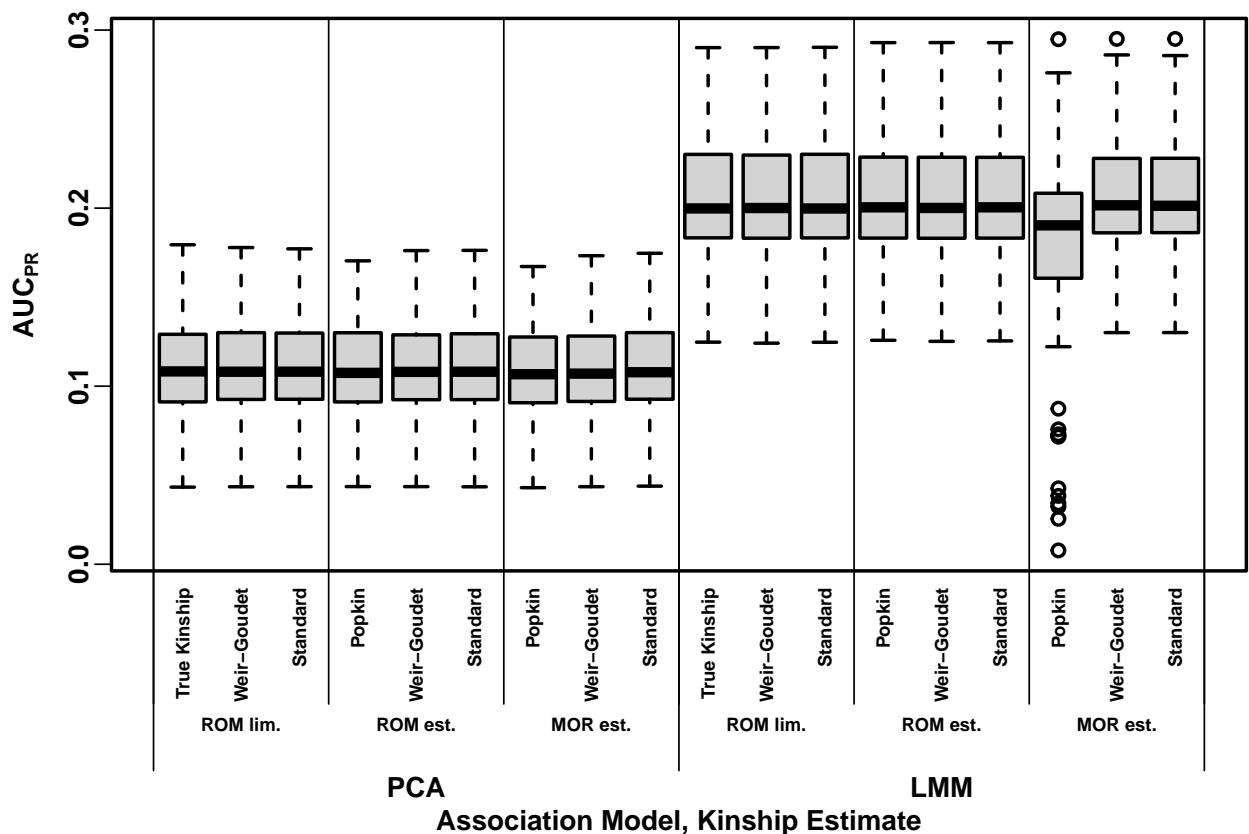


Figure 2: **Distributions of Area Under the Precision-Recall Curve (AUC<sub>PR</sub>) on the admixed family simulation with  $h^2 = 0.8$ .** Higher AUC<sub>PR</sub> is better performance. Results for 100 replicates (each a random genotype matrix and trait vector). Approaches cluster primarily by association model (LMM or PCA), and vary little across bias types.

248 We perform LMM and PCA association tests to determine how kinship biases affect association  
249 performance. Surprisingly, we find that kinship bias type does not have a discernible effect on asso-  
250 ciation performance, as summarized by  $AUC_{PR}$  (a robust proxy for power; high and low heritability  
251 in Figs. 2 and S2, respectively) and  $SRMSD_p$  (measures null statistic calibration; Figs. S3 and S4).  
252 The largest differences in performance are explained by the association model (LMM vs PCA), as  
253 expected due to our use of a family simulation where PCA performs poorly. Within association  
254 models, there are no clear differences between the performance of any of the kinship matrices, in fact  
255 many appear to have identical distributions (both statistics), the only clear exception being LMM  
256 popkin MOR with  $h^2 = 0.8$ , which has a few outlier replicates where performance is exceedingly  
257 poor (at the end of the results we show these are due to limited numerical precision exacerbated by  
258 high condition numbers of trait covariance matrices).

259 To better characterize the nearly identical performance distribution just observed, we next mea-  
260 sure the agreement between individual association p-values. We measure high correlations between  
261 p-values, near 1 for comparisons involving the same model (between LMM methods or between  
262 PCA methods, both heritabilities), and across models around 0.6 for  $h^2 = 0.8$ , which increases to  
263 0.88 for  $h^2 = 0.3$  (Figs. S5 and S6). To measure numerical agreement more stringently, we calculate  
264 the proportion of loci between two methods with p-values within 0.01 of each other, and find a  
265 remarkably high agreement between estimators of different bias types after matching association  
266 model and locus weight type or limit (Figs. 3 and S7). This is in contrast to low agreement between  
267 PCA and LMM statistics, and between LMM statistics with different locus weight types or limits.  
268 Minimum agreements are higher across PCA methods, though here the true kinship or popkin esti-  
269 mates disagree more from Standard and WG matrices. Overall, kinship matrices with different bias  
270 types (otherwise matched) result in nearly identical association statistics.

### 271 3.2 Empirical analysis using 1000 Genomes

272 Now we repeat our analysis using the real genotypes of 1000 Genomes. Kinship estimates are shown  
273 in Fig. 4 (note real data have no true kinship or estimator limits). Popkin ROM estimates display an  
274 approximate nested block structure that arises from the tree relationships between subpopulations

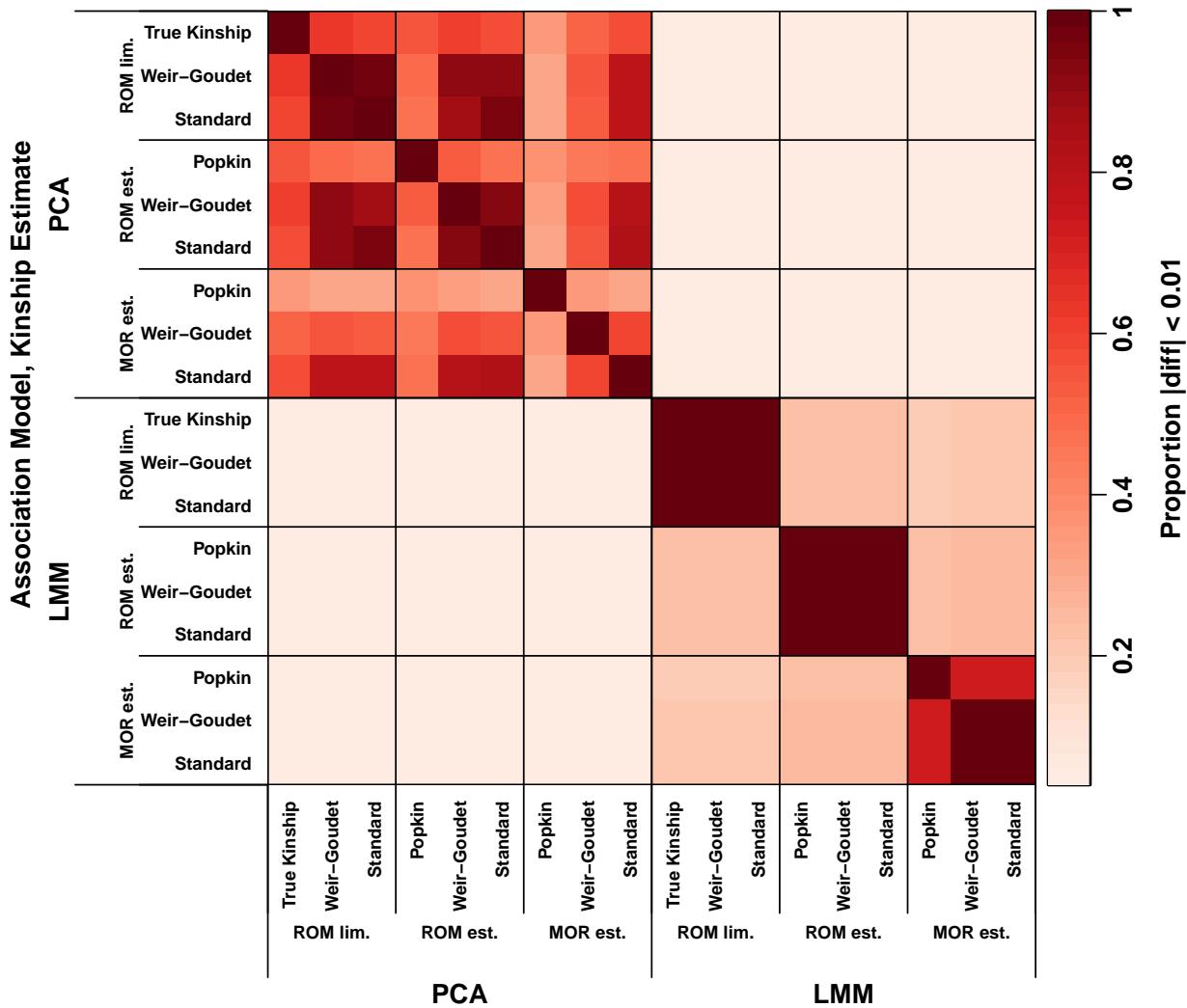


Figure 3: **Agreement between p-values on the admixed family simulation with  $h^2 = 0.8$ .** Calculated agreement (absolute difference under 0.01) averaged over loci (color) of association p-values between association models (LMM vs PCA) and kinship matrices (x and y axes). All 100 replicates are used. Different bias types (matched for association model and locus weight type) have large proportions of nearly identical p-values.

275 (Fig. 4A; trees were explicitly fit to this data in previous work (Yao and Ochoa, 2022)). However,  
 276 popkin MOR estimates do not follow the nested blocks tree structure, since kinship between African  
 277 and non-African populations is higher than kinship within African populations (Fig. 4B). Standard  
 278 estimates have values closer to zero, and a different bias for each pair of individuals, resulting in  
 279 higher relative kinship for African compared to non-African populations (Fig. 4C-D). Lastly, WG  
 280 estimates are uniformly smaller than popkin's and attain large negative values (Fig. 4E-F).

281 Our association test conclusion are similar to our simulation study:  $AUC_{PR}$  and  $SRMSD_p$   
 282 distributions are nearly identical for estimators of different bias types but same locus weight type  
 283 (ROM or MOR) and association model. However, unlike the simulation, here for  $h^2 = 0.8$  (but  
 284 not 0.3) the MOR estimates noticeably outperform ROM estimates (LMM only), in terms of both  
 285  $AUC_{PR}$  (Figs. 5 and S8) and  $SRMSD_p$  (Figs. S9 and S10). P-values are even more highly correlated  
 286 in this case (Figs. S11 and S12), and again nearly identical at a large proportion of loci between  
 287 approaches with matched association model and locus weight type (MOR or ROM), regardless of  
 288 bias type (Figs. S13 and S14).

### 289 **3.3 Proof of association invariability to common kinship biases**

290 Our empirical observations suggest that replacing a kinship matrix with either the Standard or  
 291 WG-biased version does not alter association statistics (with exceptions we attribute to numerical  
 292 limited precision artifacts); here prove a more general version of these facts mathematically. Our  
 293 constructive proof shows that only a regression model with relatedness effects as covariates and an  
 294 intercept is required, whose coefficients adapt to the bias, and no other coefficients change. This is  
 295 fortunate, as the intercept and relatedness effect coefficients are nuisance parameters that usually  
 296 go unreported, while the focal genetic association coefficient and its p-value are unchanged by these  
 297 biases.

298 The most general form we identified of the bias function, mapping a kinship matrix to its bias-  
 299 transformed version, and for which association invariability holds, is

$$300 \quad \Phi^{T'} = F(\Phi^T) = \frac{1}{c} \mathbf{B} \Phi^T \mathbf{B}^\top, \quad \mathbf{B} = \mathbf{I} - \mathbf{1}\mathbf{b}^\top, \quad (12)$$

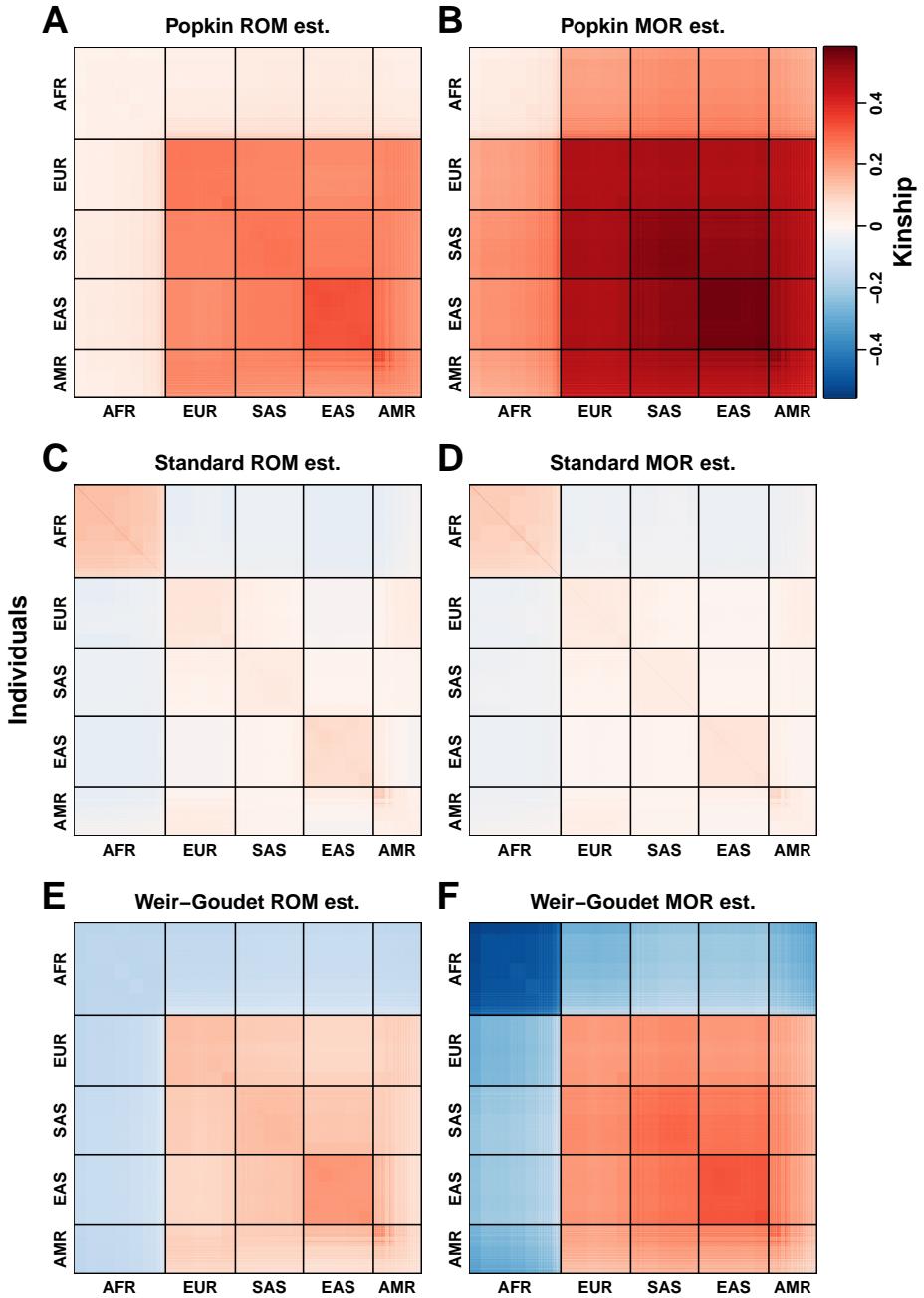


Figure 4: **Kinship estimates on 1000 Genomes.** Each panel represents a kinship matrix as a heatmap, as in Fig. 1. Superpopulation codes: AFR = African, EUR = European, SAS = South Asian, EAS = East Asian, AMR = Admixed Americans (Hispanics). Each estimator bias type (Popkin, Standard, and Weir-Goudet; rows) has two locus weight types (columns): ROM (ratio of means) and MOR (mean of ratios). In this visualization the upper range of all panels is capped to the 99 percentile of the diagonal (population inbreeding values) of the popkin MOR estimates.

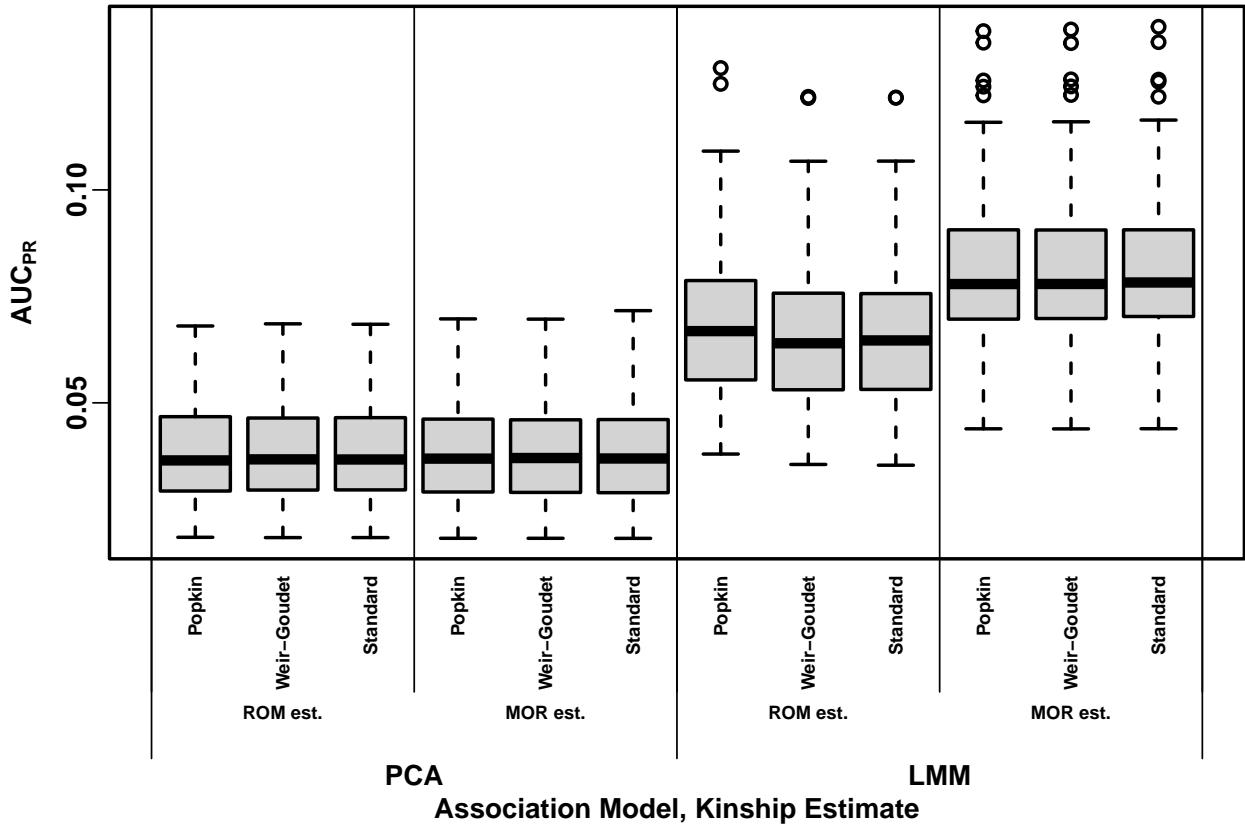


Figure 5: **Distributions of Area Under the Precision-Recall Curve (AUC<sub>PR</sub>) on 1000 Genomes with  $h^2 = 0.8$ .** Higher AUC<sub>PR</sub> is better performance. Results based on 100 simulated trait replicates (real genotype matrix is fixed). Approaches cluster primarily by association model (LMM or PCA) and locus weight type (ROM or MOR), and do not depend much at all on the bias type.

301 where  $c$  is any positive scalar and  $\mathbf{b}$  is any length- $n$  vector. The key property that the linear operator  
 302  $\mathbf{B}$  must satisfy is that it shifts the input vector by the same scalar across its values, or

303 
$$\mathbf{Bs} = \mathbf{s} - \mathbf{1}\eta, \quad (13)$$

304 where  $\mathbf{s}$  is any vector and the scalar  $\eta = \mathbf{b}^\top \mathbf{s}$  is a function of the input vector.  $\mathbf{B}$  in Eq. (12) is the  
 305 only form that results in Eq. (13).

306 The Standard bias function  $F = F^{\text{std}}$  of Eq. (4) can be written as Eq. (12) with  $c = 1 - \bar{\varphi}^T$   
 307 and  $\mathbf{b} = \frac{1}{n}\mathbf{1}$ , in which case  $\mathbf{B}$  equals the centering matrix. Further, the generalized Standard  
 308 estimator studied in Ochoa and Storey (2021) has  $\mathbf{b}$  be a vector of individual weights that sum to  
 309 one:  $\mathbf{b}^\top \mathbf{1} = 1$ . These  $\mathbf{B}$  and  $\Phi^{T'}$  are singular transformations (they are not invertible and have a  
 310 zero eigenvalue), since  $\mathbf{B}\mathbf{1} = \mathbf{0}$  and  $\mathbf{B}^\top \mathbf{b} = \mathbf{0}$ .

311 The WG bias function  $F = F^{\text{WG}}$  of Eq. (6) can be written as Eq. (12) with  $c = 1 - \tilde{\varphi}^T$  and

$$\mathbf{b} = q \frac{(\Phi^T)^{-1} \mathbf{1}}{\mathbf{1}^\top (\Phi^T)^{-1} \mathbf{1}}, \quad (14)$$

$$q = 1 \pm \sqrt{1 - \tilde{\varphi}^T \left( \mathbf{1}^\top (\Phi^T)^{-1} \mathbf{1} \right)}. \quad (15)$$

312 The derivation of this factorization is given in Appendix D. The determinant of the quadratic  
 313 solution  $q$  would be non-negative if  $\tilde{\varphi}^T$  satisfied  $\tilde{\varphi}^T \leq 1 / (\mathbf{1}^\top (\Phi^T)^{-1} \mathbf{1})$ . However, the actual  $\tilde{\varphi}^T$   
 314 does not satisfy this inequality in any of our empirical cases, and in fact  $1 / (\mathbf{1}^\top (\Phi^T)^{-1} \mathbf{1}) \leq \bar{\varphi}^T$   
 315 holds (proven in Appendix E; although  $\tilde{\varphi}^T \leq \bar{\varphi}^T$  (Appendix C), in practice those two are very close  
 316 while  $1 / (\mathbf{1}^\top (\Phi^T)^{-1} \mathbf{1})$  is much smaller than both), so  $b$  above is complex. This is a consequence  
 317 of WG estimates being non-PSD, which we elaborate in the following sections. Nevertheless, PCA  
 318 as well as the GCTA algorithms work for non-PSD matrices without invoking complex numbers  
 319 (following sections and Appendix F).

320 **3.3.1 Proof for LMM case**

321 Consider a random effect  $\mathbf{s}$  drawn using  $\Phi^T$ , as given in Eq. (9). Using the affine transformation  
 322 property of Multivariate Normal distributions (which holds even if  $\mathbf{B}$  below is singular) and Eq. (12),  
 323 then

$$324 \quad \mathbf{s}' = \mathbf{B}\mathbf{s} \sim \text{Normal}(\mathbf{0}, 2\sigma^{2l}\Phi^{T'}) ,$$

$$325 \quad \sigma^{2l} = c\sigma^2. \quad (16)$$

326 (This  $\mathbf{s}'$  has a degenerate distribution for Standard bias, since  $\Phi^{T'}$  is singular, but  $\mathbf{s}' + \boldsymbol{\epsilon}$  is usually  
 327 non-degenerate, since its covariance  $\mathbf{V}' = 2\sigma^{2l}\Phi^{T'} + \sigma_\epsilon^2\mathbf{I}$  is invertible as long as  $\sigma_\epsilon^2 \neq 0$ .) Replacing  
 328  $\mathbf{B}\mathbf{s}$  with the shift form in Eq. (13) shows that  $\mathbf{s}' = \mathbf{s} - \mathbf{1}\eta$  are equal in distribution. Therefore, the  
 329 random effect  $\mathbf{s}'$  of the biased kinship matrix differs from the random effect  $\mathbf{s}$  of the original kinship  
 330 only by  $\mathbf{1}\eta$ , a difference compensated for by adjusting the intercept coefficient in Eq. (8):

$$331 \quad \alpha' = \alpha + \eta. \quad (17)$$

332 No other regression coefficients, or the total residuals, change when  $\Phi^T$  is replaced with  $\Phi^{T'}$ ,  
 333 including the association coefficient  $\beta_i$  that is the focus of the test.

334 The above results require kinship matrices  $\Phi^{T'}$  to be PSD, as covariance matrices are generally  
 335 required to be, and which are characterized by non-negative eigenvalues and determinants. Never-  
 336 theless, for the non-PSD WG bias (has a negative eigenvalue) combined with the generalized least  
 337 squares association algorithm, which is used by GCTA and other LMMs (Kang et al., 2008; Kang  
 338 et al., 2010; Yang et al., 2014), we find a stronger result consistent with Eq. (17), namely that  
 339  $\alpha' = \alpha$ , or in other words,  $\eta = 0$  (Appendix F).

340 The LMM association p-value does not change in several common tests, including the F-test,  
 341 since it only depends on the residuals and these do not change, as well as the likelihood ratio test,  
 342 because although covariance determinants change, they cancel out in the ratio. The Wald test used  
 343 by GCTA (Yang et al., 2014) is also invariant to these kinship biases given our empirical results in  
 344 Figs. 3 and S13 and proven explicitly for WG bias in Appendix F. Lastly, using an implementation in

345 R and simulated data, we confirmed that the LMM Score test is also invariant to these kinship biases.  
 346 These arguments hold whether variance components are fit with maximum likelihood or restricted  
 347 maximum likelihood (Kang et al., 2008; Kang et al., 2010; Yang et al., 2014), since multiplying the  
 348 estimated genetic variance component  $\sigma^2$  by  $c$  and adjusting the intercept compensates for the bias  
 349 regardless of how  $\sigma^2, \sigma_\epsilon^2$  are estimated.

350 **3.3.2 Proof for PCA case**

351 We present a proof for the PCA case that relies on an approximation that holds well in practice.  
 352 Based on the PCA model of Eqs. (10) and (11), let  $\mathbf{U}_r$  be the top eigenvectors of  $\Phi^T$ , and  $\mathbf{U}'_r$  those  
 353 of  $\Phi^{T'}$ . The key approximation is that

$$354 \quad \mathbf{U}'_r \approx \mathbf{B} \mathbf{U}_r, \quad (18)$$

355 which is not strictly equal (since  $\mathbf{B} \mathbf{U}_r$  is not generally orthogonal, as eigenvectors must be), but we  
 356 have found it to be a good approximation in practice. In this case the eigenvector coefficients need  
 357 not change,  $\gamma'_r = \gamma_r$ , since the difference in scale of the kinship matrices ( $c$  in Eq. (12)) is absorbed  
 358 by the eigenvalues not present in this model. Applying the shift of Eq. (13) shows that

$$\mathbf{U}'_r \gamma'_r = \mathbf{B} \mathbf{U}_r \gamma_r = \mathbf{U}_r \gamma_r - \mathbf{1}\eta,$$

359 where  $\eta = \mathbf{b}^\top \mathbf{U}_r \gamma_r$  is a scalar. Therefore, the relatedness effects again differ only by  $\mathbf{1}\eta$ , which is  
 360 compensated for by adjusting the intercept using Eq. (17), so the association coefficient  $\beta_i$  and the  
 361 residuals are the same in both cases. This proof works if there are small numbers of zero or negative  
 362 eigenvalues in  $\Phi^{T'}$  (non-PSD cases), as those rank last and are simply ignored. The observations  
 363 from LMMs, that p-values are invariant to bias types, also hold for PCA.

364 We visualize the top PCs of our datasets in Fig. 6 to assess the validity of Eq. (18). The  
 365 approximation is equivalent to each biased PC (Standard or Weir-Goudet) being shifted from the  
 366 unbiased PC (Popkin), as described in Eq. (13). Fig. 6 indeed shows that PC1 is shifted by  
 367 noticeable amounts in each of these cases, while PC2 is less shifted. However, a rotation of the

PCs is also noticeable, particularly in the simulated data, and other large differences between MOR  
 estimators, as expected since we know the approximation cannot be exact. Also, PCs can change  
 order upon bias transformation, which we notice in the admixed family simulation, where PC2 and  
 PC3 from popkin (and true kinship) actually correspond to PC1 and PC2, respectively, in both  
 Standard and WG, and are plotted as such. No PC reordering occurs in 1000 Genomes. Overall,  
 while the approximation of Eq. (18) can be weakened to merely require that the biased PCs plus  
 intercept span the same subspace of the unbiased PCs plus intercept, the approximate PC shifts  
 better explain intuitively why the result for LMM is also observed for PCA association.

### 3.4 Proof of association invariability to change in ancestral population

The kinship matrices we used so far have values that depend on the choice of ancestral population  
 $T$ . Here we consider the effect on association of changing ancestral population, and prove that it is  
 also compensated for by the relatedness and intercept coefficients.

Start from a kinship matrix  $\Phi^S$  in terms of ancestral population  $S$ , and let  $T$  be a population  
 ancestral to  $S$ . If the inbreeding coefficient of  $S$  when  $T$  is the reference ancestral population is  $f_S^T$ ,  
 then the kinship matrix  $\Phi^T$  in terms of  $T$  is given by (Ochoa and Storey, 2021)

$$(\mathbf{J} - \Phi^T) = (\mathbf{J} - \Phi^S) (1 - f_S^T).$$

Solving for  $\Phi^T$  and simplifying results in

$$\Phi^T = (1 - f_S^T) \Phi^S + f_S^T \mathbf{J}.$$

This resembles WG bias but in reverse: whereas WG reduces and rescales kinship by  $\tilde{\varphi}^T$ , changing  
 to a more ancestral population rescales and increases kinship by  $f_S^T$ . Indeed, excluding  $f_S^T = 1$ , this

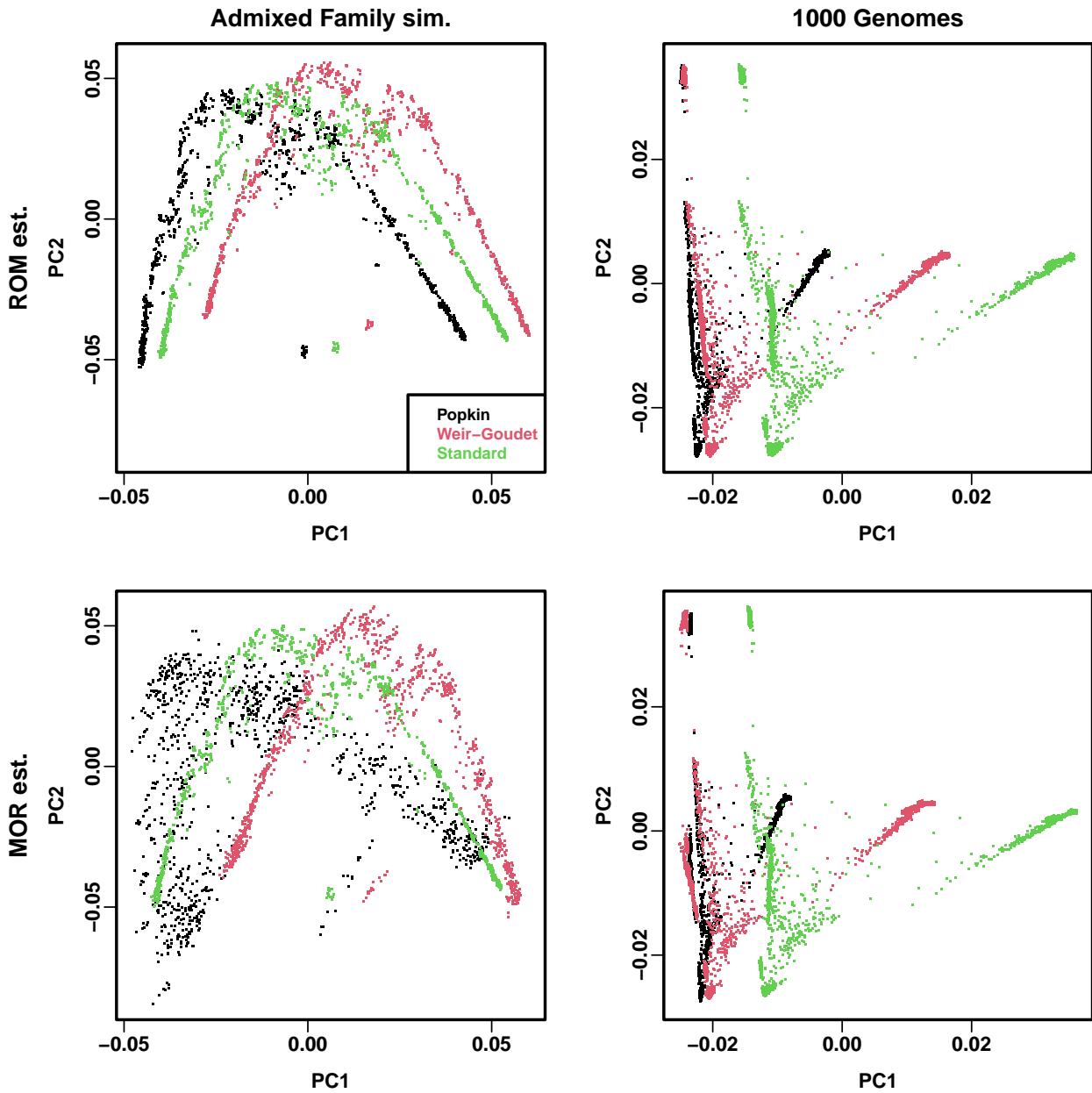


Figure 6: **Visualization of PC shift due to kinship biases.** Each panel shows three estimates (bias types): Popkin, Standard, and Weir-Goudet. ROM estimates are in first row, MOR in second row. (In admixed family, ROM limits are very similar to ROM estimates (not shown).) Columns show estimates from each dataset: admixed family simulation (first replicate) and 1000 Genomes. For popkin (both ROM and MOR estimates) in admixed family only, PC1 and PC2 are replaced with PC2 and PC3 (see text).

386 transformation can be written as Eq. (12) with  $c = (1 - f_S^T)^{-1}$  and

$$\mathbf{b} = q \frac{(\Phi^S)^{-1} \mathbf{1}}{\mathbf{1}^\top (\Phi^S)^{-1} \mathbf{1}},$$
$$q = 1 \pm \sqrt{1 + \frac{f_S^T}{1 - f_S^T} \left( \mathbf{1}^\top (\Phi^S)^{-1} \mathbf{1} \right)}.$$

387 The determinant of  $q$  is strictly positive, since  $\mathbf{1}^\top (\Phi^S)^{-1} \mathbf{1} > 0$  (since  $\Phi^S$  is positive definite, its  
388 inverse is too) and  $0 \leq f_S^T < 1$ . Thus, our previous results apply: ancestor change is compensated  
389 for by the relatedness and intercept coefficients, so the association statistics are invariant to this  
390 transformation.

391 **3.5 Characterization of non-PSD and singular kinship and trait covariance es-**  
392 **timators**

393 While attempting to validate and characterize the earlier factorization of the WG bias function  
394 (Eqs. (12) to (15)), we discovered that it does not produce PSD matrices, which covariance matrices  
395 are required to be. To characterize this problem more broadly, we calculate the eigenvalues of all  
396 kinship matrices  $\Phi^T$  and trait covariance matrices  $\mathbf{V} = 2\sigma^2\Phi^T + \sigma_\epsilon^2\mathbf{I}$ , the latter used by LMMs and  
397 which we calculate using GCTA's estimates of  $\sigma^2$  and  $\sigma_\epsilon^2$ .

398 We find that all WG matrices have very large negative minimum eigenvalues, and popkin MOR  
399 estimates also have smaller negative minimum eigenvalues (Figs. S15 and S16). Moreover, besides all  
400 WG matrices and most popkin MOR estimates, Standard matrices are also often non-PSD but only  
401 in 1000 Genomes (Figs. 7 and S17), which has missing genotypes (the admixed family simulation  
402 does not have missing genotypes). Each of these non-PSD matrices only has one negative eigenvalue.  
403 Notably, all popkin ROM estimates are PSD in every evaluation, including under missingness in  
404 1000 Genomes.

405 In order to quantify matrix singularity, as well as numerical accuracy problems caused by mul-  
406 tiplying by inverses of nearly-singular matrices, we calculate condition numbers, which equal the  
407 maximum absolute eigenvalue divided by the minimum absolute eigenvalue of our covariance ma-  
408 trices. As expected, we see that Standard kinship matrices are singular on our admixed family

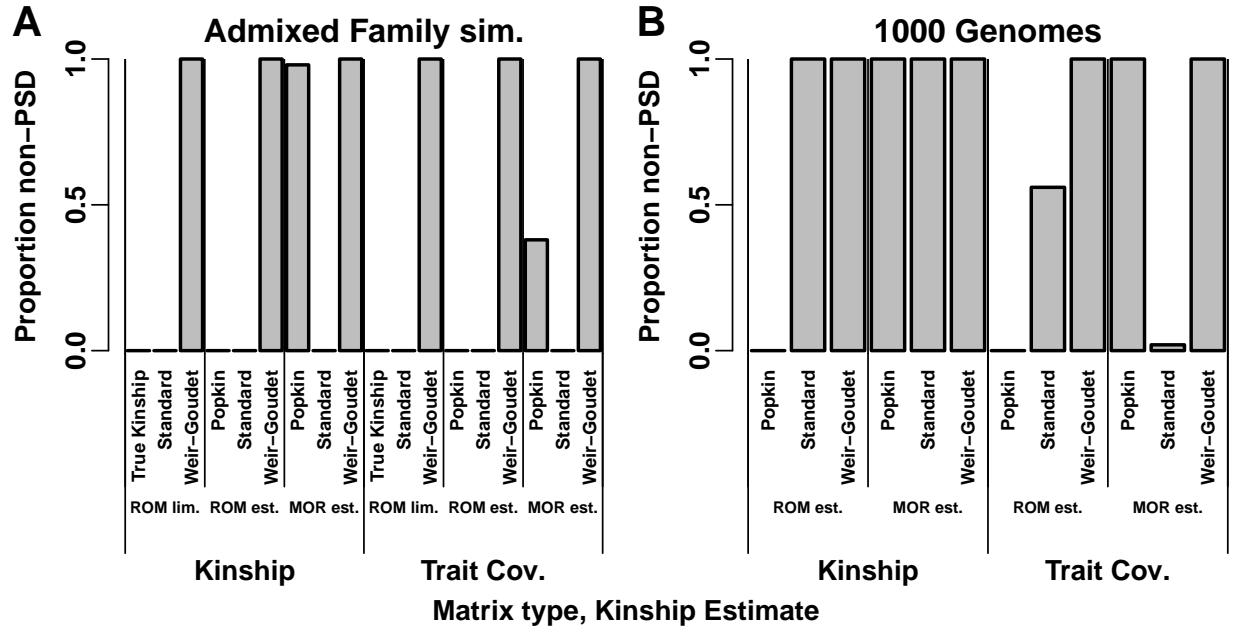


Figure 7: **Proportion of kinship and trait covariance ( $\mathbf{V}$ ) matrices with  $h^2 = 0.8$  that are not positive semidefinite (PSD).** A matrix is non-PSD if it has negative eigenvalues (below  $-10^{-7}$  to allow for limited machine precision). Proportion is calculated over 100 replicates (1000 Genomes kinship has one value since genotypes are fixed, but  $\mathbf{V}$  varies per replicate). **A.** In admixed family simulation, which does not have missing genotypes, all WG matrices and most popkin MOR estimates are non-PSD. All non-PSD kinship matrices results in non-PSD  $\mathbf{V}$  except some popkin ROM estimates yield PSD  $\mathbf{V}$ . **B.** In 1000 Genomes, which has missingness, all kinship estimates are non-PSD except popkin ROM. Of the non-PSD kinship matrices, only some Standard estimates yield PSD  $\mathbf{V}$ .

409 simulation (which lacks missingness), as reflected by extremely high condition numbers, but their  
 410 trait covariances have small condition numbers (Figs. S18 and S19). No other matrices are singular,  
 411 but popkin MOR estimates in the admixed family simulation have relatively high condition numbers  
 412 for both kinship and trait covariance for  $h^2 = 0.8$  but not 0.3.

413 Consider the theoretical connection between the eigenvalues of  $\Phi^T$  and those of  $\mathbf{V}$ . The eigen-  
 414 decomposition trick widely used to fit variance components in LMMs (Kang et al., 2008; Lippert  
 415 et al., 2011; Svishcheva et al., 2012; Zhou and Stephens, 2012; Sul et al., 2018) yields

$$\mathbf{V} = \mathbf{U} (2\sigma^2 \boldsymbol{\Lambda} + \sigma_\epsilon^2 \mathbf{I}) \mathbf{U}^\top,$$

416 where  $\mathbf{U}$  and  $\boldsymbol{\Lambda}$  are the eigenvectors and eigenvalues of  $\Phi^T$ , respectively (Eq. (11)), so the eigen-  
 417 vectors of  $\mathbf{V}$  are also  $\mathbf{U}$  and its eigenvalues are  $2\sigma^2 \boldsymbol{\Lambda} + \sigma_\epsilon^2 \mathbf{I}$ . Therefore, since  $\sigma^2, \sigma_\epsilon^2 \geq 0$ , then if  
 418  $\Phi^T$  is positive definite (all of its eigenvalues are positive) then so is  $\mathbf{V}$ , and the condition number  
 419 of  $\mathbf{V}$  is always smaller (better) or equal than that of  $\Phi^T$ . A negative kinship eigenvalue  $\lambda_k$  may  
 420 become positive for  $\mathbf{V}$  only if  $\lambda_k > -\sigma_\epsilon^2/(2\sigma^2) = -(1-h^2)/(2h^2)$ , so very large negative  $\lambda_k$  values  
 421 as observed for WG do not become positive in  $\mathbf{V}$ , in fact they can become more negative for high  
 422 heritability (Fig. S15), though they are less negative at lower heritability (Fig. S16).  $\mathbf{V}$  is always in-  
 423 vertible and well-conditioned even when  $\Phi^T$  is singular PSD (has zero eigenvalues), as the Standard  
 424 estimator is under no missingness, since a kinship zero eigenvalue becomes  $\sigma_\epsilon^2$  for  $\mathbf{V}$ . Conversely, the  
 425 above equation explains why some non-PSD kinship matrices are particularly problematic: negative  
 426 eigenvectors near the heritability-dependent value  $-\sigma_\epsilon^2/(2\sigma^2)$  can result in ill-conditioned  $\mathbf{V}$ . We  
 427 see that popkin MOR estimates are non-PSD (Fig. S15) in such a way that some of their  $\mathbf{V}$  are  
 428 ill-conditioned under high heritability (Fig. S18) but not the lower heritability (Fig. S19), and this  
 429 explains its poorer performance in the admixed family evaluations (Figs. 2 and S3), as shown in the  
 430 next subsection.

### 431 3.6 Further empirical validation of theoretical predictions

432 Seeing that WG is always non-PSD, and to query other instances where predictions are not fully  
 433 met, here we analyze estimation accuracy of various parameters to better understand theoretically

434 and empirically how broken assumptions affect them. With PCA, no deviations from expectation of  
435  $AUC_{PR}$  and  $SRMSD_p$  are observed for WG (Figs. 2 and 5, Figs. S3 and S9), which makes sense since  
436 PCA simply ignores eigenvectors with negative or zero eigenvalues. Therefore, our analysis focuses  
437 on LMM, where deviations are observed for  $h^2 = 0.8$  but not for 0.3, and clarification regarding  
438 WG is needed.

439 LMMs such as GCTA perform association testing in two steps. First is the restricted maximum  
440 likelihood step used to fit variance components. Although the eigendecomposition approaches (Kang  
441 et al., 2008; Lippert et al., 2011; Svishcheva et al., 2012; Zhou and Stephens, 2012; Sul et al., 2018)  
442 require positive definite  $\mathbf{V}$  (lest the determinant of  $\mathbf{V}$  be negative), surprisingly the GCTA average  
443 information algorithm only requires in practice that  $\mathbf{V}$  be invertible (Yang et al., 2011). Thus,  
444 the relationship between WG, Standard, and True or Popkin variance components are largely as  
445 expected from our theoretical prediction  $\sigma^{2\prime} = c\sigma^2$  in Eq. (16), with the exception of popkin ROM  
446 on 1000 Genomes  $h^2 = 0.8$  only, whose genetic variance estimates are slightly smaller than expected  
447 (Figs. S20 and S21).

448 Next we determine the effect of WG bias on coefficient estimates. In this second step of LMM  
449 association testing, once  $\mathbf{V}$  is determined, GCTA and other LMMs use generalized least squares  
450 to estimate fixed effects coefficients (Kang et al., 2008; Kang et al., 2010; Yang et al., 2014).  
451 Using the first replicate of the admixed family simulation and the true kinship matrix and the  
452 Standard and WG limits only, we recalculate the genetic effect  $\beta_i$  and intercept coefficients  $\alpha$  in R  
453 for all loci, and confirm that we recover the GCTA estimates for  $\beta_i$  to the given precision. We then  
454 compare intercept coefficients, which are not given by GCTA, and confirm our theoretical prediction  
455 (Appendix F) that they are identical whether the True or WG ROM limit kinship matrices are used  
456 (the mean absolute difference is below  $10^{-7}$ ). In contrast, intercepts fit using the Standard ROM  
457 limit kinship matrix are different than those of the true kinship (not shown), which agrees with our  
458 theoretical prediction that the intercept varies to compensate for the kinship matrix bias ( $\alpha' = \alpha + \eta$   
459 in Eq. (17)).

460 Lastly, we explain the largest deviations from our predictions of the performance metrics  $AUC_{PR}$   
461 and  $SRMSD_p$  for  $h^2 = 0.8$  (for  $h^2 = 0.3$  there are no large prediction deviations). We find that the

462 small performance errors of popkin ROM in 1000 Genomes (Figs. 5 and S9) are driven by errors  
463 in genetic variance component estimation  $\sigma^2$  (Fig. S22). However, the larger performance errors  
464 of popkin MOR in the admixed family simulation (Figs. 2 and S3) are instead explained by the  
465 condition number of  $\mathbf{V}$  (Fig. S23). This result makes sense since the condition number by definition  
466 quantifies regression coefficient estimation accuracy.

## 467 4 Discussion

468 Previous research showed that commonly used kinship estimators are biased, and that these biases  
469 can be large (Ochoa and Storey (2021); Fig. 1). Our initial hypothesis was that these kinship biases  
470 would affect association testing, but surprisingly found that association is unaffected. We then  
471 proved theoretically that it is the intercept and relatedness effect (random effect or PCs) coeffi-  
472 cients that compensate for the bias, and result in identical association coefficients and significance  
473 statistics.

474 Kinship estimates depend on the choice of ancestral population, which conditions the distribu-  
475 tions of allele frequencies and genotypes, but the effect of this choice of association testing was not  
476 only unknown but completely disregarded. A corollary of our theoretical results is that changes  
477 of ancestral population, which behave algebraically like kinship bias, are also compensated for by  
478 the relatedness and intercept coefficients, so association testing is also invariant to the choice of  
479 ancestral population. Thus, although a choice of ancestral population is always being made when  
480 estimating kinship, this choice is fortunately inconsequential to association testing, as it ought to  
481 be since relatedness is being conditioned upon in these tests.

482 Given that kinship bias type is not important for association studies, we are free to choose a  
483 kinship estimator based on other properties. Ideally, kinship matrices result in well conditioned trait  
484 covariance matrices, since that has the largest effect in numerical accuracy and power in LMMs.  
485 Well-conditioned association is guaranteed for PSD kinship matrices, and popkin ROM is the only  
486 estimator that produces PSD matrices consistently across our evaluations (Fig. 7). Popkin ROM  
487 is also the only unbiased kinship estimator (Ochoa and Storey, 2021). We observed that Standard  
488 kinship estimates are also not PSD when genotypes are missing, a well understood phenomenon

489 for related sample covariance estimators outside genetics (Jurczak and Rohde, 2017). Fortunately,  
490 non-PSD kinship estimators often perform well for association. Nevertheless, in our admixed family  
491 simulation we did see the other popkin estimator (the MOR version) perform particularly poorly due  
492 to being non-PSD, which in a heritability-dependent manner results in ill-conditioned association  
493 tests and substantial loss of accuracy and power (Figs. 2, S3 and S23). Theory predicts that the same  
494 can happen with any non-PSD estimator, depending on unknowns such as the heritability and the  
495 value of the negative eigenvalues of the kinship estimator, so it is risky to use MOR estimators (all  
496 of which are non-PSD in 1000 Genomes), as well as the WG estimator generally (which is non-PSD  
497 in all replicates of all of our evaluations). We also observe smaller numerical inaccuracies for popkin  
498 ROM, the estimator we recommend, in 1000 Genomes with  $h^2 = 0.8$  only, although the result is  
499 mixed: performance is slightly better (Fig. 5) although null p-value calibration is slightly worse  
500 (Fig. S9). The cause is variance components are poorly estimated (Fig. S20), but we did not find a  
501 more fundamental explanation. Overall, our assessment suggests that the popkin ROM estimator is  
502 the safest choice due to its guarantee of well-conditioned associations that other estimators cannot  
503 make.

504 Despite being non-PSD, we observe better performance for MOR versus ROM estimators in  
505 LMM association of 1000 Genomes with  $h^2 = 0.8$  (Fig. 5), but not under lower heritability (Fig. S8).  
506 Perhaps this is expected because we simulated larger coefficients for rare variants, while MOR  
507 estimators upweigh rare variants. This effect is not observed in the admixed family simulation,  
508 where MOR and ROM versions give similar kinship estimates (Fig. 1) and performed similarly  
509 (Fig. 2), compared to 1000 Genomes where kinship estimates are also strikingly different (Fig. 4).  
510 However, only popkin ROM is unbiased (Fig. 1B, Fig. S1). One potential explanation is that our  
511 kinship model assumes that all variants existed in the MRCA population, whereas rare variants  
512 in human data are known to be more recent mutations, and thus their effective kinship matrix is  
513 different than that of ancestral variants. Therefore, despite its biases, the popkin MOR estimator  
514 may better capture the covariance of rare variants and thus model them better in association tests,  
515 particularly in LMMs where the effect is most pronounced. Future work should focus on better  
516 approaches for upweighing rare variants or otherwise estimating their covariance structure while

517 resulting in positive definite kinship estimates.

518 Our conclusions that common kinship biases do not affect association studies extend to variations  
519 of the Standard kinship estimator that weigh loci according to linkage disequilibrium (Speed et al.,  
520 2017; Wang et al., 2017), which also have the Standard bias type since this bias is present in each  
521 locus (Ochoa and Storey, 2021). As shown in our theoretical results, another form of the Standard  
522 kinship estimator that weighs individuals to estimate ancestral allele frequencies  $\hat{p}_i^T$ , including the  
523 best unbiased linear estimator in Appendix E (Astle and Balding, 2009; Thornton and McPeek,  
524 2010), is also subject to the same conclusions.

525 In this study, we show empirically and theoretically that association tests are invariant to the  
526 use of common kinship estimators that are biased versus a more recent unbiased estimator. Since  
527 the results hold in the presence of additional covariates, they hold for multivariate tests in general,  
528 which encompasses LASSO approaches and rare variant (burden, kernel, etc) tests that include  
529 PCs or random effects from a kinship matrix. The underpinnings of our proof show that the  
530 same result holds for association with generalized linear models, since the intercept and relatedness  
531 effects interact in the same way as for linear models (the link function goes around the trait only);  
532 these models include case/control models such as logistic PCA and LMM. However, heritability  
533 estimation requires unbiased estimates of the random effect coefficient ( $\sigma^2$ ), so it is biased when  
534 the standard kinship estimator is used, as it is using GCTA (Yang et al., 2011; Yang et al., 2014).  
535 Nevertheless, heritability estimation is a complex problem and a complete analysis is beyond the  
536 scope of this work. Overall, we have described an unexpected robustness of association studies, and  
537 our theoretical understanding of this result may help guide future improvements for association and  
538 other related models.

## 539 Declaration of interests

540 The authors declare no competing interests.

541 **Acknowledgments**

542 This work was funded in part by the Duke University School of Medicine Whitehead Scholars  
543 Program, a gift from the Whitehead Charitable Foundation. The 1000 Genomes data were generated  
544 at the New York Genome Center with funds provided by NHGRI Grant 3UM1HG008901-03S1.

545 **Web resources**

546 plink2, <https://www.cog-genomics.org/plink/2.0/>  
547 GCTA, <https://yanglab.westlake.edu.cn/software/gcta/>  
548 bnpsd, <https://cran.r-project.org/package=bnpsd>  
549 simfam, <https://cran.r-project.org/package=simfam>  
550 simtrait, <https://cran.r-project.org/package=simtrait>  
551 popkin, <https://cran.r-project.org/package=popkin>  
552 popkinsuppl, <https://github.com/OchoaLab/popkinsuppl>

553 **Data and code availability**

554 The data and code generated during this study are available on GitHub at <https://github.com/>  
555 [OchoaLab/bias-assoc-paper](https://github.com/OchoaLab/bias-assoc-paper). The high-coverage version of the 1000 Genomes Project was down-  
556 loaded from [ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data\\_collections/1000G\\_2504\\_high\\_](ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000G_2504_high_)  
557 [coverage/working/20190425\\_NYGC\\_GATK/](coverage/working/20190425_NYGC_GATK/) and detailed processing instructions are found on GitHub  
558 at <https://github.com/OchoaLab/data/>.

559 **References**

560 1000 Genomes Project Consortium et al. (2012). “An integrated map of genetic variation from 1,092  
561 human genomes”. *Nature* 491(7422), pp. 56–65. DOI: 10.1038/nature11632.

- 562 Altschul, Stephen F., Raymond J. Carroll, and David J. Lipman (1989). “Weights for data related by  
563 a tree”. *Journal of Molecular Biology* 207(4), pp. 647–653. DOI: 10.1016/0022-2836(89)90234-  
564 9.
- 565 Astle, William and David J. Balding (2009). “Population Structure and Cryptic Relatedness in  
566 Genetic Association Studies”. *Statist. Sci.* 24(4), pp. 451–471. DOI: 10.1214/09-STS307.
- 567 Aulchenko, Yurii S., Dirk-Jan de Koning, and Chris Haley (2007). “Genomewide rapid associa-  
568 tion using mixed model and regression: a fast and simple method for genomewide pedigree-  
569 based quantitative trait loci association analysis”. *Genetics* 177(1), pp. 577–585. DOI: 10.1534/  
570 *genetics*.107.075614.
- 571 Balding, D. J. and R. A. Nichols (1995). “A method for quantifying differentiation between popula-  
572 tions at multi-allelic loci and its implications for investigating identity and paternity”. *Genetica*  
573 96(1-2), pp. 3–12. DOI: <https://doi.org/10.1007/BF01441146>.
- 574 Bhatia, Gaurav et al. (2013). “Estimating and interpreting FST: the impact of rare variants”.  
575 *Genome Res.* 23(9), pp. 1514–1521. DOI: 10.1101/gr.154831.113.
- 576 Chang, Christopher C. et al. (2015). “Second-generation PLINK: rising to the challenge of larger  
577 and richer datasets”. *GigaScience* 4(1), p. 7. DOI: 10.1186/s13742-015-0047-8.
- 578 Consortium, The 1000 Genomes Project (2010). “A map of human genome variation from population-  
579 scale sequencing”. *Nature* 467(7319), pp. 1061–1073. DOI: 10.1038/nature09534.
- 580 Devlin, B. and Kathryn Roeder (1999). “Genomic Control for Association Studies”. *Biometrics*  
581 55(4), pp. 997–1004. DOI: 10.1111/j.0006-341X.1999.00997.x.
- 582 Emik, L. Otis and Clair E. Terrill (1949). “Systematic procedures for calculating inbreeding coeffi-  
583 cients”. *J Hered* 40(2), pp. 51–55. DOI: 10.1093/oxfordjournals.jhered.a105986.
- 584 Fairley, Susan et al. (2020). “The International Genome Sample Resource (IGSR) collection of  
585 open human genomic variation resources”. *Nucleic Acids Research* 48(D1), pp. D941–D947. DOI:  
586 10.1093/nar/gkz836.
- 587 García-Cortés, Luis Alberto (2015). “A novel recursive algorithm for the calculation of the detailed  
588 identity coefficients”. *Genetics Selection Evolution* 47(1), p. 33. DOI: 10.1186/s12711-015-  
589 0108-6.

- 590 Hoffman, Gabriel E. (2013). "Correcting for population structure and kinship using the linear mixed  
591 model: theory and extensions". *PLoS ONE* 8(10), e75707. DOI: 10.1371/journal.pone.0075707.
- 592 Jacquard, Albert (1970). *Structures génétiques des populations*. Paris: Masson et Cie.
- 593 Jurczak, Kamil and Angelika Rohde (2017). "Spectral analysis of high-dimensional sample covariance  
594 matrices with missing observations". *Bernoulli* 23(4A), pp. 2466–2532. DOI: 10.3150/16-BEJ815.
- 595 Kang, Hyun Min et al. (2008). "Efficient control of population structure in model organism associ-  
596 ation mapping". *Genetics* 178(3), pp. 1709–1723. DOI: 10.1534/genetics.107.080101.
- 597 Kang, Hyun Min et al. (2010). "Variance component model to account for sample structure in  
598 genome-wide association studies". *Nat. Genet.* 42(4), pp. 348–354. DOI: 10.1038/ng.548.
- 599 Lippert, Christoph et al. (2011). "FaST linear mixed models for genome-wide association studies".  
600 *Nat. Methods* 8(10), pp. 833–835. DOI: 10.1038/nmeth.1681.
- 601 Loh, Po-Ru et al. (2015). "Efficient Bayesian mixed-model analysis increases association power in  
602 large cohorts". *Nat. Genet.* 47(3), pp. 284–290. DOI: 10.1038/ng.3190.
- 603 Malécot, Gustave (1948). *Mathématiques de l'hérédité*. Masson et Cie.
- 604 Ochoa, Alejandro and John D. Storey (2021). "Estimating FST and kinship for arbitrary population  
605 structures". *PLoS Genet* 17(1), e1009241. DOI: 10.1371/journal.pgen.1009241.
- 606 Price, Alkes L. et al. (2006). "Principal components analysis corrects for stratification in genome-  
607 wide association studies". *Nat. Genet.* 38(8), pp. 904–909. DOI: 10.1038/ng1847.
- 608 Rakovski, Cyril S. and Daniel O. Stram (2009). "A kinship-based modification of the armitage trend  
609 test to address hidden population structure and small differential genotyping errors". *PLoS ONE*  
610 4(6), e5825. DOI: 10.1371/journal.pone.0005825.
- 611 Sherman, Jack and Winifred J. Morrison (1950). "Adjustment of an Inverse Matrix Corresponding  
612 to a Change in One Element of a Given Matrix". *The Annals of Mathematical Statistics* 21(1),  
613 pp. 124–127. DOI: 10.1214/aoms/1177729893.
- 614 Speed, Doug and David J. Balding (2015). "Relatedness in the post-genomic era: is it still useful?"  
615 *Nat. Rev. Genet.* 16(1), pp. 33–44. DOI: 10.1038/nrg3821.
- 616 Speed, Doug et al. (2012). "Improved heritability estimation from genome-wide SNPs". *Am. J. Hum.  
617 Genet.* 91(6), pp. 1011–1021. DOI: 10.1016/j.ajhg.2012.10.010.

- 618 Speed, Doug et al. (2017). "Reevaluation of SNP heritability in complex human traits". *Nat Genet*  
619 49(7), pp. 986–992. DOI: 10.1038/ng.3865.
- 620 Sul, Jae Hoon, Lana S. Martin, and Eleazar Eskin (2018). "Population structure in genetic studies:  
621 Confounding factors and mixed models". *PLoS Genet.* 14(12), e1007309. DOI: 10.1371/journal.  
622 pgen.1007309.
- 623 Svishcheva, Gulnara R. et al. (2012). "Rapid variance components-based method for whole-genome  
624 association analysis". *Nat Genet* 44(10), pp. 1166–1170. DOI: 10.1038/ng.2410.
- 625 Thornton, Timothy and Mary Sara McPeek (2010). "ROADTRIPS: case-control association testing  
626 with partially or completely unknown population and pedigree structure". *Am. J. Hum. Genet.*  
627 86(2), pp. 172–184. DOI: 10.1016/j.ajhg.2010.01.001.
- 628 Voight, Benjamin F. and Jonathan K. Pritchard (2005). "Confounding from Cryptic Relatedness  
629 in Case-Control Association Studies". *PLOS Genetics* 1(3), e32. DOI: 10.1371/journal.pgen.  
630 0010032.
- 631 Wang, Bowen, Serge Sverdlov, and Elizabeth Thompson (2017). "Efficient Estimation of Realized  
632 Kinship from SNP Genotypes". *Genetics, genetics*.116.197004. DOI: 10.1534/genetics.116.  
633 197004.
- 634 Weir, Bruce S. and Jérôme Goudet (2017). "A Unified Characterization of Population Structure and  
635 Relatedness". *Genetics* 206(4), pp. 2085–2103. DOI: 10.1534/genetics.116.198424.
- 636 Wright, Sewall (1922). "Coefficients of Inbreeding and Relationship". *The American Naturalist*  
637 56(645), pp. 330–338.
- 638 — (1949). "The Genetical Structure of Populations". *Annals of Eugenics* 15(1), pp. 323–354. DOI:  
639 10.1111/j.1469-1809.1949.tb02451.x.
- 640 Xie, C., D. D. Gessler, and S. Xu (1998). "Combining different line crosses for mapping quantitative  
641 trait loci using the identical by descent-based variance component method". *Genetics* 149(2),  
642 pp. 1139–1146.
- 643 Yang, Jian et al. (2010). "Common SNPs explain a large proportion of the heritability for human  
644 height". *Nat. Genet.* 42(7), pp. 565–569. DOI: 10.1038/ng.608.

- 645 Yang, Jian et al. (2011). “GCTA: a tool for genome-wide complex trait analysis”. *Am. J. Hum. Genet.* 88(1), pp. 76–82. DOI: 10.1016/j.ajhg.2010.11.011.
- 647 Yang, Jian et al. (2014). “Advantages and pitfalls in the application of mixed-model association methods”. *Nat Genet* 46(2), pp. 100–106. DOI: 10.1038/ng.2876.
- 649 Yao, Yiqi and Alejandro Ochoa (2022). *Limitations of principal components in quantitative genetic association models for human studies*. Tech. rep. bioRxiv, p. 2022.03.25.485885. DOI: 10.1101/2022.03.25.485885.
- 652 Yu, Jianming et al. (2006). “A unified mixed-model method for association mapping that accounts for multiple levels of relatedness”. *Nat. Genet.* 38(2), pp. 203–208. DOI: 10.1038/ng1702.
- 654 Zhou, Xiang and Matthew Stephens (2012). “Genome-wide efficient mixed-model analysis for association studies”. *Nat. Genet.* 44(7), pp. 821–824. DOI: 10.1038/ng.2310.

## 656 Appendices

### 657 A Justification for popkin generalizations

658 The popkin estimator in Eq. (1) has been generalized in this work to include locus weights  $w_i$ . The  
 659 original ROM formulation had  $w_i = 1$  for all loci  $i$  (Ochoa and Storey, 2021). Recalling from that  
 660 original work that

$$\mathbb{E} [(x_{ij} - 1)(x_{ik} - 1) - 1 | T] = 4p_i^T (1 - p_i^T) (\varphi_{jk}^T - 1),$$

661 then for fixed  $w_i$  we get

$$\begin{aligned} \mathbb{E} [A_{jk} | T] &= v_m^T (\varphi_{jk}^T - 1), \\ v_m^T &= \frac{4}{m} \sum_{i=1}^m w_i p_i^T (1 - p_i^T). \end{aligned}$$

662 Therefore, as before all the unknowns  $p_i^T$  and now also the known weights  $w_i$  collapse into a single  
 663 parameter  $v_m^T$ , which is estimated under the assumption that the minimum kinship is zero, giving  
 664  $\hat{A}_{\min} = -v_m^T$ , so that

$$\hat{\varphi}_{jk}^{T,\text{popkin-ROM}} = 1 - \frac{A_{jk}}{\hat{A}_{\min}} \xrightarrow[m \rightarrow \infty]{\text{a.s.}} \varphi_{jk}^T$$

665 as desired.

666 The MOR case of  $w_i = (\hat{p}_i^T (1 - \hat{p}_i^T))^{-1}$  does not fit the previous case because this  $w_i$  is a  
 667 random variable (it is a function of the genotypes). The term of interest  $w_i((x_{ij} - 1)(x_{ik} - 1) - 1)$   
 668 is a ratio of random variables whose expectation does not have a closed form. In this case, we rely  
 669 on the first-order approximation to this expectation, namely

$$\begin{aligned} \mathbb{E} \left[ \frac{(x_{ij} - 1)(x_{ik} - 1) - 1}{\hat{p}_i^T (1 - \hat{p}_i^T)} \middle| T \right] &\approx \frac{\mathbb{E}[(x_{ij} - 1)(x_{ik} - 1) - 1 | T]}{\mathbb{E}[\hat{p}_i^T (1 - \hat{p}_i^T) | T]} \\ &= \frac{4p_i^T (1 - p_i^T) (\varphi_{jk}^T - 1)}{p_i^T (1 - p_i^T) (1 - \bar{\varphi}^T)} \\ &= \frac{4(\varphi_{jk}^T - 1)}{1 - \bar{\varphi}^T}, \end{aligned}$$

670 where the expectation of  $\hat{p}_i^T (1 - \hat{p}_i^T)$  was calculated previously (Ochoa and Storey, 2021). In this  
 671 case the expectation of  $A_{jk}$ , summing across loci, is also approximated by

$$\mathbb{E}[A_{jk}|T] \approx \frac{4(\varphi_{jk}^T - 1)}{1 - \bar{\varphi}^T}.$$

672 The same strategy as before applies to estimate the unknown factor  $4/(1 - \bar{\varphi}^T)$ , namely that if the  
 673 minimum kinship is zero then  $\hat{A}_{\min} \approx -4/(1 - \bar{\varphi}^T)$ , resulting in

$$\hat{\varphi}_{jk}^{T,\text{popkin-MOR}} = 1 - \frac{A_{jk}}{\hat{A}_{\min}} \approx \varphi_{jk}^T.$$

## 674 B Connection between popkin and standard kinship estimator

675 Since the connection we discovered holds when data are complete, but not under missingness, to  
 676 determine necessary conditions we introduce more complete forms of the estimators that handle  
 677 missingness. Popkin (with locus weights) has the following parts updated:

$$A_{ijk} = I_{ij}I_{ik}((x_{ij} - 1)(x_{ik} - 1) - 1),$$

$$A_{jk} = \frac{1}{m_{jk}} \sum_{i=1}^m w_i A_{ijk},$$

$$m_{jk} = \sum_{i=1}^m I_{ij}I_{ik},$$

678 where  $I_{ij} = 1$  if  $x_{ij}$  is not missing, 0 otherwise (this way missing  $x_{ij}$  can have any value and not  
 679 contribute to the estimator). Only loci with both genotypes ( $x_{ij}$  and  $x_{ik}$ ) non-missing are included  
 680 in the above average, and  $m_{jk}$  counts the total number of such loci. The ancestral allele frequency  
 681 estimator with missingness is

$$\hat{p}_i^T = \frac{1}{2n_i} \sum_{j=1}^n I_{ij}x_{ij},$$

$$n_i = \sum_{j=1}^n I_{ij},$$

682 which averages over individuals rather than loci, so its denominator is the number of non-missing  
 683 individuals at this locus. Let us compute some averages of the popkin estimator. Since the result  
 684 we want holds at every locus separately, let us formulate the averages of interest at locus  $i$  only:

$$\bar{A}_{ij} = \frac{1}{n} \sum_{k=1}^n A_{ijk} = I_{ij} \frac{n_i}{n} ((x_{ij} - 1)(2\hat{p}_i^T - 1) - 1),$$

$$\bar{A}_i = \frac{1}{n} \sum_{k=1}^n \bar{A}_{ij} = - \left( \frac{n_i}{n} \right)^2 4\hat{p}_i^T (1 - \hat{p}_i^T).$$

685 Therefore, the combination of interest is:

$$\begin{aligned}
A_{ijk} + \bar{A}_i - \bar{A}_{ij} - \bar{A}_{ik} &= I_{ij}I_{ik} (x_{ij} - 2\hat{p}_i^T) (x_{ik} - 2\hat{p}_i^T) \\
&\quad + \frac{n_i}{n} \left( I_{ij} - \frac{n_i}{n} \right) 4\hat{p}_i^T + \left( \left( \frac{n_i}{n} \right)^2 - I_{ij}I_{ik} \right) 4(\hat{p}_i^T)^2 \\
&\quad + I_{ij} \left( I_{ik} - \frac{n_i}{n} \right) x_{ij} (2\hat{p}_i^T - 1) + I_{ik} \left( I_{ij} - \frac{n_i}{n} \right) x_{ik} (2\hat{p}_i^T - 1).
\end{aligned}$$

686 For the above to equal  $I_{ij}I_{ik} (x_{ij} - 2\hat{p}_i^T) (x_{ik} - 2\hat{p}_i^T)$ , which is the first term above, the rest of the  
687 terms must vanish for arbitrary values of  $\hat{p}_i^T$ ,  $x_{ij}$ , and  $x_{ik}$ . Since  $n_i > 0$  (there is at least one  
688 non-missing individual at every locus), the term  $\frac{n_i}{n}(I_{ij} - \frac{n_i}{n})4\hat{p}_i^T$  vanishes if and only if  $I_{ij} = \frac{n_i}{n}$ , and  
689 since  $I_{jk} = 0$  does not solve this equation (because  $n_i > 0$ ), then  $I_{jk} = 1$ , which requires  $n_i = n$ ,  
690 so no individuals can have missing data at this locus (the rest of the terms vanish when this is so).

691 Thus,

$$A_{ijk} + \bar{A}_i - \bar{A}_{ij} - \bar{A}_{ik} = I_{ij}I_{ik} (x_{ij} - 2\hat{p}_i^T) (x_{ik} - 2\hat{p}_i^T)$$

692 if and only if there is no missing data at locus  $i$ . The other desired result of

$$\bar{A}_i = -4\hat{p}_i^T (1 - \hat{p}_i^T)$$

693 also requires  $n_i = n$ .

694 Assuming now no missingness, transforming the popkin estimates using the Standard bias func-  
695 tion of Eq. (4) gives

$$\begin{aligned}
\frac{\hat{\varphi}_{jk}^{T,\text{popkin}} + \bar{\varphi}^{T,\text{popkin}} - \bar{\varphi}_j^{T,\text{popkin}} - \bar{\varphi}_k^{T,\text{popkin}}}{1 - \bar{\varphi}^{T,\text{popkin}}} &= \frac{A_{jk} + \bar{A} - \bar{A}_j - \bar{A}_k}{-\bar{A}} \\
&= \frac{\sum_{i=1}^m w_i (A_{ijk} + \bar{A}_i - \bar{A}_{ij} - \bar{A}_{ik})}{-\sum_{i=1}^m w_i \bar{A}_i} \\
&= \frac{\sum_{i=1}^m w_i (x_{ij} - 2\hat{p}_i^T) (x_{ik} - 2\hat{p}_i^T)}{\sum_{i=1}^m w_i 4\hat{p}_i^T (1 - \hat{p}_i^T)}.
\end{aligned}$$

696 Therefore, if popkin ROM is input ( $w_i = 1$ ), this transformation yields Standard ROM. On the  
697 other hand, if popkin MOR is used ( $w_i^{-1} = \hat{p}_i^T (1 - \hat{p}_i^T)$ ), the transformation yields Standard MOR.

698 **C Mean kinship inequalities**

699 Denote the mean of the diagonal kinship terms as  $\bar{\delta}^T = \frac{1}{n} \sum_{j=1}^n \varphi_{jj}^T$ . Here we prove that

$$0 \leq \tilde{\varphi}^T \leq \bar{\varphi}^T \leq \bar{\delta}^T \leq 1,$$

700 with each of  $\tilde{\varphi}^T = \bar{\varphi}^T$  and  $\bar{\varphi}^T = \bar{\delta}^T$  if and only if all kinship values are equal.

701 The inequalities  $0 \leq \tilde{\varphi}^T \leq \bar{\varphi}^T \leq \bar{\delta}^T \leq 1$  follow directly from previous work, applied to a kinship  
702 matrix rather than a coancestry matrix as done originally, as the proof required solely a covariance  
703 matrix with values between 0 and 1 (Ochoa and Storey, 2021).  $\tilde{\varphi}^T$  is defined in Eq. (7).  $0 \leq \tilde{\varphi}^T$   
704 follows since every kinship value is non-negative.  $\bar{\varphi}^T$  and  $\tilde{\varphi}^T$  are related by

$$705 \quad \bar{\varphi}^T = \frac{\tilde{\varphi}^T(n-1) + \bar{\delta}^T}{n}. \quad (19)$$

706 Applying  $\bar{\varphi}^T \leq \bar{\delta}^T$  to Eq. (19) and simplifying yields  $\tilde{\varphi}^T \leq \bar{\delta}^T$ . Lastly, since  $\bar{\varphi}^T - \tilde{\varphi}^T = (\bar{\delta}^T - \tilde{\varphi}^T)/n$   
707 (from rearranging Eq. (19)), it also follows that  $\tilde{\varphi}^T \leq \bar{\varphi}^T$ , as desired. Furthermore,  $\tilde{\varphi}^T = \bar{\varphi}^T$  holds  
708 if and only if all  $\varphi_{jk}^T = \bar{\delta}^T$ , since that is necessary and sufficient for  $\bar{\varphi}^T = \bar{\delta}^T$ .

709 **D Derivation of WG bias factorization**

710 Here we rewrite the WG bias function of Eq. (6) as a factorization of the form of Eq. (12). It is  
711 easy to see that  $c = 1 - \tilde{\varphi}^T$ . Expanding Eq. (12) gives

$$\begin{aligned} \mathbf{B}\Phi^T\mathbf{B}^\top &= (\mathbf{I} - \mathbf{1}\mathbf{b}^\top)\Phi^T(\mathbf{I} - \mathbf{b}\mathbf{1}^\top) \\ &= \Phi^T - \mathbf{1}(\Phi^T\mathbf{b})^\top - (\Phi^T\mathbf{b})\mathbf{1}^\top + \mathbf{J}(\mathbf{b}^\top\Phi^T\mathbf{b}), \end{aligned}$$

712 where  $\mathbf{b}^\top\Phi^T\mathbf{b}$  is a scalar and  $\Phi^T\mathbf{b}$  a vector. Equating the above to Eq. (6) and rearranging, we  
713 obtain

$$\mathbf{J}(\tilde{\varphi}^T + (\mathbf{b}^\top\Phi^T\mathbf{b})) = \mathbf{1}(\Phi^T\mathbf{b})^\top + (\Phi^T\mathbf{b})\mathbf{1}^\top.$$

714 Since  $\tilde{\varphi}^T + (\mathbf{b}^\top \Phi^T \mathbf{b})$  is a scalar and  $\mathbf{J} = \mathbf{1}\mathbf{1}^\top$ , we can see that the solution requires the right side to  
 715 also be a constant matrix, which is only achieved if  $\Phi^T \mathbf{b} \propto \mathbf{1}$ . We choose the scaling factor for the  
 716 last  $\mathbf{1}$  to be  $q \left( \mathbf{1}^\top (\Phi^T)^{-1} \mathbf{1} \right)^{-1}$  as this simplifies notation later, and solving for  $\mathbf{b}$  results in Eq. (14).  
 717 To solve for  $q$ , we replace  $\mathbf{b}$  from Eq. (14) into the above equation, which after rearranging results  
 718 in

$$q^2 - 2q + \tilde{\varphi}^T \left( \mathbf{1}^\top (\Phi^T)^{-1} \mathbf{1} \right) = 0.$$

719 The solution to the above quadratic equation is given by Eq. (15), as desired.

## 720 E Minimum weighted mean kinship

721 Consider the weighted mean kinship value  $\mathbf{w}^\top \Phi^T \mathbf{w}$ , where  $\mathbf{w}$  are weights that sum to one ( $\mathbf{w}^\top \mathbf{1} = 1$ ).  
 722 The ordinary mean kinship  $\bar{\varphi}^T$  is the special case with  $\mathbf{w} = \frac{1}{n} \mathbf{1}$ . The weights that minimize the  
 723 weighted mean kinship are the solution of the Lagrangian multiplier problem

$$G = \mathbf{w}^\top \Phi^T \mathbf{w} + \lambda(\mathbf{w}^\top \mathbf{1} - 1).$$

724 The derivatives are the constraint and  $\frac{dG}{d\mathbf{w}} = 2\Phi^T \mathbf{w} + \lambda \mathbf{1} = \mathbf{0}$ . The optimal weights thus satisfy  
 725  $\mathbf{w} = \frac{-\lambda}{2} (\Phi^T)^{-1} \mathbf{1}$ . Multiplying by  $\mathbf{1}^\top$ , since  $\mathbf{1}^\top \mathbf{w} = 1$ , allows us to solve for  $\lambda^{-1} = -\frac{1}{2} \mathbf{1}^\top (\Phi^T)^{-1} \mathbf{1}$ .  
 726 Thus, the optimal weights are

$$\mathbf{w} = \frac{(\Phi^T)^{-1} \mathbf{1}}{\mathbf{1}^\top (\Phi^T)^{-1} \mathbf{1}},$$

727 a solution that recurs in related settings, and applied to genotypes as  $\hat{p}_i^T = \mathbf{w}^\top \mathbf{x}_i / 2$  yields the  
 728 best linear unbiased estimator of  $p_i^T$  (Altschul et al., 1989; Astle and Balding, 2009; Thornton and  
 729 McPeek, 2010). Therefore, the minimum weighted mean kinship is, and satisfies,

$$\mathbf{w}^\top \Phi^T \mathbf{w} = \frac{1}{\mathbf{1}^\top (\Phi^T)^{-1} \mathbf{1}} \leq \bar{\varphi}^T \approx \tilde{\varphi}^T.$$

730 **F Proof that WG bias results in zero intercept shift under LMM**  
 731 **generalized least squares estimation**

732 For this section suppose that variance components have been estimated, so  $\mathbf{V} = 2\sigma^2 \Phi^T + \sigma_\epsilon^2 \mathbf{I}$  is  
 733 given, assume it is invertible, and rewrite the LMM as

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\epsilon}_V, \quad \boldsymbol{\epsilon}_V \sim \text{Normal}(\mathbf{0}, \mathbf{V}),$$

734 where the design matrix  $\mathbf{Z} = (\mathbf{1}, \mathbf{x}_i, \dots)$  contains the intercept, genotype and now additional covari-  
 735 ates, and  $\boldsymbol{\beta} = (\alpha, \beta_i, \dots)$  are their coefficients. The generalized least squares coefficients estimate,  
 736 used by GCTA and other LMMs, is

$$\hat{\boldsymbol{\beta}} = (\mathbf{Z}^\top \mathbf{V}^{-1} \mathbf{Z})^{-1} \mathbf{Z}^\top \mathbf{V}^{-1} \mathbf{y}.$$

737 Now suppose  $\mathbf{V}$  corresponds to some kinship matrix  $\Phi^T$  while  $\mathbf{V}'$  corresponds to  $\Phi^{T'} = F^{\text{WG}}(\Phi^T)$ ,  
 738 and  $\mathbf{V}'$  is also invertible. Our strategy involves repeated application of the Sherman-Morrison for-  
 739 mula for calculating inverses of matrices after a rank-1 update, which for a symmetric update of a  
 740 matrix  $\mathbf{A}$  with a vector  $\mathbf{z}$  and a scalar  $b$  takes the form (Sherman and Morrison, 1950)

$$(\mathbf{A} + b\mathbf{z}\mathbf{z}^\top)^{-1} = \mathbf{A}^{-1} - \frac{b}{1 + b(\mathbf{z}^\top \mathbf{A}^{-1} \mathbf{z})} (\mathbf{A}^{-1} \mathbf{z}) (\mathbf{A}^{-1} \mathbf{z})^\top.$$

741 Since  $F^{\text{WG}}(\Phi^T)$  is a rank-1 update of  $\Phi^T$  by Eq. (6), then  $\mathbf{V}'$  is also a rank-1 update of  $\mathbf{V}$ :

$$\begin{aligned} \mathbf{V}' &= 2\sigma^{2'} \Phi^{T'} + \sigma_\epsilon^2 \mathbf{I} \\ &= 2\sigma^2 (\Phi^T - \tilde{\varphi}^T \mathbf{1} \mathbf{1}^\top) + \sigma_\epsilon^2 \mathbf{I} \\ &= \mathbf{V} - d \mathbf{1} \mathbf{1}^\top, \end{aligned}$$

742 where  $d = 2\sigma^2 \tilde{\varphi}^T$  and we used  $\sigma^{2'} = (1 - \tilde{\varphi}^T) \sigma^2$ . Therefore,

$$(\mathbf{V}')^{-1} = \mathbf{V}^{-1} + e \mathbf{V}^{-1} \mathbf{1} (\mathbf{V}^{-1} \mathbf{1})^\top,$$

<sup>743</sup> where  $e = d / (1 - d(\mathbf{1}^\top \mathbf{V}^{-1} \mathbf{1}))$ . Therefore the following remains a rank-1 update,

$$\mathbf{Z}^\top (\mathbf{V}')^{-1} \mathbf{Z} = \mathbf{Z}^\top \mathbf{V}^{-1} \mathbf{Z} + e \mathbf{u} \mathbf{u}^\top,$$

<sup>744</sup> where  $\mathbf{u} = \mathbf{Z}^\top \mathbf{V}^{-1} \mathbf{1}$  is a column vector the length of the number of covariates (including intercept

<sup>745</sup> and genotype). Therefore,

$$(\mathbf{Z}^\top (\mathbf{V}')^{-1} \mathbf{Z})^{-1} = (\mathbf{Z}^\top \mathbf{V}^{-1} \mathbf{Z})^{-1} - g \mathbf{v} \mathbf{v}^\top,$$

<sup>746</sup> where  $\mathbf{v} = (\mathbf{Z}^\top \mathbf{V}^{-1} \mathbf{Z})^{-1} \mathbf{u}$  and  $g = e / (1 + e(\mathbf{u}^\top \mathbf{v}))$ . Noting that  $\mathbf{Z}^\top \mathbf{V}^{-1} \mathbf{1}$  is the first column of

<sup>747</sup>  $\mathbf{Z}^\top \mathbf{V}^{-1} \mathbf{Z}$ , then  $\mathbf{v}$  is the first column of the identity matrix:

$$\mathbf{v} = (\mathbf{Z}^\top \mathbf{V}^{-1} \mathbf{Z})^{-1} \mathbf{Z}^\top \mathbf{V}^{-1} \mathbf{1} = \begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix},$$

<sup>748</sup> where  $\mathbf{0}$  is a vector the length of the number of covariates minus one (exclude the intercept). As a

<sup>749</sup> consequence,  $\mathbf{Z} \mathbf{v} = \mathbf{1}$ , so  $\mathbf{u}^\top \mathbf{v} = \mathbf{1}^\top \mathbf{V}^{-1} \mathbf{1}$  and

$$\begin{aligned} g &= \frac{e}{1 + e(\mathbf{1}^\top \mathbf{V}^{-1} \mathbf{1})} \\ &= \frac{\frac{d}{1 - d(\mathbf{1}^\top \mathbf{V}^{-1} \mathbf{1})}}{1 + \frac{d}{1 - d(\mathbf{1}^\top \mathbf{V}^{-1} \mathbf{1})}(\mathbf{1}^\top \mathbf{V}^{-1} \mathbf{1})} \\ &= d. \end{aligned}$$

<sup>750</sup> The final step yields the coefficient estimates as a rank-1 update:

$$\begin{aligned} \hat{\beta}' &= (\mathbf{Z}^\top (\mathbf{V}')^{-1} \mathbf{Z})^{-1} \mathbf{Z}^\top (\mathbf{V}')^{-1} \mathbf{y} \\ &= ((\mathbf{Z}^\top \mathbf{V}^{-1} \mathbf{Z})^{-1} - d \mathbf{v} \mathbf{v}^\top) \mathbf{Z}^\top (\mathbf{V}^{-1} + e \mathbf{V}^{-1} \mathbf{1} (\mathbf{V}^{-1} \mathbf{1})^\top) \mathbf{y} \\ &= \hat{\beta} + e \mathbf{v} (\mathbf{1}^\top \mathbf{V}^{-1} \mathbf{y}) - d \mathbf{v} (\mathbf{1}^\top \mathbf{V}^{-1} \mathbf{y}) - d e \mathbf{v} (\mathbf{1}^\top \mathbf{V}^{-1} \mathbf{1}) (\mathbf{1}^\top \mathbf{V}^{-1} \mathbf{y}) \\ &= \hat{\beta} + \mathbf{v} (\mathbf{1}^\top \mathbf{V}^{-1} \mathbf{y}) (e - d - d e (\mathbf{1}^\top \mathbf{V}^{-1} \mathbf{1})). \end{aligned}$$

751 The last factor above vanishes:

$$\begin{aligned} e - d - de(\mathbf{1}^\top \mathbf{V}^{-1} \mathbf{1}) &= \frac{d}{1 - d(\mathbf{1}^\top \mathbf{V}^{-1} \mathbf{1})} - d - d \frac{d}{1 - d(\mathbf{1}^\top \mathbf{V}^{-1} \mathbf{1})} (\mathbf{1}^\top \mathbf{V}^{-1} \mathbf{1}) \\ &= 0. \end{aligned}$$

752 Therefore,  $\hat{\beta}' = \hat{\beta}$ , which shows that all fixed effect coefficients, including the intercept, are invariant  
753 to using a WG-biased kinship matrix instead of the unbiased one when the coefficients are estimated  
754 with generalized least squares.

755 Furthermore, since the diagonal values of  $(\mathbf{Z}^\top (\mathbf{V}')^{-1} \mathbf{Z})^{-1}$ , which correspond to  $\text{Var}(\hat{\beta}'_k)$  for each  
756  $k$ , are the same as those of  $(\mathbf{Z}^\top \mathbf{V}^{-1} \mathbf{Z})^{-1}$  except for the first one corresponding to the intercept, then  
757 the Wald test statistic of the  $k$ th covariate coefficients, given by  $\hat{\beta}_k^2 / \text{Var}(\hat{\beta}_k)$ , and their p-values,  
758 are also the same for  $k \neq 1$  for WG bias as for the unbiased kinship matrix.

## Supplemental figures

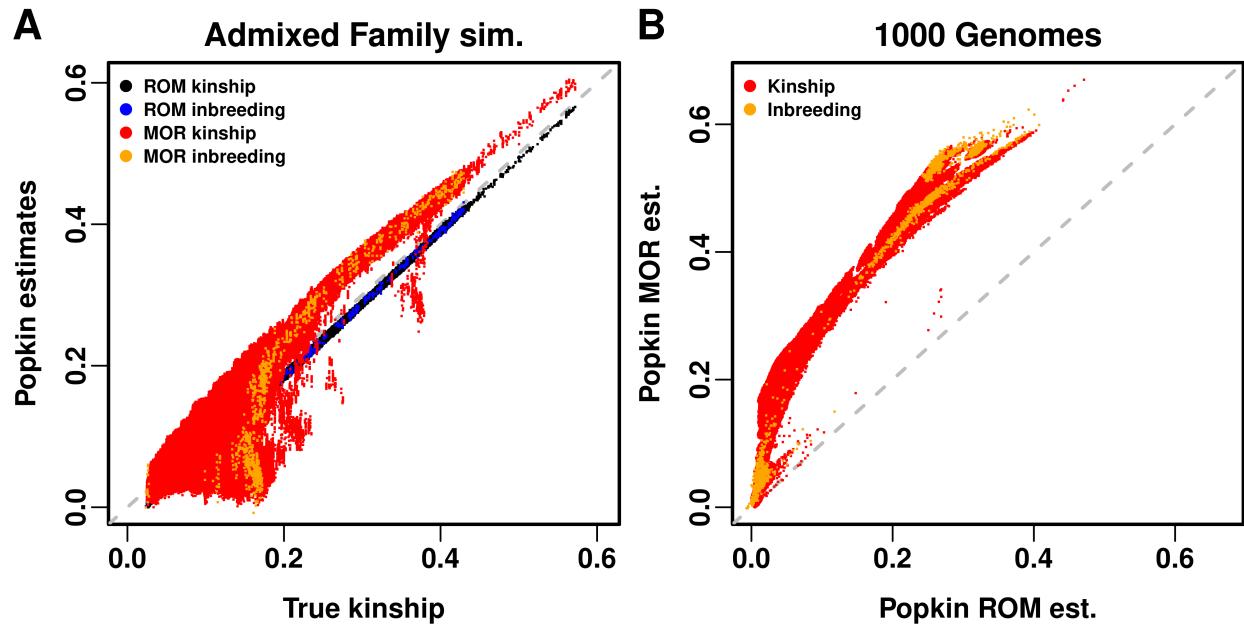


Figure S1: **Comparison of popkin ROM and MOR estimates.** Kinship (off-diagonal of matrix) and inbreeding (transformed diagonal) are plotted in different colors, which shows that their biases (if any) overlap. **A.** In admixed family simulation, both estimates are compared against true kinship. Popkin ROM has a negligible bias, due to the minimum true kinship of the simulation being slightly larger than zero. Popkin MOR has considerable biases, tending to be upward though not always. **B.** In 1000 Genomes, since true kinship is unknown, popkin ROM takes its place. Popkin MOR biases take on a similar shape as panel A.

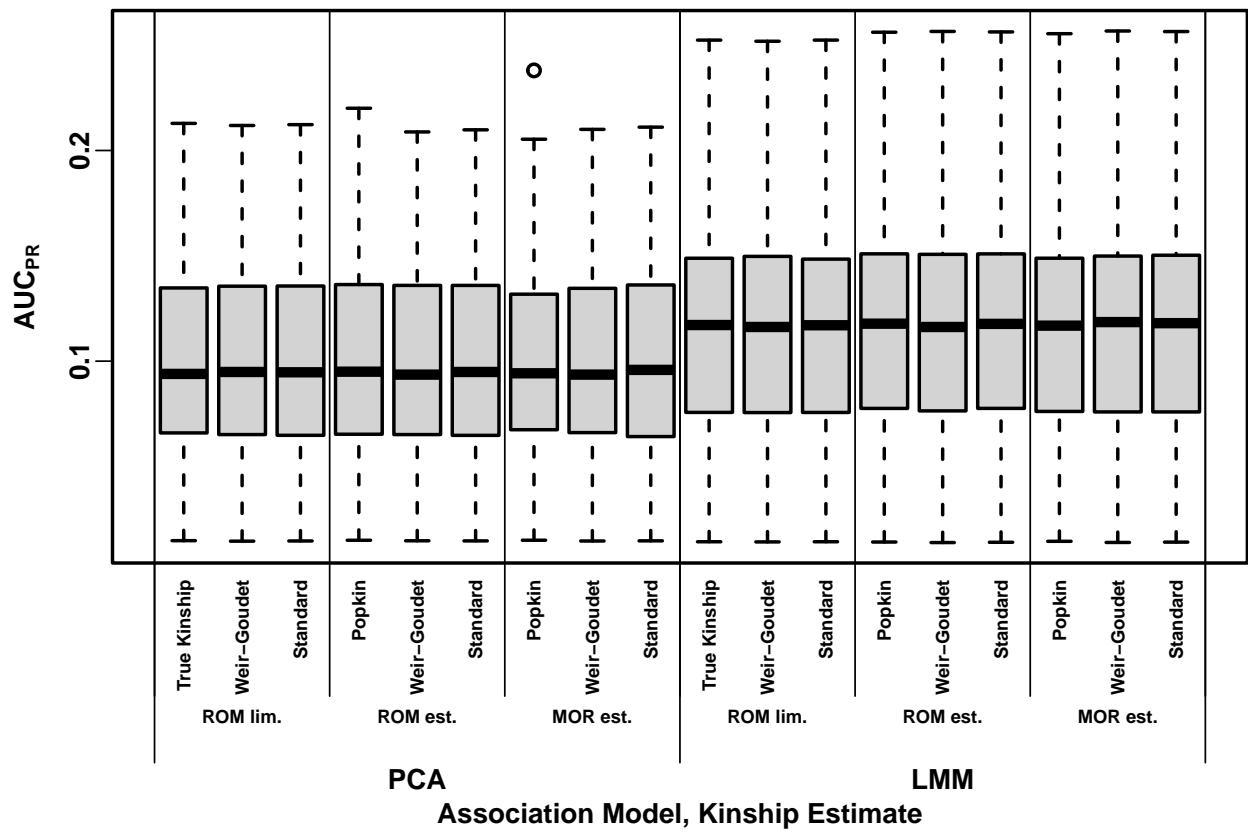


Figure S2: Distributions of Area Under the Precision-Recall Curve (AUC<sub>PR</sub>) on the admixed family simulation with  $h^2 = 0.3$ . Like Fig. 2 except simulated with lower heritability.

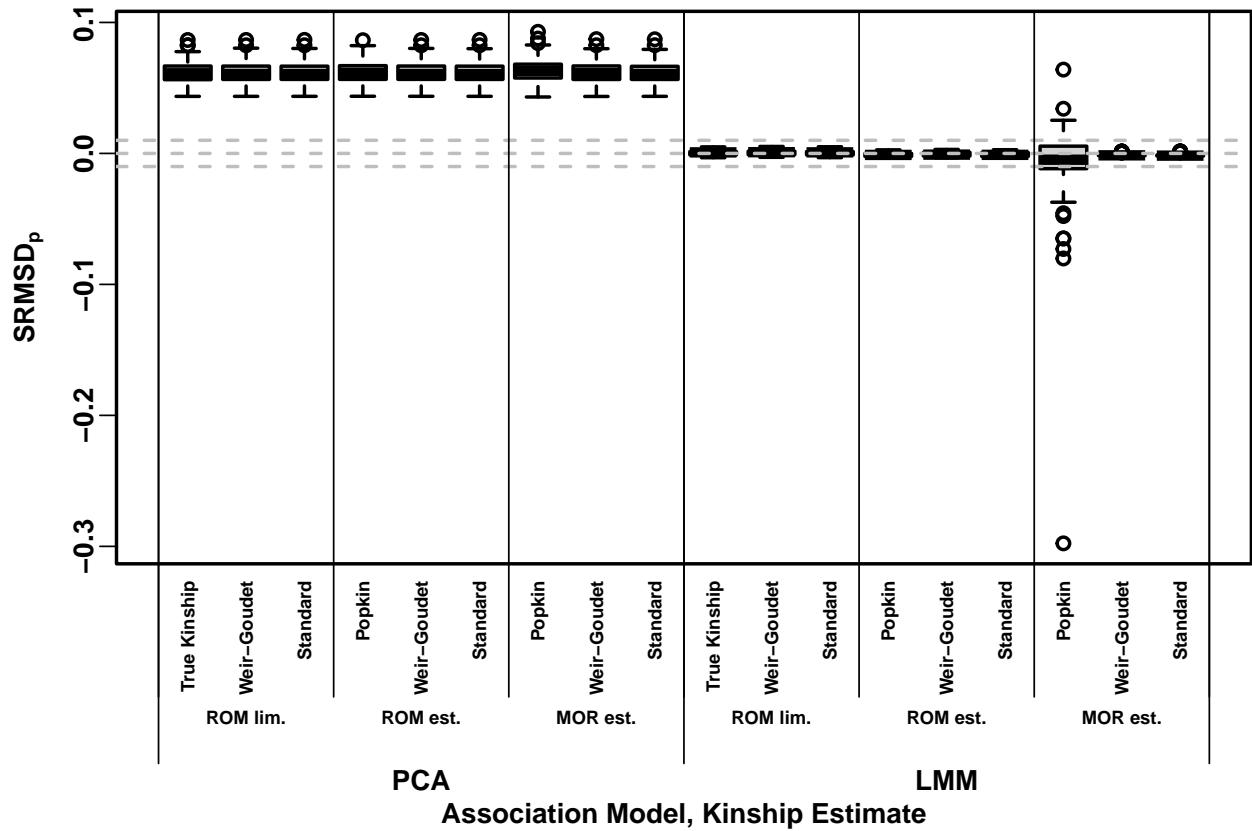


Figure S3: Signed Root Mean Square Deviation of null p-values ( $\text{SRMSD}_p$ ) on the admixed family simulation with  $h^2 = 0.8$ . Same methods and simulation as Fig. 2, see that for more information.  $|\text{SRMSD}_p| < 0.01$  (area between gray dashed lines) is considered calibrated. All PCA runs are miscalibrated by similar amounts, whereas most LMM runs are calibrated with few exceptions.

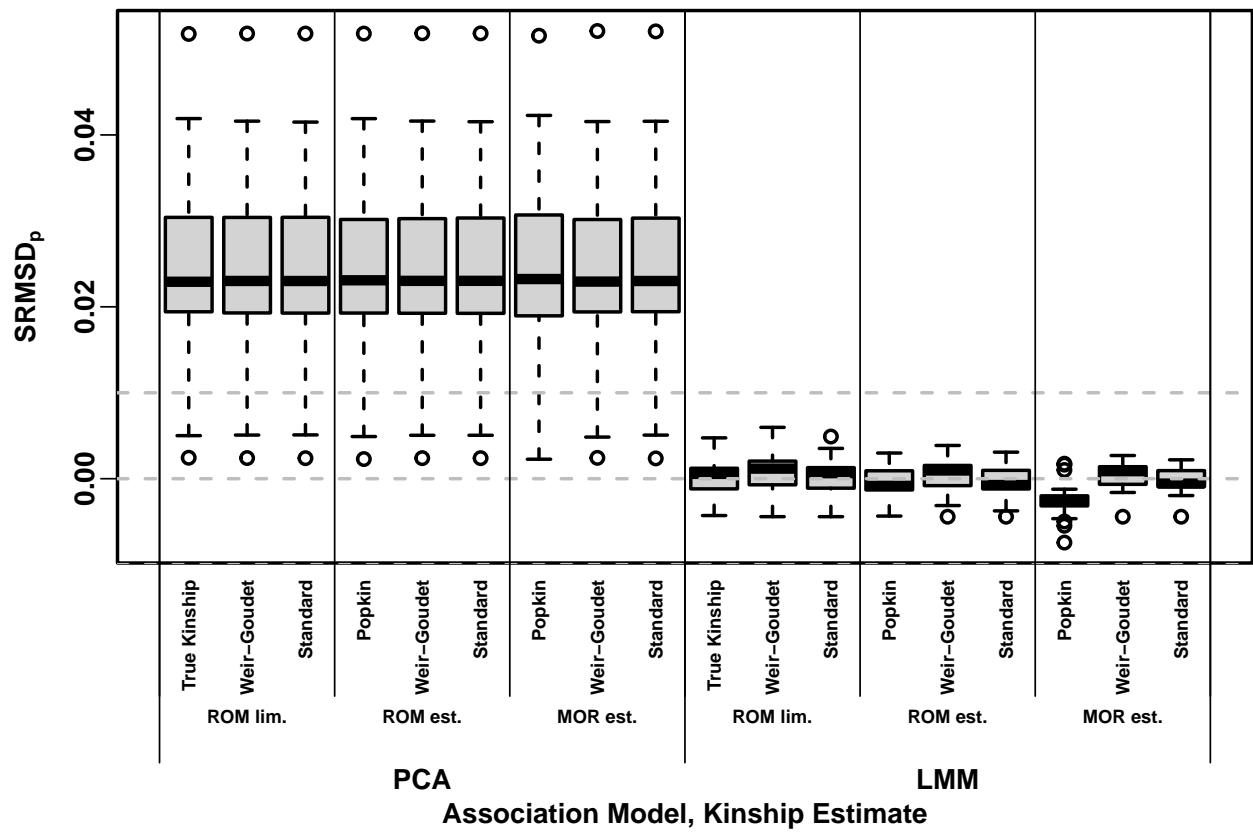


Figure S4: Signed Root Mean Square Deviation of null p-values ( $\text{SRMSD}_p$ ) on the admixed family simulation with  $h^2 = 0.3$ . Like Fig. S3 except simulated with lower heritability.

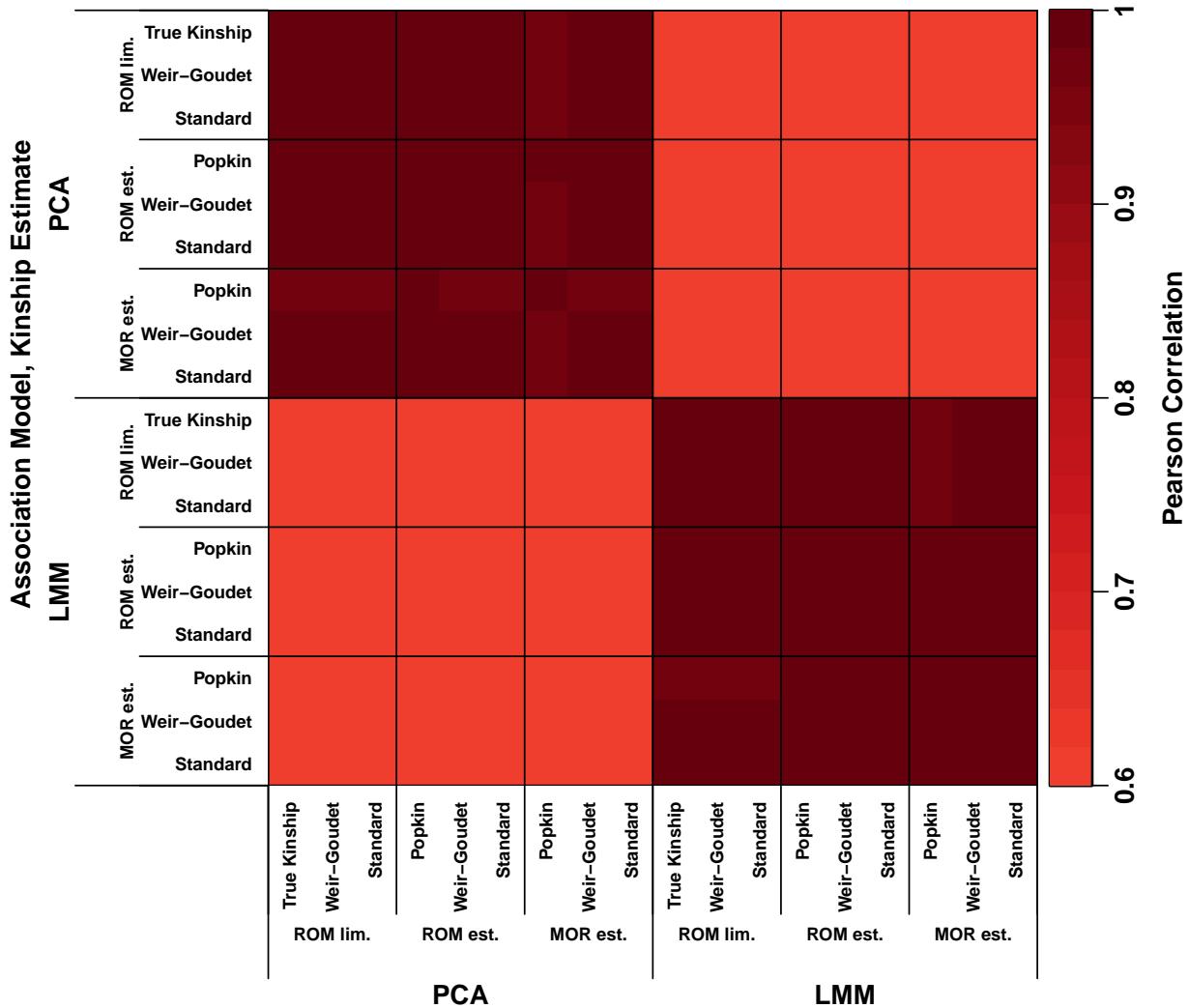


Figure S5: Correlation between p-values on the admixed family simulation with  $h^2 = 0.8$ . Like Fig. 3 but measuring correlation instead of agreement.

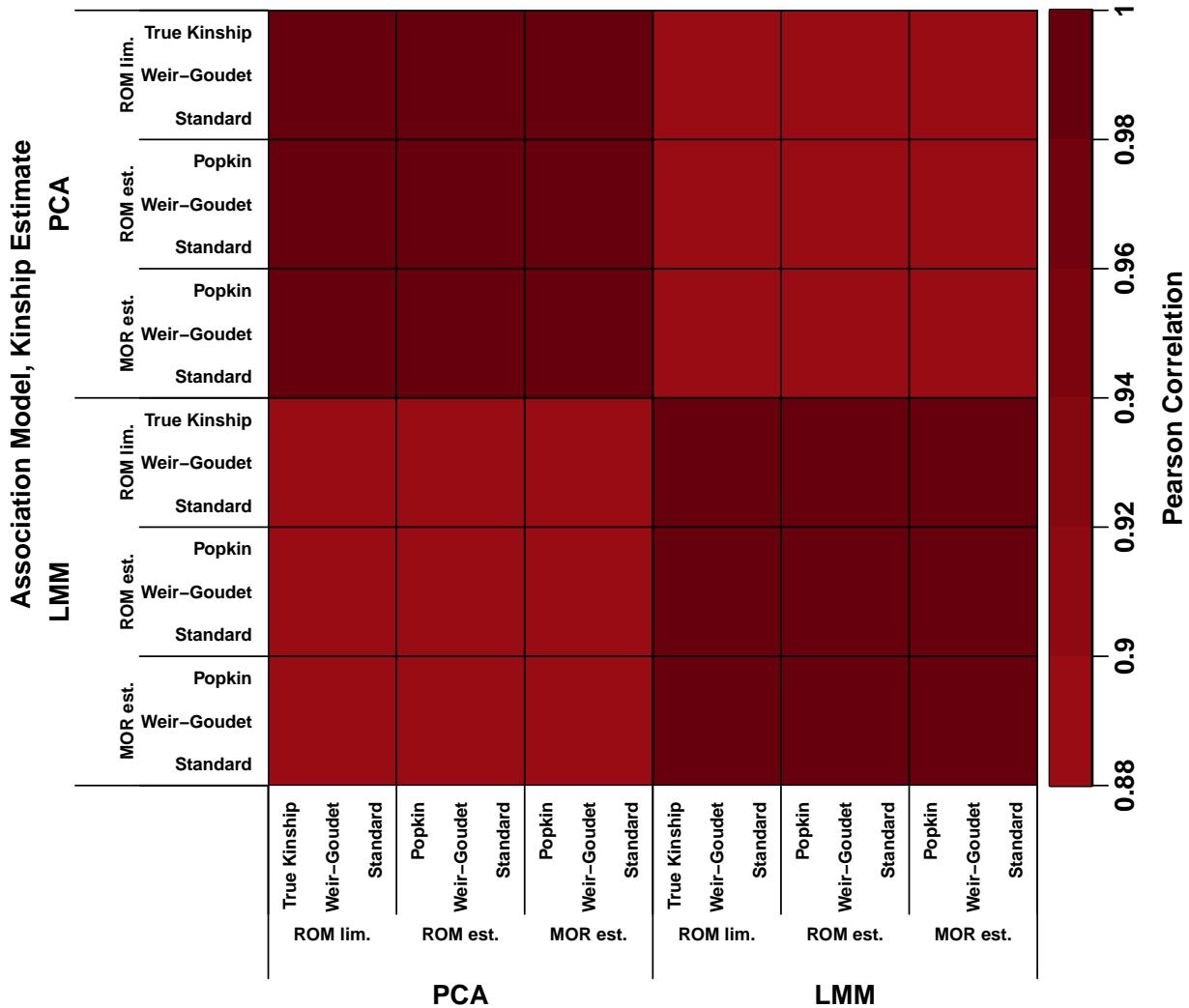


Figure S6: Correlation between p-values on the admixed family simulation with  $h^2 = 0.3$ . Like Fig. S5 except simulated with lower heritability.

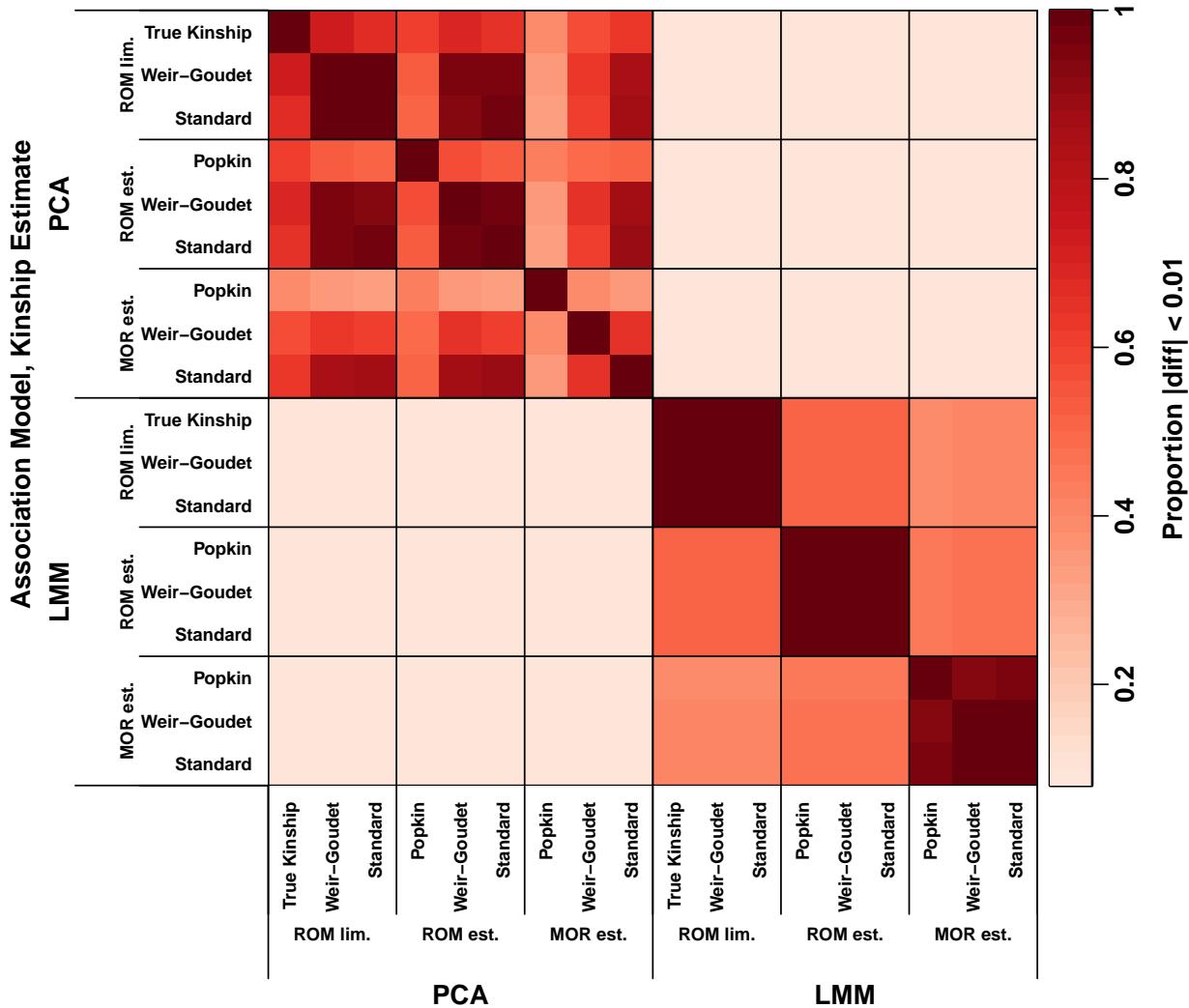


Figure S7: **Agreement between p-values on the admixed family simulation with  $h^2 = 0.3$ .**  
Like Fig. 3 except simulated with lower heritability.

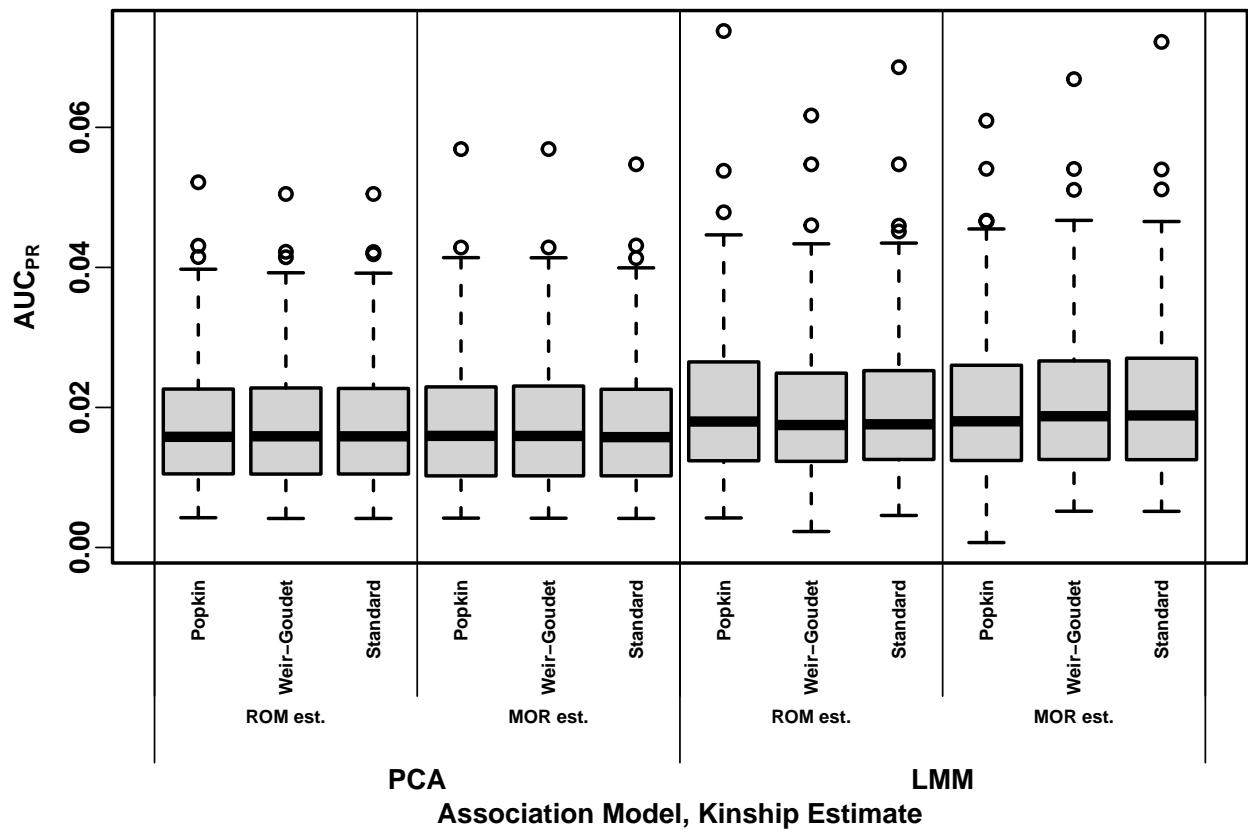


Figure S8: **Distributions of Area Under the Precision-Recall Curve (AUC<sub>PR</sub>) on 1000 Genomes with  $h^2 = 0.3$ .** Like Fig. 5 except simulated with lower heritability.

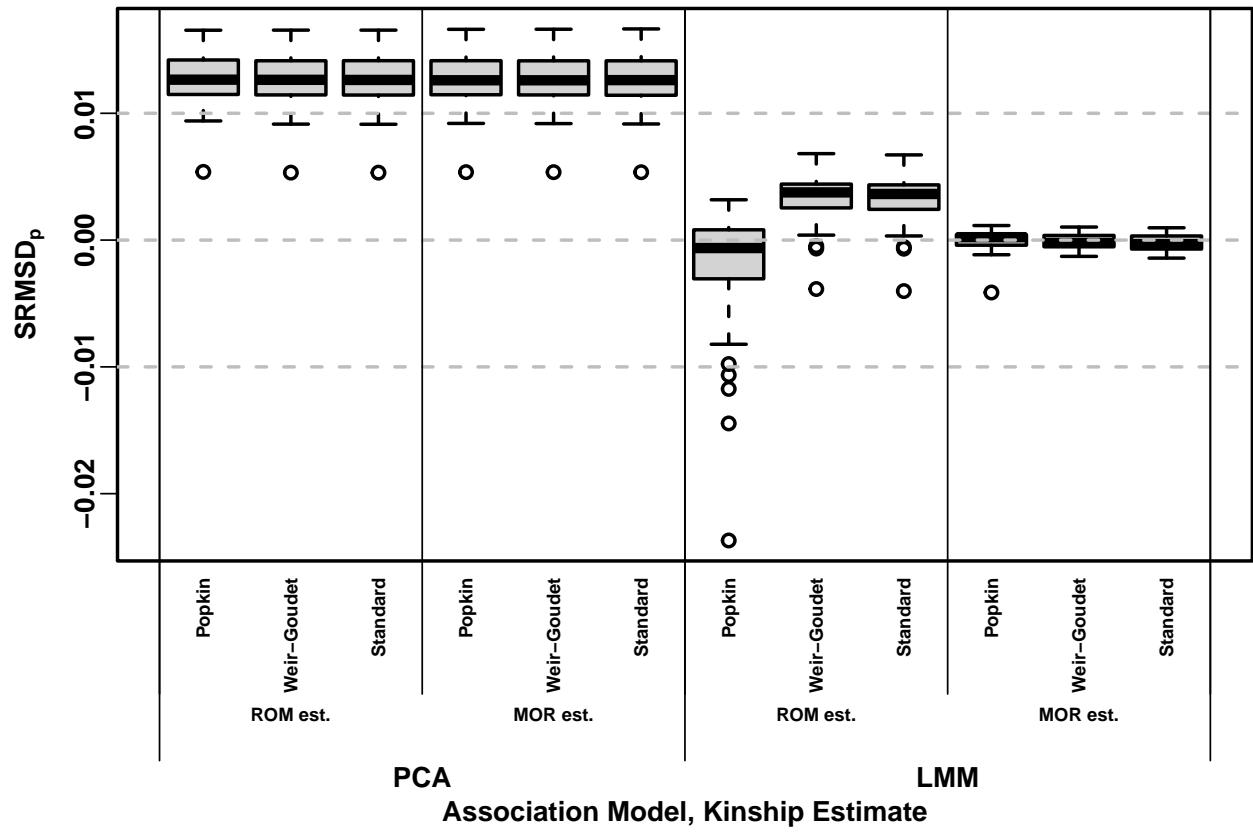


Figure S9: Signed Root Mean Square Deviation of null p-values ( $\text{SRMSD}_p$ ) on 1000 Genomes with  $h^2 = 0.8$ . Same methods and simulation as Fig. 5, and y-axis statistic and conclusions of Fig. S3, see those for more information.

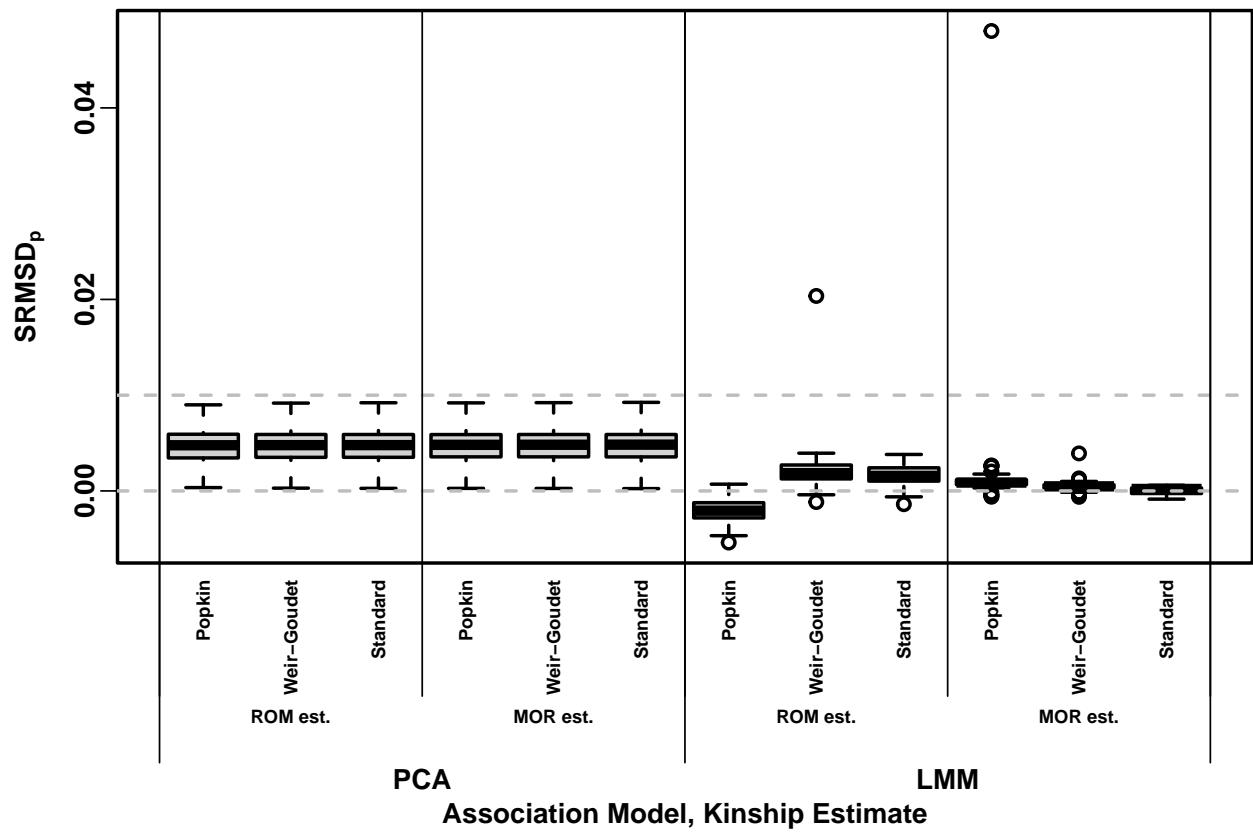


Figure S10: Signed Root Mean Square Deviation of null p-values ( $\text{SRMSD}_p$ ) on 1000 Genomes with  $h^2 = 0.3$ . Like Fig. S9 except simulated with lower heritability.

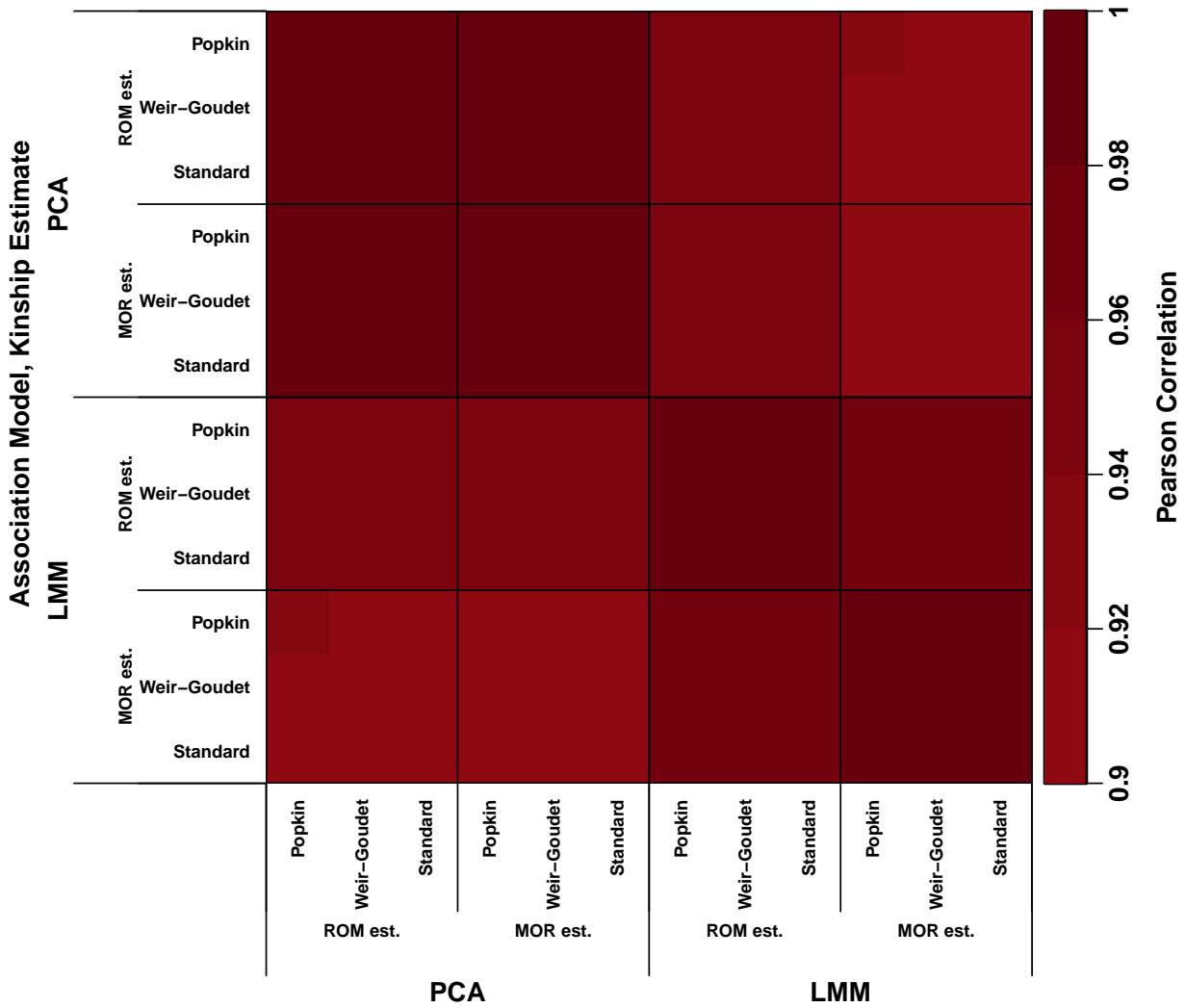


Figure S11: Correlation between p-values on 1000 Genomes with  $h^2 = 0.8$ . See Fig. S5 for more details.

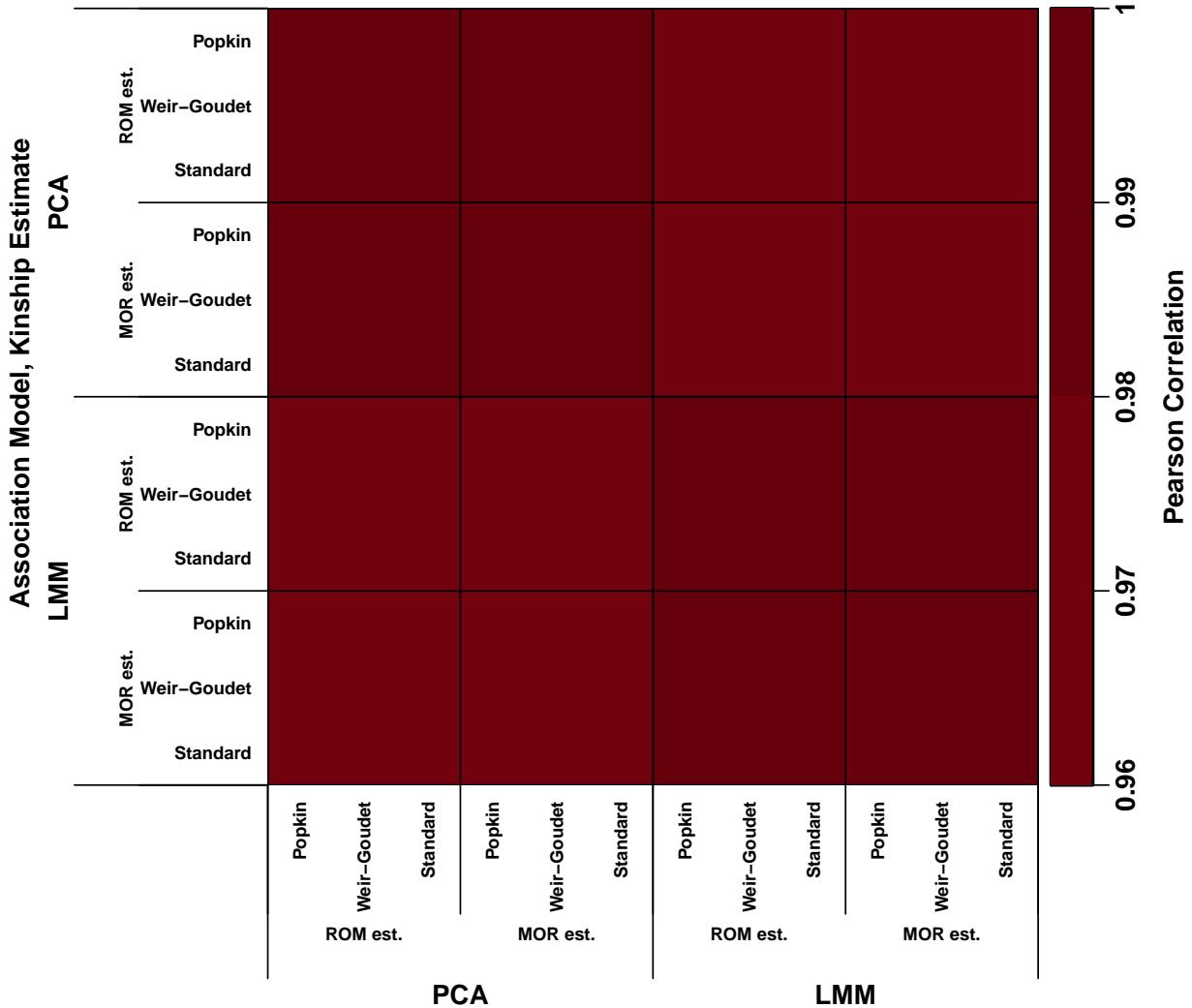


Figure S12: Correlation between p-values on 1000 Genomes with  $h^2 = 0.3$ . Like Fig. S11 except simulated with lower heritability.

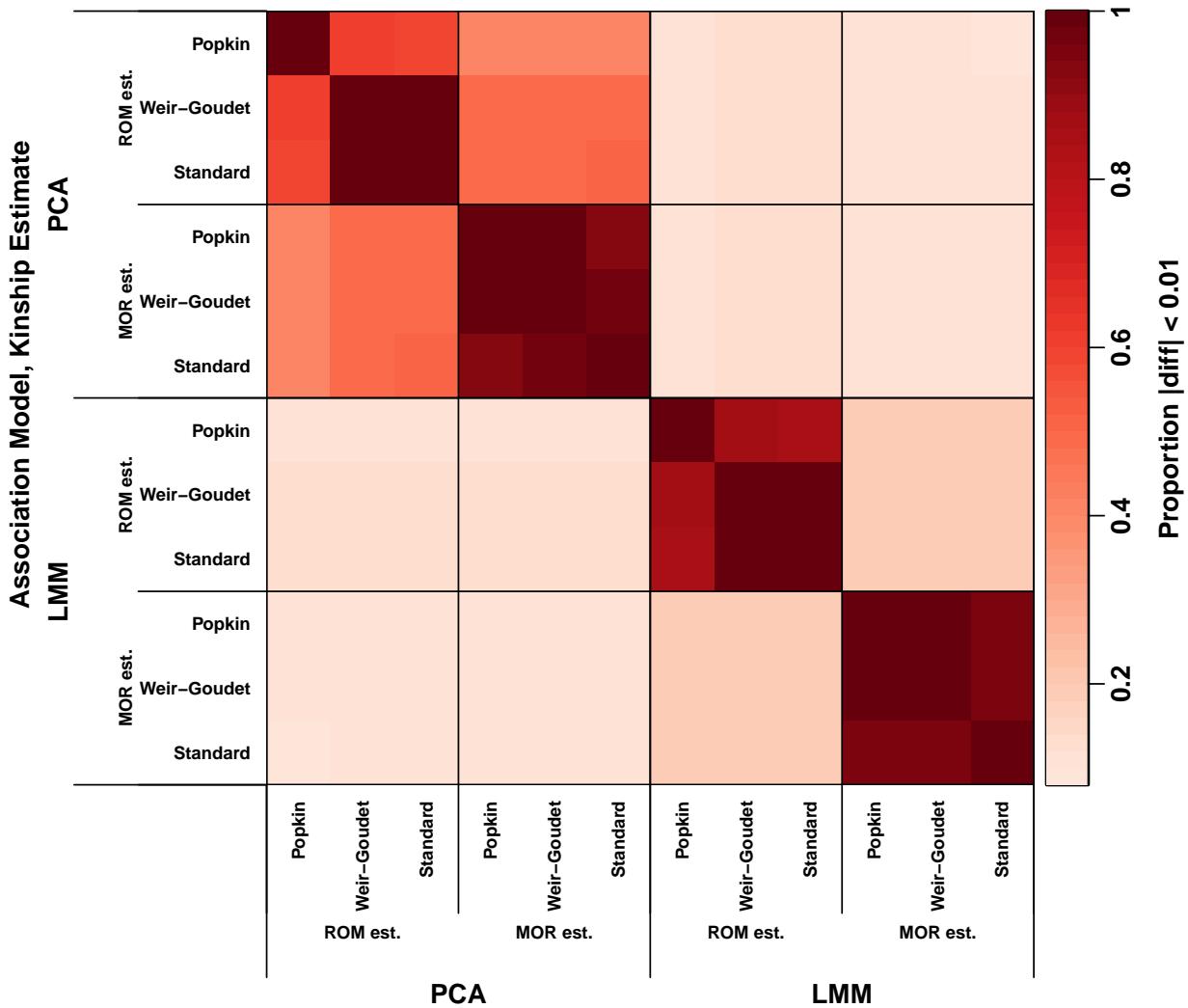


Figure S13: Agreement between p-values on 1000 Genomes with  $h^2 = 0.8$ . See Fig. 3 for more details.

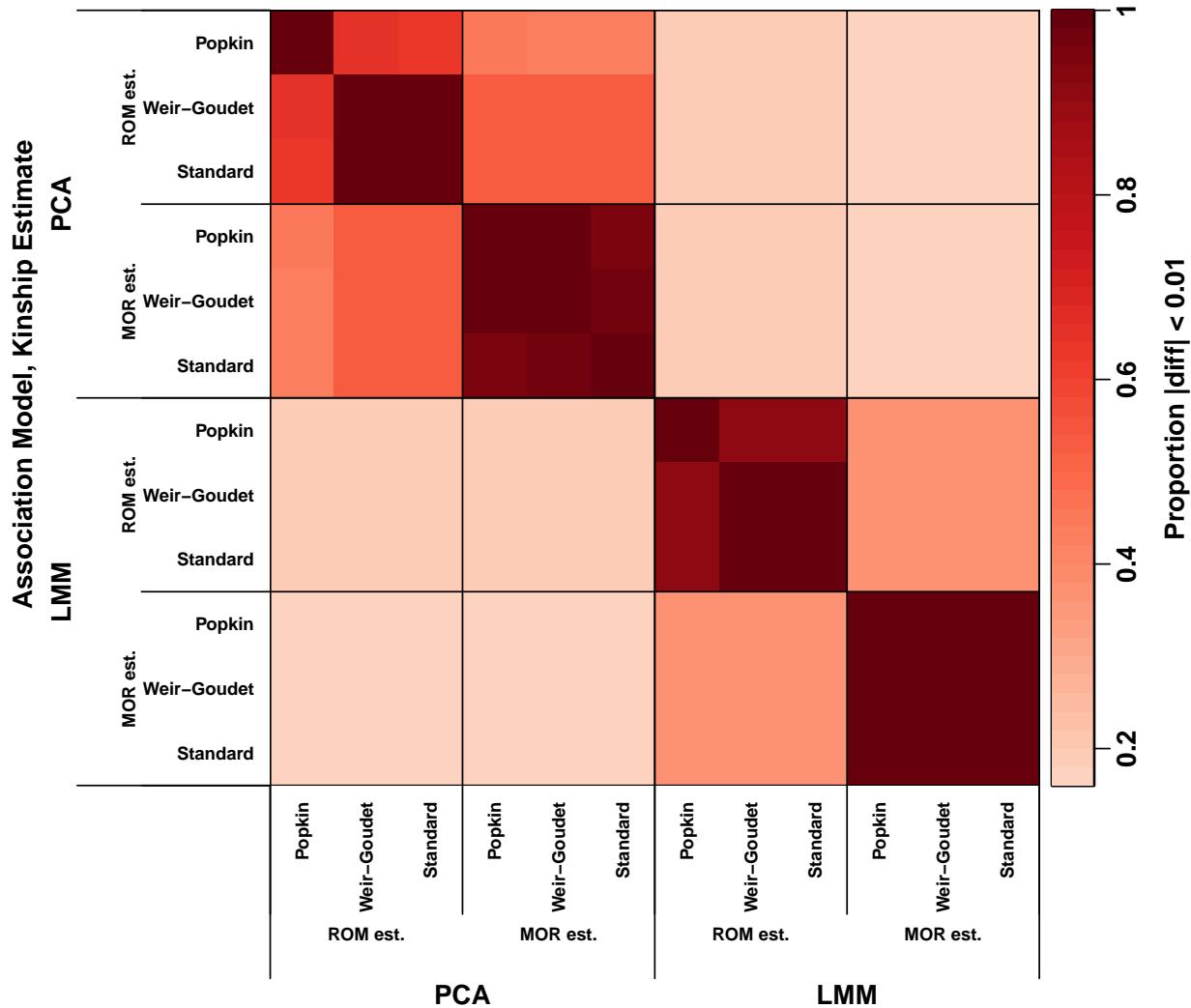


Figure S14: **Agreement between p-values on 1000 Genomes with  $h^2 = 0.3$ .** Like Fig. S13 except simulated with lower heritability.

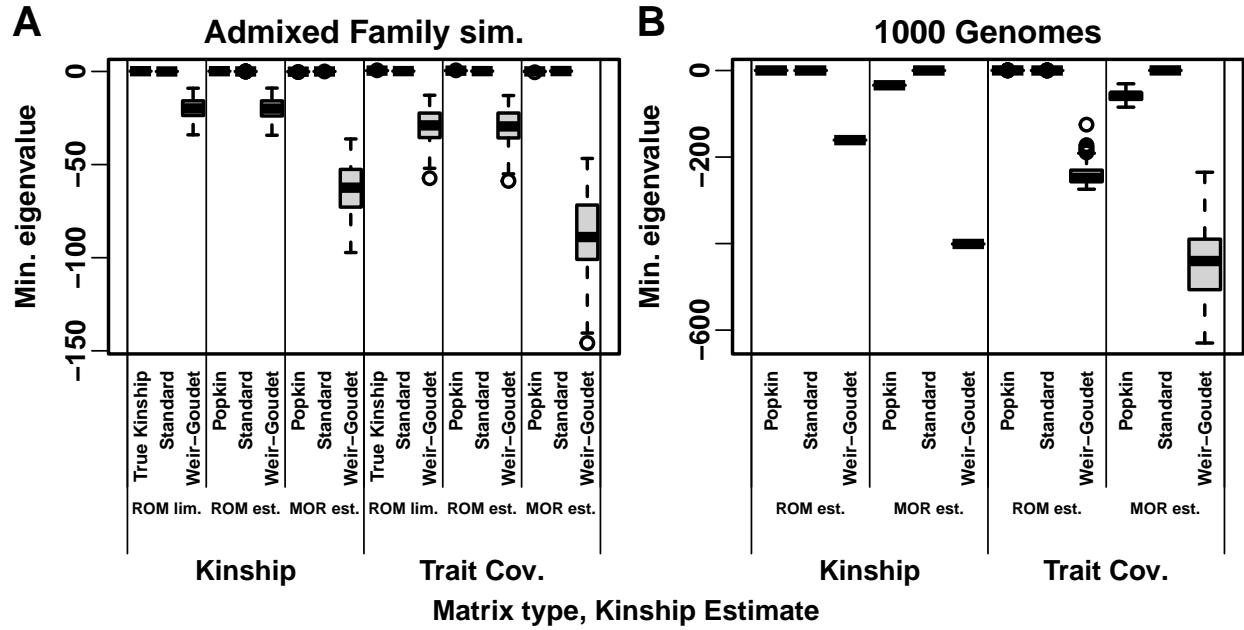


Figure S15: Minimum eigenvalue of kinship and trait covariance ( $\mathbf{V}$ ) matrices with  $h^2 = 0.8$ . Each distribution is over 100 replicates (1000 Genomes kinship has one value since genotypes are fixed, but  $\mathbf{V}$  varies per replicate). All WG matrices has very large negative eigenvalues, and Popkin MOR has negative eigenvalues as well; in these cases  $\mathbf{V}$  always has negative eigenvalues too.

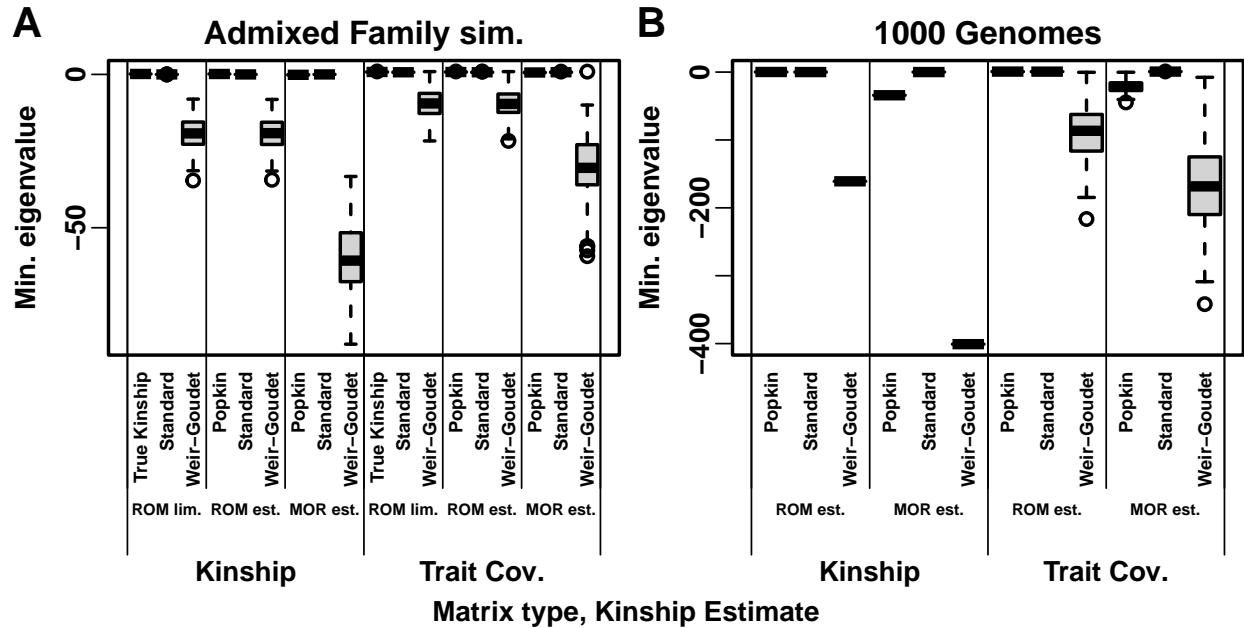


Figure S16: Minimum eigenvalue of kinship and trait covariance ( $\mathbf{V}$ ) matrices with  $h^2 = 0.3$ . Like Fig. S15 except simulated with lower heritability.

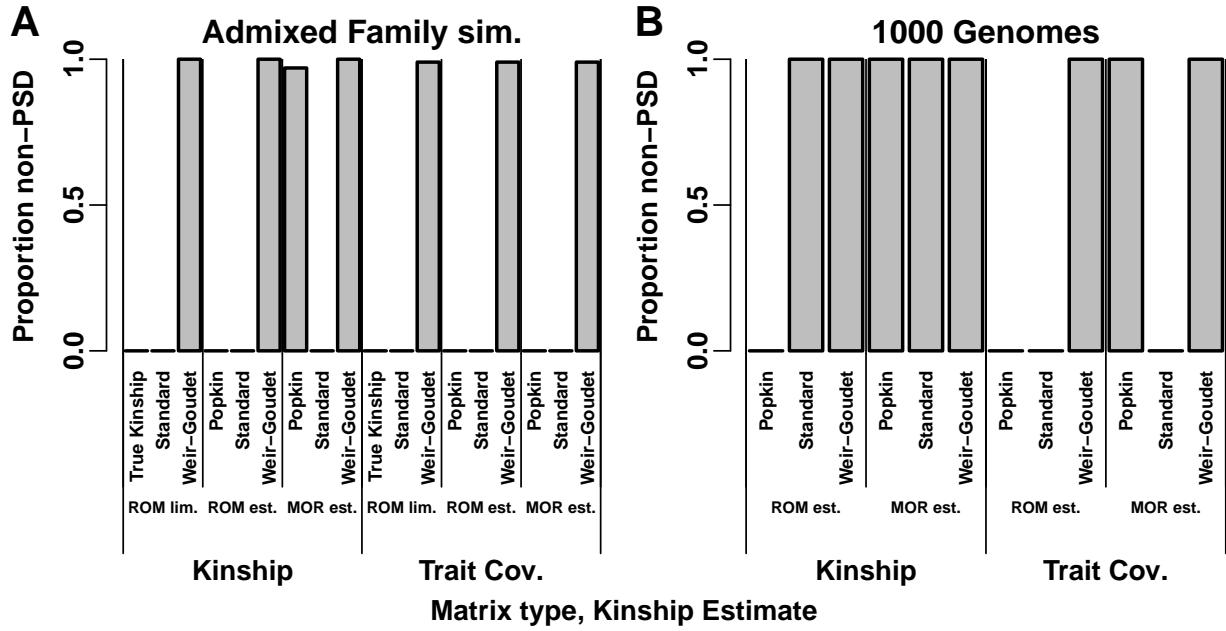


Figure S17: Proportion of kinship and trait covariance ( $\mathbf{V}$ ) matrices with  $h^2 = 0.3$  that are not positive semidefinite (PSD). Like Fig. 7 except simulated with lower heritability.

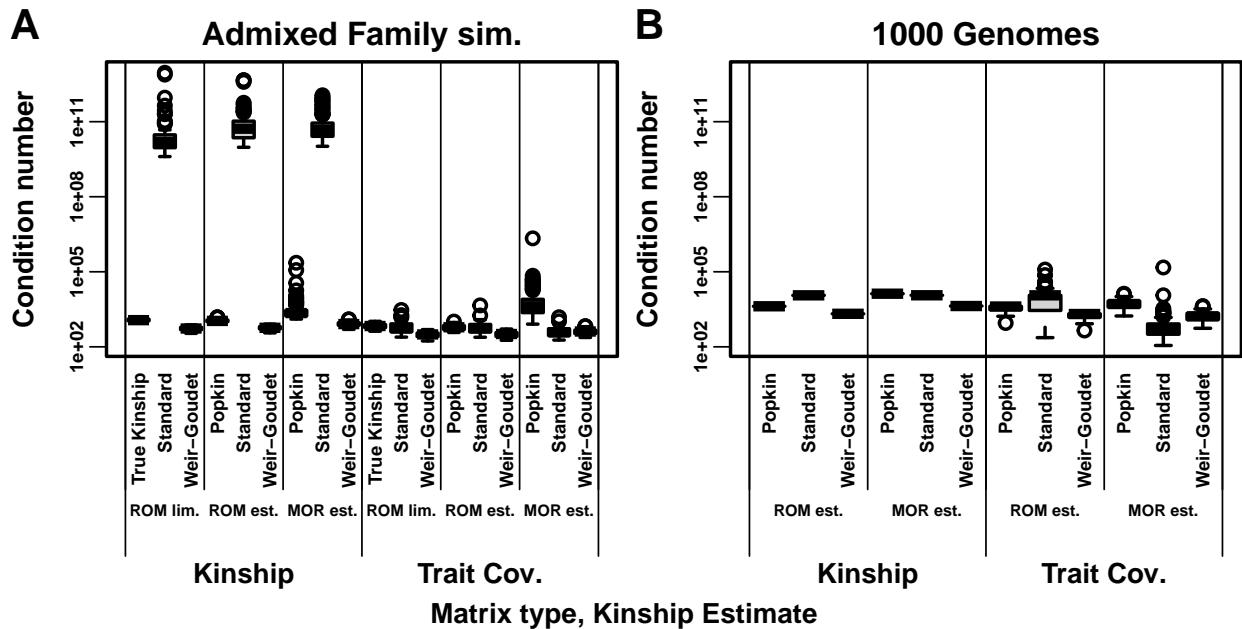


Figure S18: Condition numbers of kinship and trait covariance ( $\mathbf{V}$ ) matrices with  $h^2 = 0.8$ . Larger condition numbers reflect ill-conditioned problems such as near singularity. Each distribution is over 100 replicates (1000 Genomes kinship has one value since genotypes are fixed, but  $\mathbf{V}$  varies per replicate).

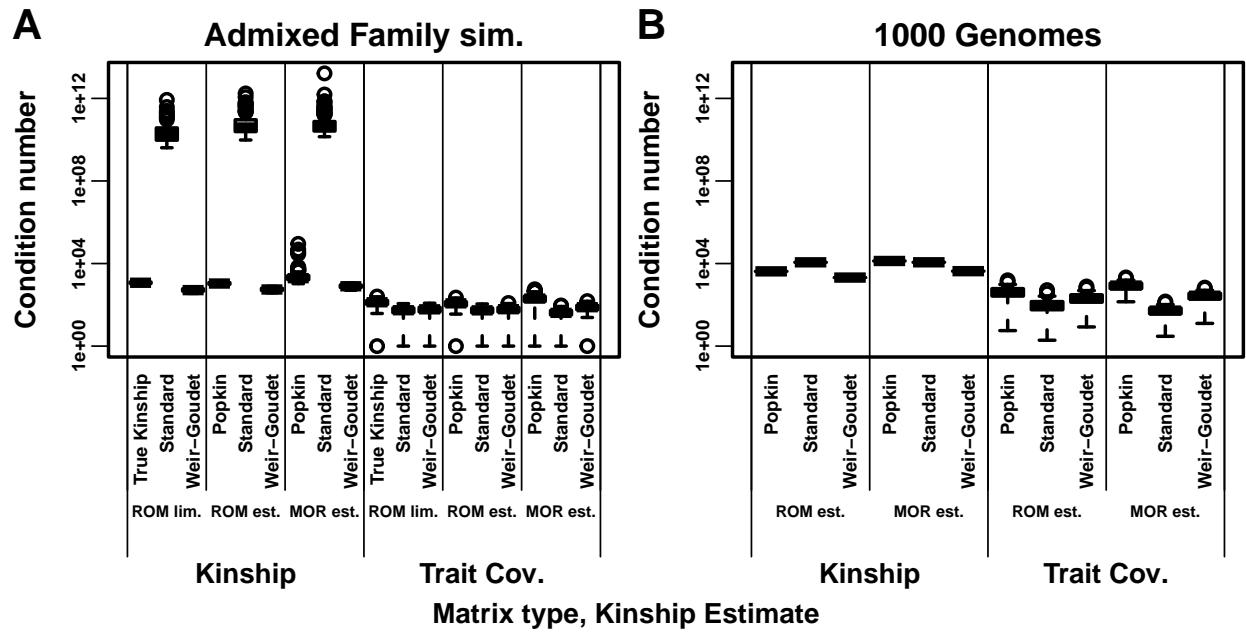


Figure S19: Condition numbers of kinship and trait covariance ( $V$ ) matrices with  $h^2 = 0.3$ . Like Fig. S18 except simulated with lower heritability.

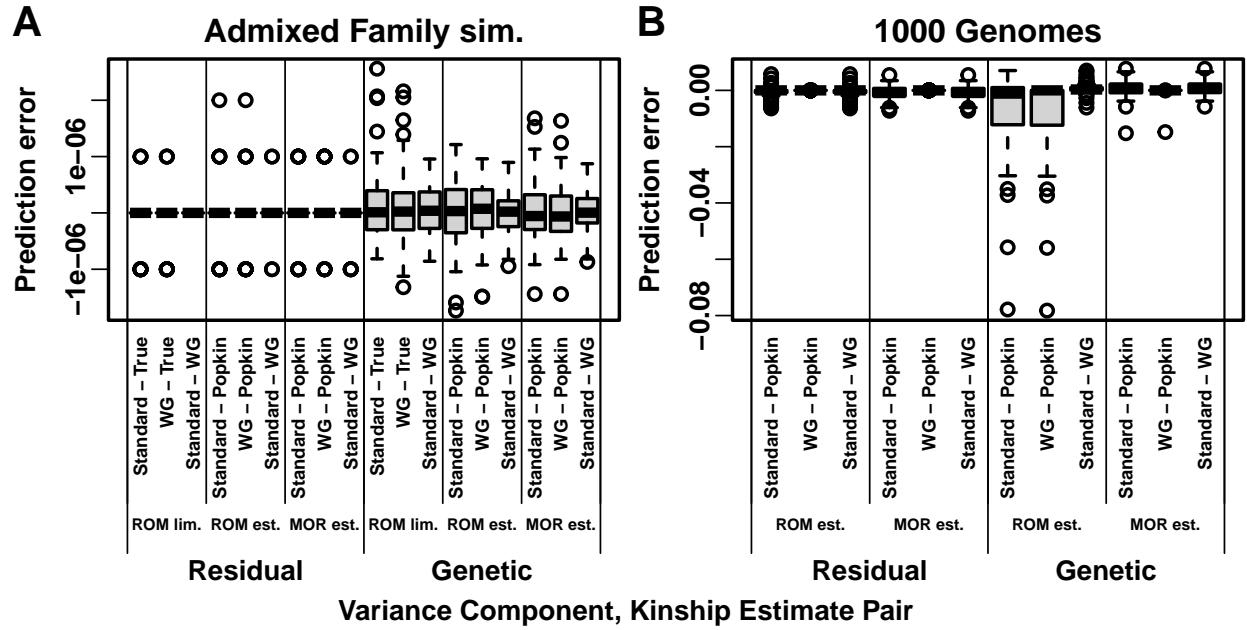


Figure S20: **Variance component prediction errors across evaluations with  $h^2 = 0.8$ .** Here we test the predictions that  $\sigma_{\epsilon}^{2l} = \sigma_{\epsilon}^2$  and  $\sigma^{2l} = c\sigma^2$  in Eq. (16). For Residual, prediction error (y-axis) is  $\sigma_{\epsilon}^{2l} - \sigma_{\epsilon}^2$  between pairs of estimates as listed. For Genetic, prediction error is  $\sigma^{2l} - c\sigma^2$ : The biased-unbiased pairs use  $c = 1 - \bar{\varphi}^T$  for Standard,  $c = 1 - \tilde{\varphi}^T$  for WG,  $\sigma^{2l}$  is their estimate and  $\sigma^2$  is True or Popkin; The Standard-WG pair uses  $\sigma^2$  for WG and  $c = (1 - \bar{\varphi}^T) / (1 - \tilde{\varphi}^T)$ . Each distribution is over the 100 replicates of each simulation. **A.** In admixed family simulation, all errors are zero within machine precision. Excess perfect zero residual prediction errors are due to limited precision of GCTA outputs. **B.** In 1000 Genomes, popkin ROM estimates has large errors compared to Standard and WG.

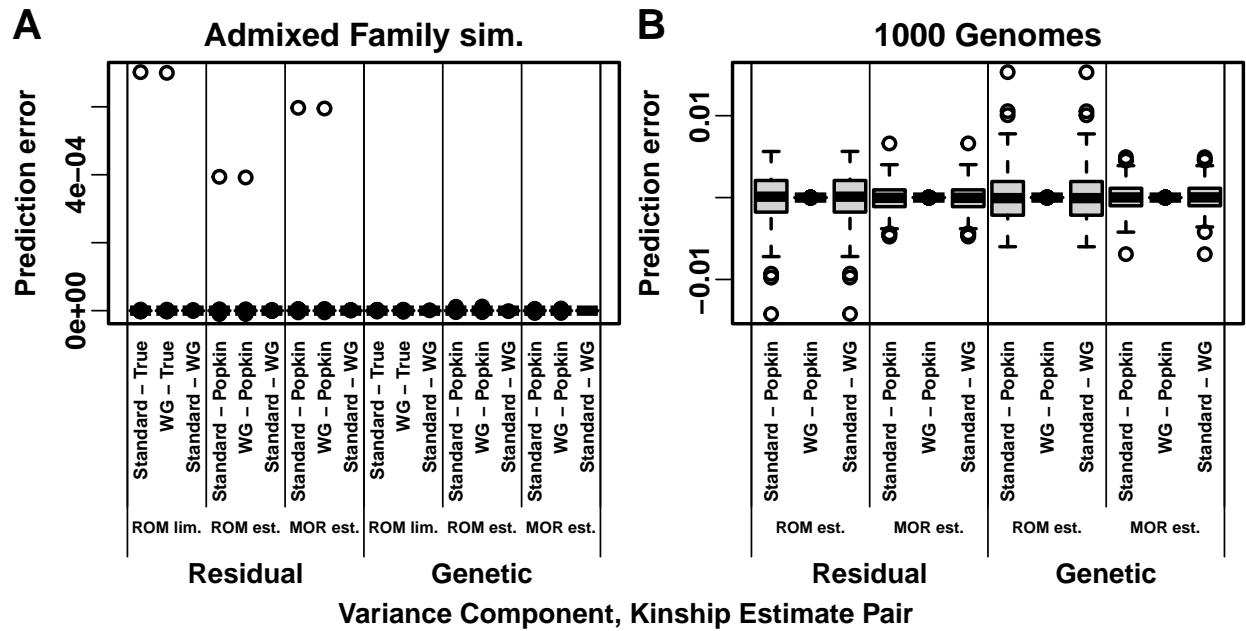


Figure S21: **Variance component prediction errors across evaluations with  $h^2 = 0.3$ .** Like Fig. S20 except simulated with lower heritability. All errors are zero within the reported precision.

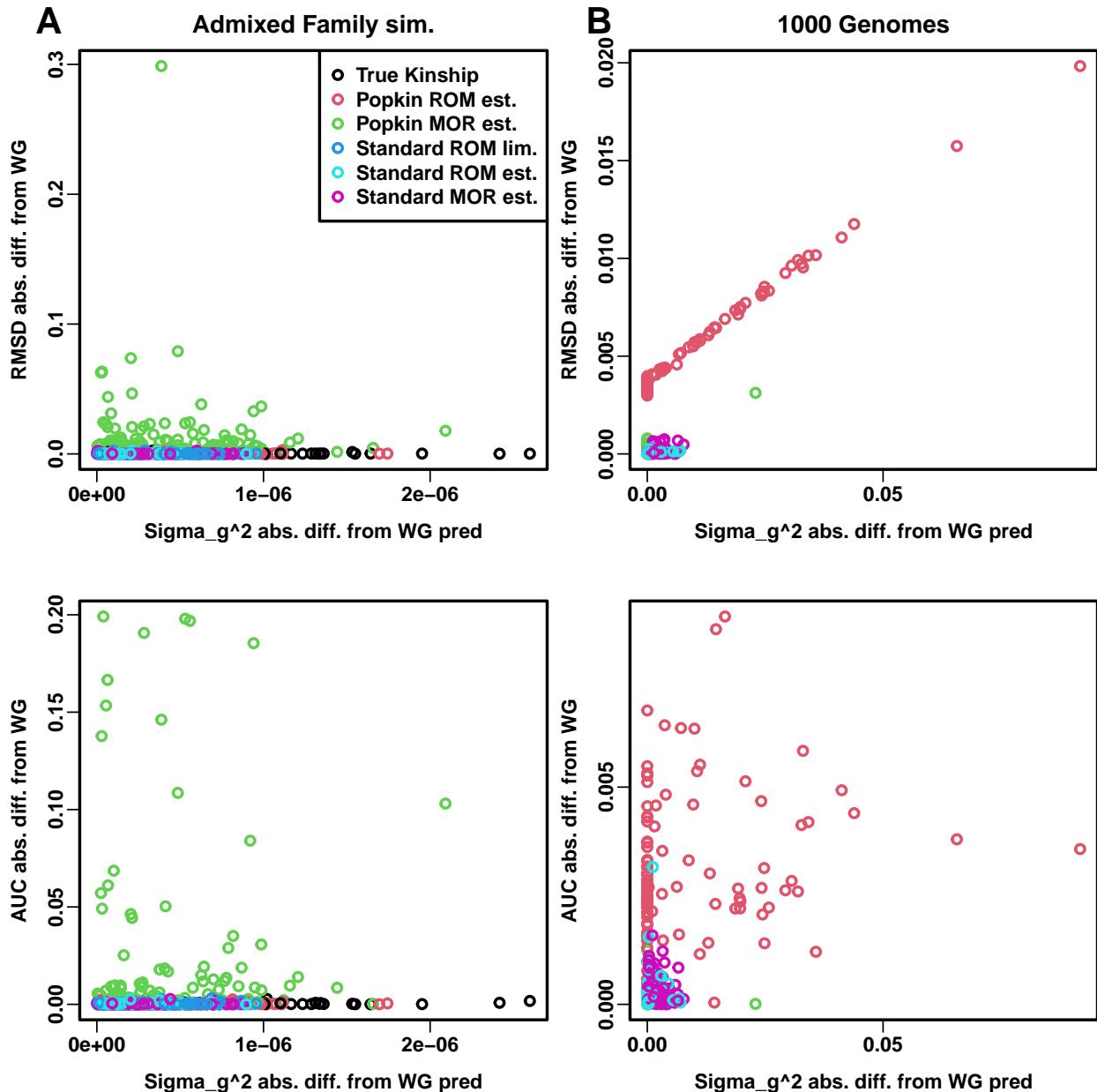


Figure S22: **AUC<sub>PR</sub>** and SRMSD<sub>p</sub> prediction errors explained by variance component errors. Evaluations with  $h^2 = 0.8$  only. Genetic variance component ( $\sigma^2$ ) absolute error is calculated with the formulas in Fig. S20 using WG as reference since its  $\mathbf{V}$  had the lowest condition numbers (Fig. S18). AUC<sub>PR</sub> and SRMSD<sub>p</sub> are expected to be the same between WG, Standard, and True or Popkin (within each locus weight type). **A.** Large errors in the admixed family simulation are not explained by high  $\sigma^2$  error. **B.** Smaller popkin ROM prediction errors in 1000 Genomes are explained by high  $\sigma^2$  error.

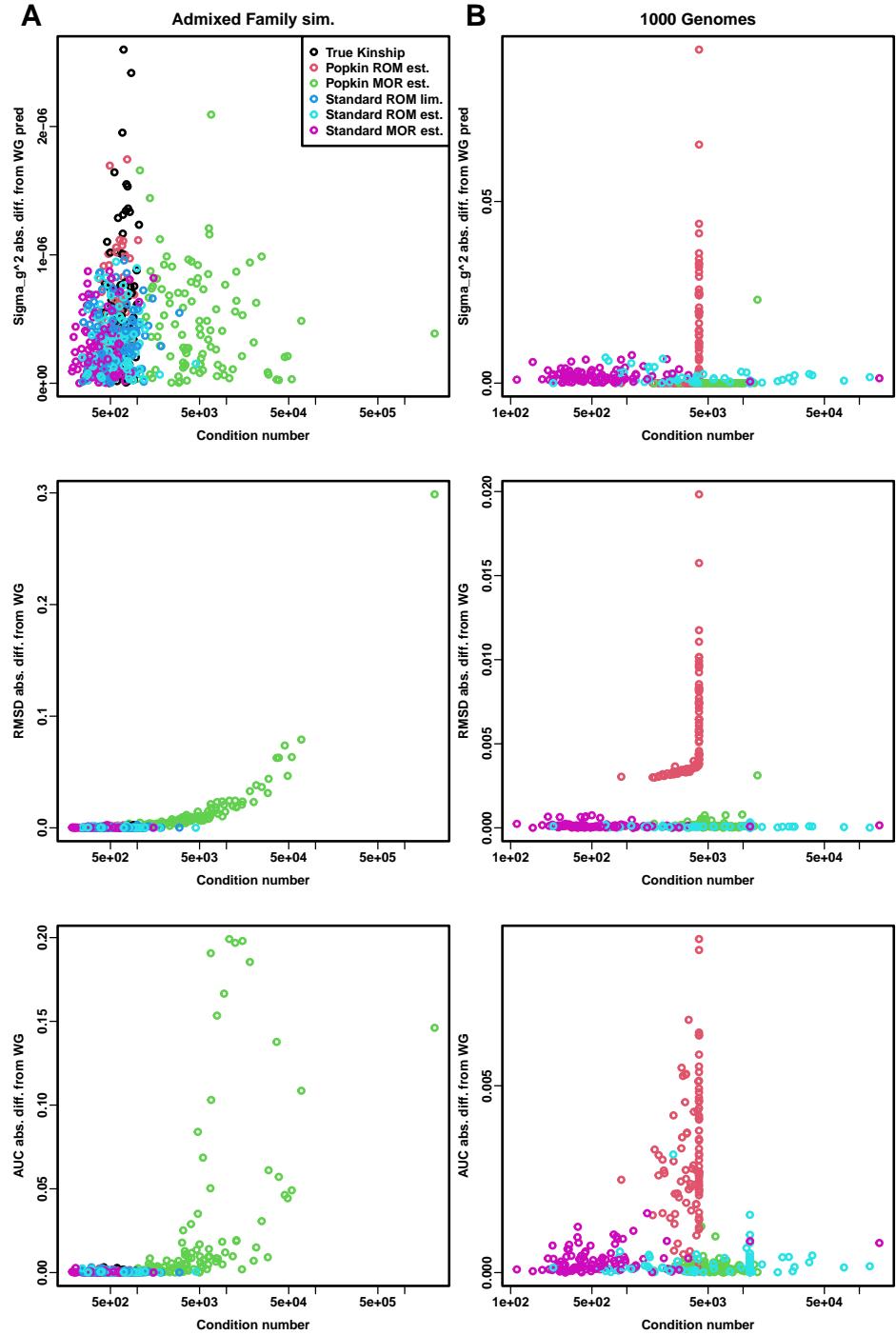


Figure S23: **AUC<sub>PR</sub> and SRMSD<sub>p</sub> prediction errors explained by the condition number of  $\mathbf{V}$ .** Evaluations with  $h^2 = 0.8$  only. AUC<sub>PR</sub> and SRMSD<sub>p</sub> are expected to be the same between WG, Standard, and True or Popkin (within each locus weight type). WG was used as reference since its  $\mathbf{V}$  had the lowest condition numbers (Fig. S18). **A.** The large popkin MOR prediction errors (AUC<sub>PR</sub>, SRMSD<sub>p</sub>, but not  $\sigma^2$ ) in the admixed family simulation are explained by the condition number of  $\mathbf{V}$ . **B.** Smaller errors in 1000 Genomes are not explained by the condition number of  $\mathbf{V}$ .