

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 Materials and 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)

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

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

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

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

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

<sup>147</sup> In Appendix C we prove that

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

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

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

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

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

### <sup>151</sup> 2.3 Association models

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

<sup>153</sup> Yao and Ochoa, 2022):

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

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

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

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

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

<sup>155</sup> 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 simulated from the admixed  
197 family model separately per 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 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 the model of Eqs. (8) and (9), is used to run all LMM associations  
226 (Yang et al., 2011; Yang et al., 2014). We pass  $2\Phi^T$  for all kinship matrices tested (the same scale  
227 as its own kinship estimate). `plink2`, which implements the model of Eqs. (10) and (11), performs  
228 the PCA associations (Chang et al., 2015). We use  $r = K - 1 = 2$  PCs for the admixed family  
229 simulations, and  $r = 10$  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).

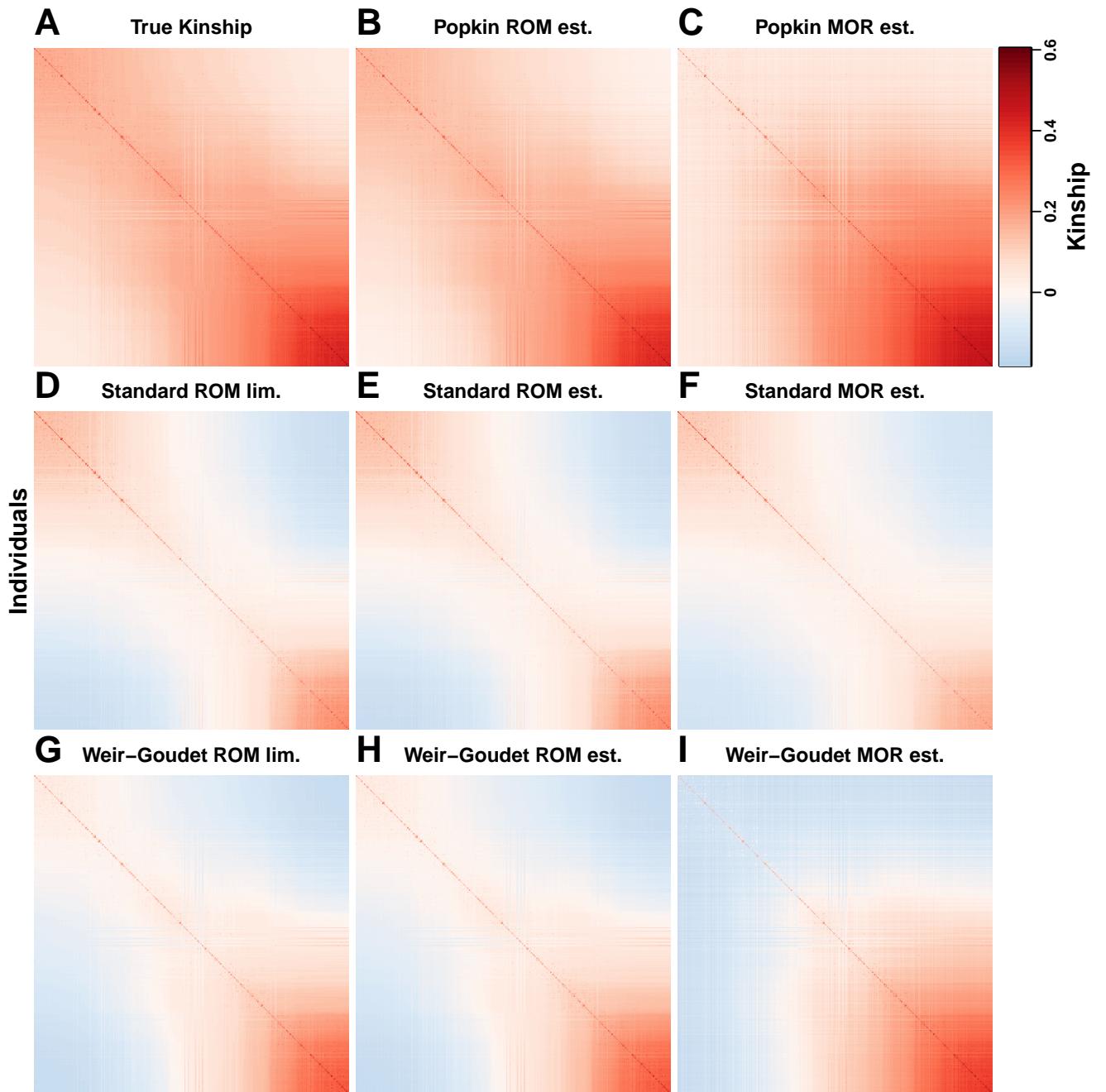


Figure 1: **Kinship estimates and limits on the admixed family simulation.** Each panel shows a kinship matrix as a heatmap, with each of the  $n = 1000$  individuals along both x and y axes, color represents kinship: positive values in red, negative in blue. Diagonal contains inbreeding values. Each 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  
 241 popkin MOR has a slight upward bias that varies across the matrix (Fig. S1A). In contrast, the  
 242 Standard and Weir-Goudet (WG) estimates have large downward biases overall, resulting in abun-  
 243 dant negative values; Standard biases vary for every pair of individuals (as described in Eq. (4);  
 244 Fig. S2), while WG has a uniform bias (following Eq. (6)). The difference is most noticeable near  
 245 the diagonal: the true kinship matrix has monotonically increasing values, WG has smaller values  
 246 but which are still monotonically increasing, and Standard estimates follow a U-shaped pattern  
 247 (decreasing at first, then increasing again in Fig. 1D-F).

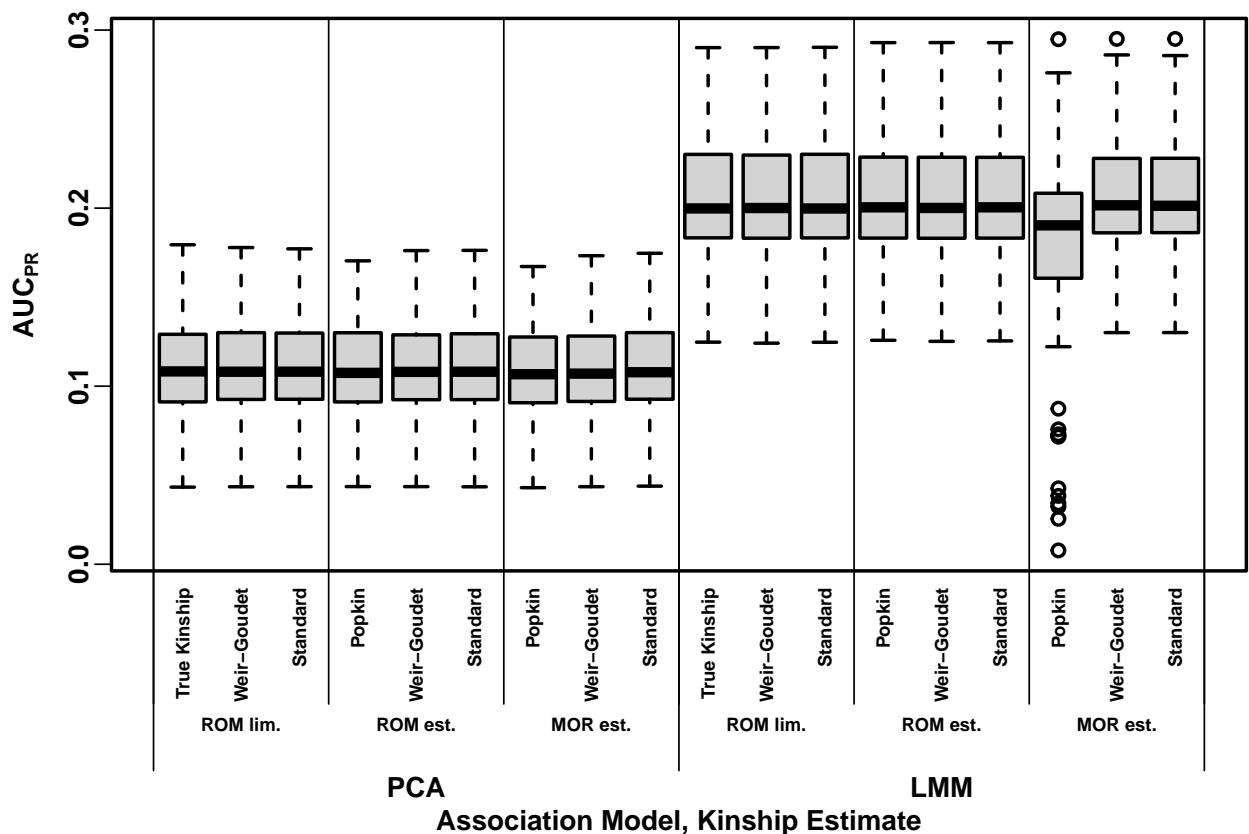


Figure 2: **Distributions of Area Under the Precision-Recall Curve (AUC<sub>PR</sub>) on the admixed family simulation with  $h^2 = 0.8$ .** Higher AUC<sub>PR</sub> is better performance. Results for 100 replicates (each a random genotype matrix and trait vector). Approaches cluster primarily by association 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 S3, respectively) and  $SRMSD_p$  (measures null statistic calibration; Figs. S4 and S5).  
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. S6 and S7). 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 S8). 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

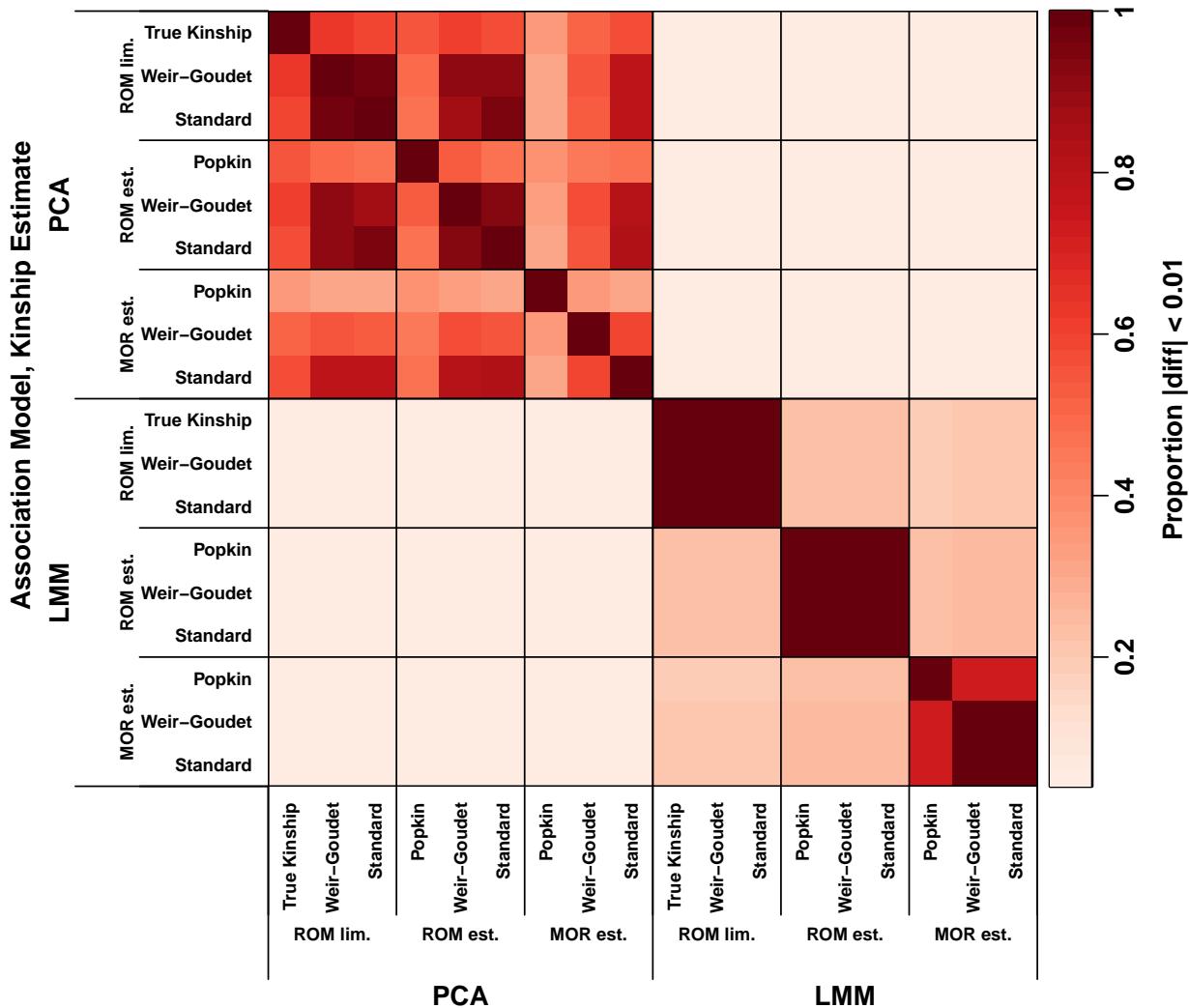


Figure 3: **Agreement between p-values on the admixed family simulation with  $h^2 = 0.8$ .**  
 Calculated agreement (absolute difference under 0.01) averaged over loci (color) of association p-values between association models (LMM vs PCA) and kinship matrices (x and y axes). All 100 replicates are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

(Fig. 4A; trees were explicitly fit to this data in previous work (Yao and Ochoa, 2022)). However, popkin MOR estimates do not follow the nested blocks tree structure, since kinship between African and non-African populations is higher than kinship within African populations (Fig. 4B, Fig. S1). Standard estimates have values closer to zero, and a different bias for each pair of individuals (Fig. S2), resulting in higher relative kinship for African compared to non-African populations (Fig. 4C-D), whereas kinship in African populations is the lowest in the unbiased estimate (Fig. 4A). Lastly, WG estimates are uniformly smaller than popkin's and attain large negative values (Fig. 4E-F).

Our association test conclusion are similar to our simulation study:  $AUC_{PR}$  and  $SRMSD_p$  distributions are nearly identical for estimators of different bias types but same locus weight type (ROM or MOR) and association model. However, unlike the simulation, here for  $h^2 = 0.8$  (but not 0.3) the MOR estimates noticeably outperform ROM estimates (LMM only), in terms of both  $AUC_{PR}$  (Figs. 5 and S9) and  $SRMSD_p$  (Figs. S10 and S11). P-values are even more highly correlated in this case (Figs. S12 and S13), and again nearly identical at a large proportion of loci between approaches with matched association model and locus weight type (MOR or ROM), regardless of bias type (Figs. S14 and S15).

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

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

The most general form we identified of the bias function, mapping a kinship matrix to its bias-

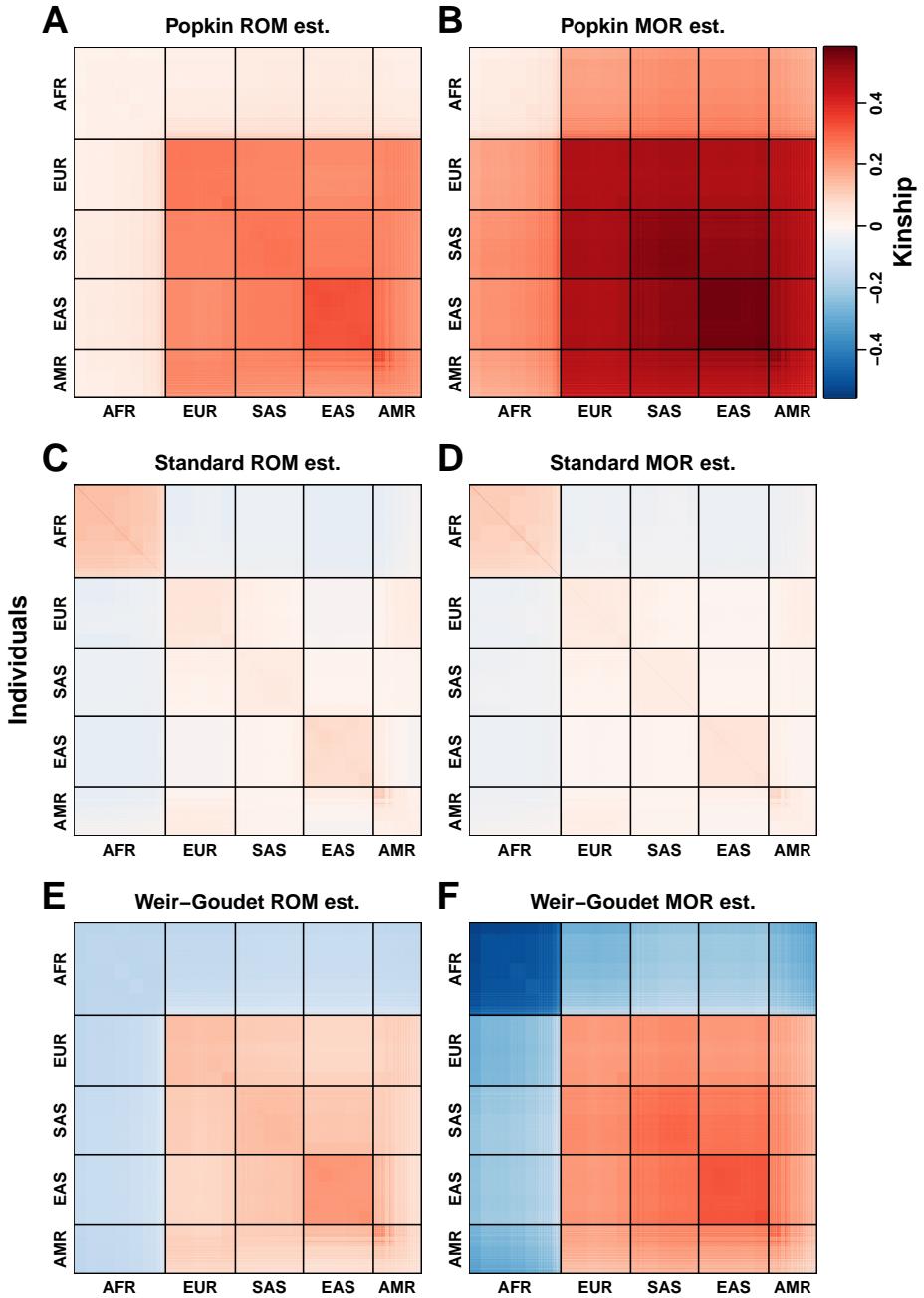


Figure 4: **Kinship estimates on 1000 Genomes.** Each panel represents a kinship matrix as a heatmap, as in Fig. 1. Superpopulation codes: AFR = African, EUR = European, SAS = South Asian, EAS = East Asian, AMR = Admixed Americans (Hispanics). Each estimator bias type (Popkin, Standard, and Weir-Goudet; 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.

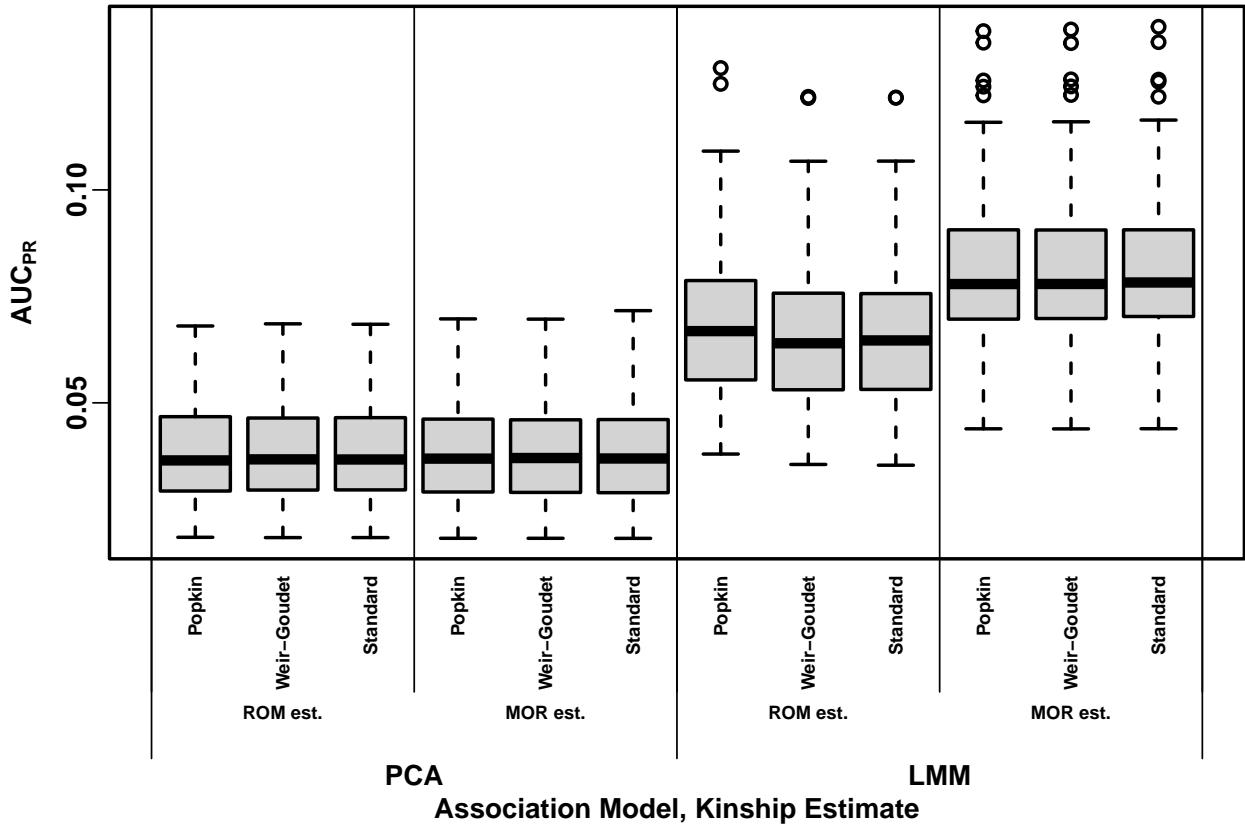


Figure 5: **Distributions of Area Under the Precision-Recall Curve (AUC<sub>PR</sub>) on 1000 Genomes with  $h^2 = 0.8$ .** Higher AUC<sub>PR</sub> is better performance. Results based on 100 simulated trait replicates (real genotype matrix is fixed). Approaches cluster primarily by association model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

301 transformed version, and for which association invariability holds, is

302

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

303 where  $c$  is any positive scalar and  $\mathbf{b}$  is any length- $n$  vector. The key property that the linear operator

304  $\mathbf{B}$  must satisfy is that it shifts the input vector by the same scalar across its values, or

305

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

306 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  
307 only form that results in Eq. (13).

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

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

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

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

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

322 **3.3.1 Proof for LMM case**

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

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

326

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

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

333

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

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

336 The above results require kinship matrices  $\Phi^{T'}$  to be PSD, as covariance matrices are generally  
337 required to be. PSD matrices are characterized by non-negative eigenvalues and determinants.  
338 Nevertheless, for the non-PSD WG bias (has a negative eigenvalue) combined with the generalized  
339 least squares association algorithm, which is used by GCTA and other LMMs (Kang et al., 2008;  
340 Kang et al., 2010; Yang et al., 2014), we find a stronger result consistent with Eq. (17), namely that  
341  $\alpha' = \alpha$ , or in other words,  $\eta = 0$  (Appendix F).

342 The LMM association p-value does not change in several common tests, including the F-test,  
343 since it only depends on the residuals and these do not change, as well as the likelihood ratio test,  
344 because although covariance determinants change, they cancel out in the ratio. The Wald test used  
345 by GCTA (Yang et al., 2014) is also invariant to these kinship biases given our empirical results in  
346 Figs. 3 and S14 and proven explicitly for WG bias in Appendix F. Lastly, using an implementation in

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

352 **3.3.2 Proof for PCA case**

353 We present a proof for the PCA case that relies on an approximation that holds well in practice.  
 354 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  
 355 of  $\Phi^{T'}$ . The key approximation is that

$$356 \quad \mathbf{U}'_r \approx \mathbf{B} \mathbf{U}_r, \quad (18)$$

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

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

361 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  
 362 compensated for by adjusting the intercept using Eq. (17), so the association coefficient  $\beta_i$  and the  
 363 residuals are the same in both cases. This proof works if there are small numbers of zero or negative  
 364 eigenvalues in  $\Phi^{T'}$  (non-PSD cases), as those rank last and are simply ignored. The observations  
 365 from LMMs, that p-values are invariant to bias types, also hold for PCA.

366 We visualize the top PCs of our datasets in Fig. 6 to assess the validity of Eq. (18). The  
 367 approximation is equivalent to each biased PC (Standard or Weir-Goudet) being shifted from the  
 368 unbiased PC (Popkin), as described in Eq. (13). Fig. 6 indeed shows that PC1 is shifted by  
 369 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

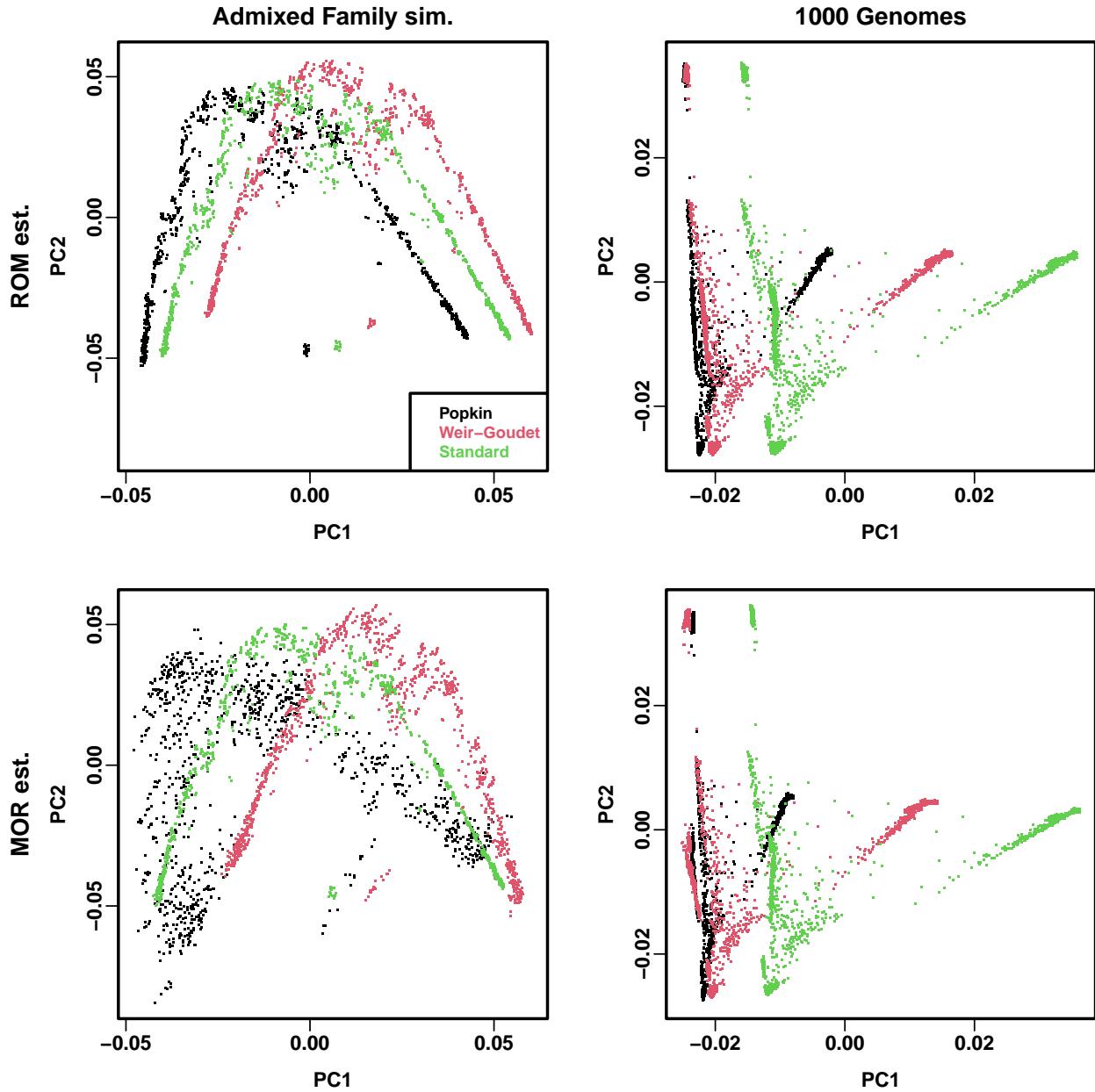


Figure 6: **Visualization of PC shift due to kinship biases.** Each panel shows three estimates (bias types): Popkin, Standard, and Weir-Goudet. ROM estimates are in first row, MOR in second row. (In admixed family, ROM limits are very similar to ROM estimates (not shown).) Columns show estimates from 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).

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

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

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

393 **3.5 Characterization of non-PSD and singular kinship and trait covariance es-  
394 timators**

395 While attempting to validate and characterize the earlier factorization of the WG bias function  
396 (Eqs. (12) to (15)), we discovered that it does not produce PSD matrices, which covariance matrices  
397 are required to be. To characterize this problem more broadly, we calculate the eigenvalues of all  
398 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  
399 which we calculate using GCTA's estimates of  $\sigma^2$  and  $\sigma_\epsilon^2$ .

400 We find that all WG matrices have very large negative minimum eigenvalues, and popkin MOR  
401 estimates also have smaller negative minimum eigenvalues (Figs. S16 and S17). Moreover, besides all  
402 WG matrices and most popkin MOR estimates, Standard matrices are also often non-PSD but only  
403 in 1000 Genomes (Figs. 7 and S18), which has missing genotypes (the admixed family simulation  
404 does not have missing genotypes). Each of these non-PSD matrices only has one negative eigenvalue.  
405 Notably, all popkin ROM estimates are PSD in every evaluation, including under missingness in  
406 1000 Genomes.

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

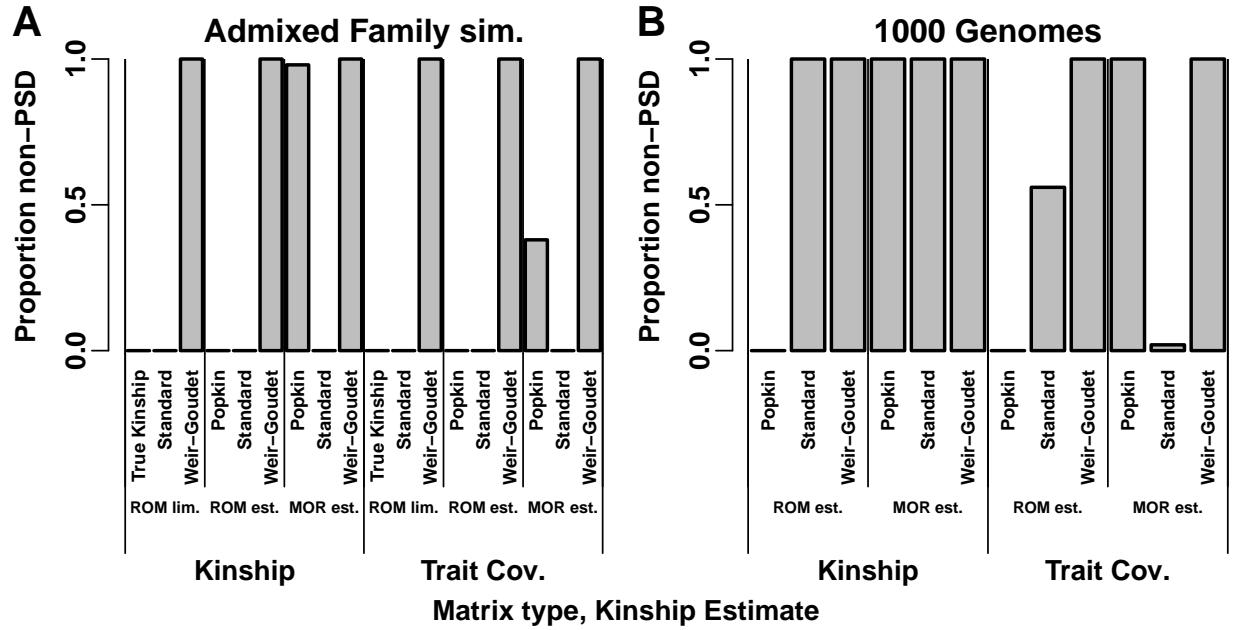


Figure 7: **Proportion of kinship and trait covariance ( $\mathbf{V}$ ) matrices with  $h^2 = 0.8$  that are not positive semidefinite (PSD).** A matrix is non-PSD if it has negative eigenvalues (below  $-10^{-7}$  to allow for limited machine precision). Proportion is calculated over 100 replicates (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  $\mathbf{V}$ .

411 simulation (which lacks missingness), as reflected by extremely high condition numbers, but their  
 412 trait covariances have small condition numbers (Figs. S19 and S20). No other matrices are singular,  
 413 but popkin MOR estimates in the admixed family simulation have relatively high condition numbers  
 414 for kinship, as well as trait covariance for  $h^2 = 0.8$  but not 0.3.

415 To explain these observations, consider the theoretical connection between the eigenvalues of  
 416  $\Phi^T$  and those of  $\mathbf{V}$ . The eigendecomposition trick widely used to fit variance components in LMMs  
 417 (Kang et al., 2008; Lippert et al., 2011; Svishcheva et al., 2012; Zhou and Stephens, 2012; Sul et al.,  
 418 2018) yields

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

419 where  $\mathbf{U}$  and  $\mathbf{\Lambda}$  are the eigenvectors and eigenvalues of  $\Phi^T$ , respectively (Eq. (11)), so the eigen-  
 420 vectors of  $\mathbf{V}$  are also  $\mathbf{U}$  and its eigenvalues are  $2\sigma^2 \mathbf{\Lambda} + \sigma_\epsilon^2 \mathbf{I}$ . Therefore, since  $\sigma^2, \sigma_\epsilon^2 \geq 0$ , then if  
 421  $\Phi^T$  is positive definite (all of its eigenvalues are positive) then so is  $\mathbf{V}$ , and the condition number  
 422 of  $\mathbf{V}$  is always smaller (better) or equal than that of  $\Phi^T$ . A negative kinship eigenvalue  $\lambda_k$  may  
 423 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  
 424 as observed for WG do not become positive in  $\mathbf{V}$ , in fact they can become more negative for high  
 425 heritability (Fig. S16), though they are less negative at lower heritability (Fig. S17).  $\mathbf{V}$  is always in-  
 426 vertible and well-conditioned even when  $\Phi^T$  is singular PSD (has zero eigenvalues), as the Standard  
 427 estimator is under no missingness, since a kinship zero eigenvalue becomes  $\sigma_\epsilon^2$  for  $\mathbf{V}$ . Conversely, the  
 428 above equation explains why some non-PSD kinship matrices are particularly problematic: negative  
 429 eigenvectors near the heritability-dependent value  $-\sigma_\epsilon^2/(2\sigma^2)$  can result in ill-conditioned  $\mathbf{V}$ . We  
 430 see that popkin MOR estimates are non-PSD (Fig. S16) in such a way that some of their  $\mathbf{V}$  are  
 431 ill-conditioned under high heritability (Fig. S19) but not the lower heritability (Fig. S20), and this  
 432 explains its poorer performance in the admixed family evaluations with high heritability (Figs. 2  
 433 and S4), as shown in the next subsection.

### 434 3.6 Further empirical validation of theoretical predictions

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

437 and empirically how broken assumptions affect them. With PCA, no deviations from expectation  
438 of  $AUC_{PR}$  and  $SRMSD_p$  are observed for WG (Figs. 2 and 5, Figs. S4 and S10), which makes sense  
439 since PCA simply ignores eigenvectors with negative or zero eigenvalues. Therefore, our analysis  
440 focuses on LMM and  $h^2 = 0.8$  only, where large deviations are observed, and clarification regarding  
441 WG is needed.

442 LMMs such as GCTA perform association testing in two steps. First is the restricted maximum  
443 likelihood step used to fit variance components. Although the eigendecomposition approaches (Kang  
444 et al., 2008; Lippert et al., 2011; Svishcheva et al., 2012; Zhou and Stephens, 2012; Sul et al., 2018)  
445 require positive definite  $\mathbf{V}$  (lest the determinant of  $\mathbf{V}$  be negative), surprisingly the GCTA average  
446 information algorithm only requires in practice that  $\mathbf{V}$  be invertible (Yang et al., 2011). Thus,  
447 the relationship between WG, Standard, and True or Popkin variance components are largely as  
448 expected from our theoretical prediction  $\sigma^{2\prime} = c\sigma^2$  in Eq. (16), with the exception of popkin ROM  
449 on 1000 Genomes  $h^2 = 0.8$  only, whose genetic variance estimates are slightly smaller than expected  
450 (Figs. S21 and S22).

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

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

465 small performance errors of popkin ROM in 1000 Genomes (Figs. 5 and S10) are driven by errors  
466 in genetic variance component estimation  $\sigma^2$  (Fig. S23). However, the larger performance errors  
467 of popkin MOR in the admixed family simulation (Figs. 2 and S4) are instead explained by the  
468 condition number of  $\mathbf{V}$  (Fig. S24). This result makes sense since the condition number by definition  
469 quantifies regression coefficient estimation accuracy.

## 470 4 Discussion

471 Previous research showed that commonly used kinship estimators are biased, and that these biases  
472 can be large (Ochoa and Storey (2021); Figs. 1 and S2). Our initial hypothesis was that these kinship  
473 biases would affect association testing, but surprisingly found that association is unaffected. We  
474 then proved theoretically that it is the intercept and relatedness effect (random effect or PCs) coef-  
475 ficients that compensate for the bias, and result in identical association coefficients and significance  
476 statistics.

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

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

492 for related sample covariance estimators outside genetics (Jurczak and Rohde, 2017). Fortunately,  
493 non-PSD kinship estimators often perform well for association. Nevertheless, in our admixed family  
494 simulation we did see the other popkin estimator (the MOR version) perform particularly poorly due  
495 to being non-PSD, which in a heritability-dependent manner results in ill-conditioned association  
496 tests and substantial loss of accuracy and power (Figs. 2, S4 and S24). Theory predicts that the same  
497 can happen with any non-PSD estimator, depending on unknowns such as the heritability and the  
498 value of the negative eigenvalues of the kinship estimator, so it is risky to use MOR estimators (all  
499 of which are non-PSD in 1000 Genomes), as well as the WG estimator generally (which is non-PSD  
500 in all replicates of all of our evaluations). We also observe smaller numerical inaccuracies for popkin  
501 ROM, the estimator we recommend, in 1000 Genomes with  $h^2 = 0.8$  only, although the result is  
502 mixed: performance is slightly better (Fig. 5) although null p-value calibration is slightly worse  
503 (Fig. S10). The cause is variance components are poorly estimated (Fig. S21), but we did not find a  
504 more fundamental explanation. Overall, our assessment suggests that the popkin ROM estimator is  
505 the safest choice due to its guarantee of well-conditioned associations that other estimators cannot  
506 make.

507 Despite being non-PSD, we observe better performance for MOR versus ROM estimators in  
508 LMM association of 1000 Genomes with  $h^2 = 0.8$  (Fig. 5; the approaches were tied under lower  
509 heritability in Fig. S9). Perhaps this is expected because we simulated larger coefficients for rare  
510 variants, while MOR estimators upweigh rare variants. This effect is not observed in the admixed  
511 family simulation, where MOR and ROM versions give similar kinship estimates (Fig. 1) and per-  
512 formed similarly (Fig. 2), compared to 1000 Genomes where kinship estimates are also strikingly  
513 different (Fig. 4). However, only popkin ROM is unbiased (Fig. 1B, Fig. S1). One potential expla-  
514 nation is that our kinship model assumes that all variants existed in the MRCA population, whereas  
515 rare variants in human data are known to be more recent mutations, and thus their effective kinship  
516 matrix is different than that of ancestral variants. Therefore, despite its biases, the popkin MOR  
517 estimator may better capture the covariance of rare variants and thus model them better in associa-  
518 tion tests, particularly in LMMs where the effect is most pronounced. Future work should focus on  
519 better approaches for upweighing rare variants or otherwise estimating their covariance structure

520 while resulting in positive definite kinship estimates.

521 Our conclusions that common kinship biases do not affect association studies extend to variations  
522 of the Standard kinship estimator that weigh loci according to linkage disequilibrium (Speed et al.,  
523 2017; Wang et al., 2017), which also have the Standard bias type since this bias is present in each  
524 locus (Ochoa and Storey, 2021). As shown in our theoretical results, another form of the Standard  
525 kinship estimator that weighs individuals to estimate ancestral allele frequencies  $\hat{p}_i^T$ , including the  
526 best unbiased linear estimator in Appendix E (Astle and Balding, 2009; Thornton and McPeek,  
527 2010), is also subject to the same conclusions. Our proof holds for arbitrary kinship matrices and  
528 their biased counterparts, making no assumptions about how they are related to the loci being  
529 tested for association, so they hold whether the test locus was included in the kinship estimate (the  
530 scenario tested empirically) or not, such as in the leave-one-chromosome-out variant of association  
531 testing with LMMs (Lippert et al., 2011; Yang et al., 2014). In addition, the proof does not make  
532 any assumptions about the tested loci, so it does not depend on allele frequency, and thus holds for  
533 rare and common variants.

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

**548 Conflicts of interest**

**549** The authors declare no conflicts of interest.

**550 Funding**

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

**554 Web resources**

**555** plink2, <https://www.cog-genomics.org/plink/2.0/>  
**556** GCTA, <https://yanglab.westlake.edu.cn/software/gcta/>  
**557** bnpsd, <https://cran.r-project.org/package=bnpsd>  
**558** simfam, <https://cran.r-project.org/package=simfam>  
**559** simtrait, <https://cran.r-project.org/package=simtrait>  
**560** popkin, <https://cran.r-project.org/package=popkin>  
**561** popkinsuppl, <https://github.com/OchoaLab/popkinsuppl>

**562 Data availability**

**563** The data and code generated during this study are available on GitHub at <https://github.com/OchoaLab/bias-assoc-paper>. The high-coverage version of the 1000 Genomes Project was down-  
**564** 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_)  
**565** coverage/working/20190425\_NYGC\_GATK/ and detailed processing instructions are found on GitHub  
**566** at <https://github.com/OchoaLab/data>.

568 **References**

- 569 1000 Genomes Project Consortium et al. (2012). “An integrated map of genetic variation from 1,092  
570 human genomes”. *Nature* 491(7422), pp. 56–65. DOI: 10.1038/nature11632.
- 571 Altschul, Stephen F., Raymond J. Carroll, and David J. Lipman (1989). “Weights for data related by  
572 a tree”. *Journal of Molecular Biology* 207(4), pp. 647–653. DOI: 10.1016/0022-2836(89)90234-  
573 9.
- 574 Astle, William and David J. Balding (2009). “Population Structure and Cryptic Relatedness in  
575 Genetic Association Studies”. *Statist. Sci.* 24(4), pp. 451–471. DOI: 10.1214/09-STS307.
- 576 Aulchenko, Yurii S., Dirk-Jan de Koning, and Chris Haley (2007). “Genomewide rapid associa-  
577 tion using mixed model and regression: a fast and simple method for genomewide pedigree-  
578 based quantitative trait loci association analysis”. *Genetics* 177(1), pp. 577–585. DOI: 10.1534/  
579 genetics.107.075614.
- 580 Balding, D. J. and R. A. Nichols (1995). “A method for quantifying differentiation between popula-  
581 tions at multi-allelic loci and its implications for investigating identity and paternity”. *Genetica*  
582 96(1-2), pp. 3–12. DOI: <https://doi.org/10.1007/BF01441146>.
- 583 Bhatia, Gaurav et al. (2013). “Estimating and interpreting FST: the impact of rare variants”.  
584 *Genome Res.* 23(9), pp. 1514–1521. DOI: 10.1101/gr.154831.113.
- 585 Chang, Christopher C. et al. (2015). “Second-generation PLINK: rising to the challenge of larger  
586 and richer datasets”. *GigaScience* 4(1), p. 7. DOI: 10.1186/s13742-015-0047-8.
- 587 Consortium, The 1000 Genomes Project (2010). “A map of human genome variation from population-  
588 scale sequencing”. *Nature* 467(7319), pp. 1061–1073. DOI: 10.1038/nature09534.
- 589 Devlin, B. and Kathryn Roeder (1999). “Genomic Control for Association Studies”. *Biometrics*  
590 55(4), pp. 997–1004. DOI: 10.1111/j.0006-341X.1999.00997.x.
- 591 Emik, L. Otis and Clair E. Terrill (1949). “Systematic procedures for calculating inbreeding coeffi-  
592 cients”. *J Hered* 40(2), pp. 51–55. DOI: 10.1093/oxfordjournals.jhered.a105986.
- 593 Fairley, Susan et al. (2020). “The International Genome Sample Resource (IGSR) collection of  
594 open human genomic variation resources”. *Nucleic Acids Research* 48(D1), pp. D941–D947. DOI:  
595 10.1093/nar/gkz836.

- 596 García-Cortés, Luis Alberto (2015). “A novel recursive algorithm for the calculation of the detailed  
597 identity coefficients”. *Genetics Selection Evolution* 47(1), p. 33. DOI: 10.1186/s12711-015-  
598 0108-6.
- 599 Hoffman, Gabriel E. (2013). “Correcting for population structure and kinship using the linear mixed  
600 model: theory and extensions”. *PLoS ONE* 8(10), e75707. DOI: 10.1371/journal.pone.0075707.
- 601 Jacquard, Albert (1970). *Structures génétiques des populations*. Paris: Masson et Cie.
- 602 Jurczak, Kamil and Angelika Rohde (2017). “Spectral analysis of high-dimensional sample covariance  
603 matrices with missing observations”. *Bernoulli* 23(4A), pp. 2466–2532. DOI: 10.3150/16-BEJ815.
- 604 Kang, Hyun Min et al. (2008). “Efficient control of population structure in model organism associ-  
605 ation mapping”. *Genetics* 178(3), pp. 1709–1723. DOI: 10.1534/genetics.107.080101.
- 606 Kang, Hyun Min et al. (2010). “Variance component model to account for sample structure in  
607 genome-wide association studies”. *Nat. Genet.* 42(4), pp. 348–354. DOI: 10.1038/ng.548.
- 608 Lippert, Christoph et al. (2011). “FaST linear mixed models for genome-wide association studies”.  
609 *Nat. Methods* 8(10), pp. 833–835. DOI: 10.1038/nmeth.1681.
- 610 Loh, Po-Ru et al. (2015). “Efficient Bayesian mixed-model analysis increases association power in  
611 large cohorts”. *Nat. Genet.* 47(3), pp. 284–290. DOI: 10.1038/ng.3190.
- 612 Malécot, Gustave (1948). *Mathématiques de l'hérédité*. Masson et Cie.
- 613 Ochoa, Alejandro and John D. Storey (2021). “Estimating FST and kinship for arbitrary population  
614 structures”. *PLoS Genet* 17(1), e1009241. DOI: 10.1371/journal.pgen.1009241.
- 615 Price, Alkes L. et al. (2006). “Principal components analysis corrects for stratification in genome-  
616 wide association studies”. *Nat. Genet.* 38(8), pp. 904–909. DOI: 10.1038/ng1847.
- 617 Rakovski, Cyril S. and Daniel O. Stram (2009). “A kinship-based modification of the armitage trend  
618 test to address hidden population structure and small differential genotyping errors”. *PLoS ONE*  
619 4(6), e5825. DOI: 10.1371/journal.pone.0005825.
- 620 Sherman, Jack and Winifred J. Morrison (1950). “Adjustment of an Inverse Matrix Corresponding  
621 to a Change in One Element of a Given Matrix”. *The Annals of Mathematical Statistics* 21(1),  
622 pp. 124–127. DOI: 10.1214/aoms/1177729893.

- 623 Speed, Doug and David J. Balding (2015). “Relatedness in the post-genomic era: is it still useful?”
- 624     *Nat. Rev. Genet.* 16(1), pp. 33–44. DOI: 10.1038/nrg3821.
- 625 Speed, Doug et al. (2012). “Improved heritability estimation from genome-wide SNPs”. *Am. J. Hum.*
- 626     *Genet.* 91(6), pp. 1011–1021. DOI: 10.1016/j.ajhg.2012.10.010.
- 627 Speed, Doug et al. (2017). “Reevaluation of SNP heritability in complex human traits”. *Nat Genet*
- 628     49(7), pp. 986–992. DOI: 10.1038/ng.3865.
- 629 Sul, Jae Hoon, Lana S. Martin, and Eleazar Eskin (2018). “Population structure in genetic studies:
- 630     Confounding factors and mixed models”. *PLoS Genet.* 14(12), e1007309. DOI: 10.1371/journal.
- 631 pgen.1007309.
- 632 Svishcheva, Gulnara R. et al. (2012). “Rapid variance components-based method for whole-genome
- 633 association analysis”. *Nat Genet* 44(10), pp. 1166–1170. DOI: 10.1038/ng.2410.
- 634 Thornton, Timothy and Mary Sara McPeek (2010). “ROADTRIPS: case-control association testing
- 635 with partially or completely unknown population and pedigree structure”. *Am. J. Hum. Genet.*
- 636 86(2), pp. 172–184. DOI: 10.1016/j.ajhg.2010.01.001.
- 637 Voight, Benjamin F. and Jonathan K. Pritchard (2005). “Confounding from Cryptic Relatedness
- 638 in Case-Control Association Studies”. *PLOS Genetics* 1(3), e32. DOI: 10.1371/journal.pgen.
- 639 0010032.
- 640 Wang, Bowen, Serge Sverdlov, and Elizabeth Thompson (2017). “Efficient Estimation of Realized
- 641 Kinship from SNP Genotypes”. *Genetics*, genetics.116.197004. DOI: 10.1534/genetics.116.
- 642 197004.
- 643 Weir, Bruce S. and Jérôme Goudet (2017). “A Unified Characterization of Population Structure and
- 644 Relatedness”. *Genetics* 206(4), pp. 2085–2103. DOI: 10.1534/genetics.116.198424.
- 645 Wright, Sewall (1922). “Coefficients of Inbreeding and Relationship”. *The American Naturalist*
- 646 56(645), pp. 330–338.
- 647 — (1949). “The Genetical Structure of Populations”. *Annals of Eugenics* 15(1), pp. 323–354. DOI:
- 648 10.1111/j.1469-1809.1949.tb02451.x.

- 649 Xie, C., D. D. Gessler, and S. Xu (1998). “Combining different line crosses for mapping quantitative  
 650 trait loci using the identical by descent-based variance component method”. *Genetics* 149(2),  
 651 pp. 1139–1146.
- 652 Yang, Jian et al. (2010). “Common SNPs explain a large proportion of the heritability for human  
 653 height”. *Nat. Genet.* 42(7), pp. 565–569. DOI: 10.1038/ng.608.
- 654 Yang, Jian et al. (2011). “GCTA: a tool for genome-wide complex trait analysis”. *Am. J. Hum.  
 655 Genet.* 88(1), pp. 76–82. DOI: 10.1016/j.ajhg.2010.11.011.
- 656 Yang, Jian et al. (2014). “Advantages and pitfalls in the application of mixed-model association  
 657 methods”. *Nat Genet* 46(2), pp. 100–106. DOI: 10.1038/ng.2876.
- 658 Yao, Yiqi and Alejandro Ochoa (2022). *Limitations of principal components in quantitative genetic  
 659 association models for human studies*. Tech. rep. bioRxiv, p. 2022.03.25.485885. DOI: 10.1101/  
 660 2022.03.25.485885.
- 661 Yu, Jianming et al. (2006). “A unified mixed-model method for association mapping that accounts  
 662 for multiple levels of relatedness”. *Nat. Genet.* 38(2), pp. 203–208. DOI: 10.1038/ng1702.
- 663 Zhou, Xiang and Matthew Stephens (2012). “Genome-wide efficient mixed-model analysis for asso-  
 664 ciation studies”. *Nat. Genet.* 44(7), pp. 821–824. DOI: 10.1038/ng.2310.

## 665 Appendices

### 666 A Justification for popkin generalizations

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

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

670 then for fixed  $w_i$  we get

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

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

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

674 as desired.

675 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  
 676 random variable (it is a function of the genotypes). The term of interest  $w_i((x_{ij} - 1)(x_{ik} - 1) - 1)$   
 677 is a ratio of random variables whose expectation does not have a closed form. In this case, we rely  
 678 on the first-order approximation to this expectation, namely

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

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

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

681 The same strategy as before applies to estimate the unknown factor  $4/(1 - \bar{\varphi}^T)$ , namely that if the

682 minimum kinship is zero then  $\hat{A}_{\min} \approx -4/(1 - \bar{\varphi}^T)$ , resulting in

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

## 683 B Connection between popkin and standard kinship estimator

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

$$\begin{aligned} A_{ijk} &= I_{ij}I_{ik}((x_{ij} - 1)(x_{ik} - 1) - 1), \\ A_{jk} &= \frac{1}{m_{jk}} \sum_{i=1}^m w_i A_{ijk}, \\ m_{jk} &= \sum_{i=1}^m I_{ij}I_{ik}, \end{aligned}$$

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

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

691 which averages over individuals rather than loci, so its denominator is the number of non-missing  
 692 individuals at this locus. Let us compute some averages of the popkin estimator. Since the result

693 we want holds at every locus separately, let us formulate the averages of interest at locus  $i$  only:

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

694 Therefore, the combination of interest is:

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

695 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  
696 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  
697 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  
698 since  $I_{jk} = 0$  does not solve this equation (because  $n_i > 0$ ), then  $I_{jk} = 1$ , which requires  $n_i = n$ ,  
699 so no individuals can have missing data at this locus (the rest of the terms vanish when this is so).

700 Thus,

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

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

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

702 also requires  $n_i = n$ .

703 Assuming now no missingness, transforming the popkin estimates using the Standard bias func-

704 tion of Eq. (4) gives

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

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

## 707 C Mean kinship inequalities

708 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

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

709 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.

710 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  
711 matrix rather than a coancestry matrix as done originally, as the proof required solely a covariance  
712 matrix with values between 0 and 1 (Ochoa and Storey, 2021).  $\tilde{\varphi}^T$  is defined in Eq. (7).  $0 \leq \tilde{\varphi}^T$   
713 follows since every kinship value is non-negative.  $\bar{\varphi}^T$  and  $\tilde{\varphi}^T$  are related by

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

715 Applying  $\bar{\varphi}^T \leq \bar{\delta}^T$  to Eq. (19) and simplifying yields  $\tilde{\varphi}^T \leq \bar{\delta}^T$ . Lastly, since  $\bar{\varphi}^T - \tilde{\varphi}^T = (\bar{\delta}^T - \tilde{\varphi}^T)/n$   
716 (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  
717 if and only if all  $\varphi_{jk}^T = \bar{\delta}^T$ , since that is necessary and sufficient for  $\bar{\varphi}^T = \bar{\delta}^T$ .

## 718 D Derivation of WG bias factorization

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

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

721 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  
 722 obtain

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

723 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  
 724 also be a constant matrix, which is only achieved if  $\Phi^T\mathbf{b} \propto \mathbf{1}$ . We choose the scaling factor for the  
 725 last  $\mathbf{1}$  to be  $q(\mathbf{1}^\top(\Phi^T)^{-1}\mathbf{1})^{-1}$  as this simplifies notation later, and solving for  $\mathbf{b}$  results in Eq. (14).  
 726 To solve for  $q$ , we replace  $\mathbf{b}$  from Eq. (14) into the above equation, which after rearranging results  
 727 in

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

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

## 729 E Minimum weighted mean kinship

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

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

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

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

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

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

## 739 F Proof that WG bias results in zero intercept shift under LMM 740 generalized least squares estimation

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

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

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

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

746 Now suppose  $\mathbf{V}$  corresponds to some kinship matrix  $\Phi^T$  while  $\mathbf{V}'$  corresponds to  $\Phi^{T'} = F^{\text{WG}}(\Phi^T)$ ,  
 747 and  $\mathbf{V}'$  is also invertible. Our strategy involves repeated application of the Sherman-Morrison for-  
 748 mula for calculating inverses of matrices after a rank-1 update, which for a symmetric update of a

<sup>749</sup> matrix  $\mathbf{A}$  with a vector  $\mathbf{z}$  and a scalar  $b$  takes the form (Sherman and Morrison, 1950)

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

<sup>750</sup> 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  $\mathbf{V}$ :

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

<sup>751</sup> where  $d = 2\sigma^2 \tilde{\varphi}^T$  and we used  $\sigma^{2\prime} = (1 - \tilde{\varphi}^T) \sigma^2$ . Therefore,

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

<sup>752</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>753</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>754</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>755</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>756</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>757</sup> where  $\mathbf{0}$  is a vector the length of the number of covariates minus one (exclude the intercept). As a

758 consequence,  $\mathbf{Z}\mathbf{v} = \mathbf{1}$ , so  $\mathbf{u}^\top \mathbf{v} = \mathbf{1}^\top \mathbf{V}^{-1} \mathbf{1}$  and

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

759 The final step yields the coefficient estimates as a rank-1 update:

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

760 The last factor above vanishes:

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

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

764 Furthermore, since the diagonal values of  $\left( \mathbf{Z}^\top (\mathbf{V}')^{-1} \mathbf{Z} \right)^{-1}$ , which correspond to  $\text{Var}(\hat{\beta}'_k)$  for each  
765  $k$ , 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  
766 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,  
767 are also the same for  $k \neq 1$  for WG bias as for the unbiased kinship matrix.

## Supplemental figures

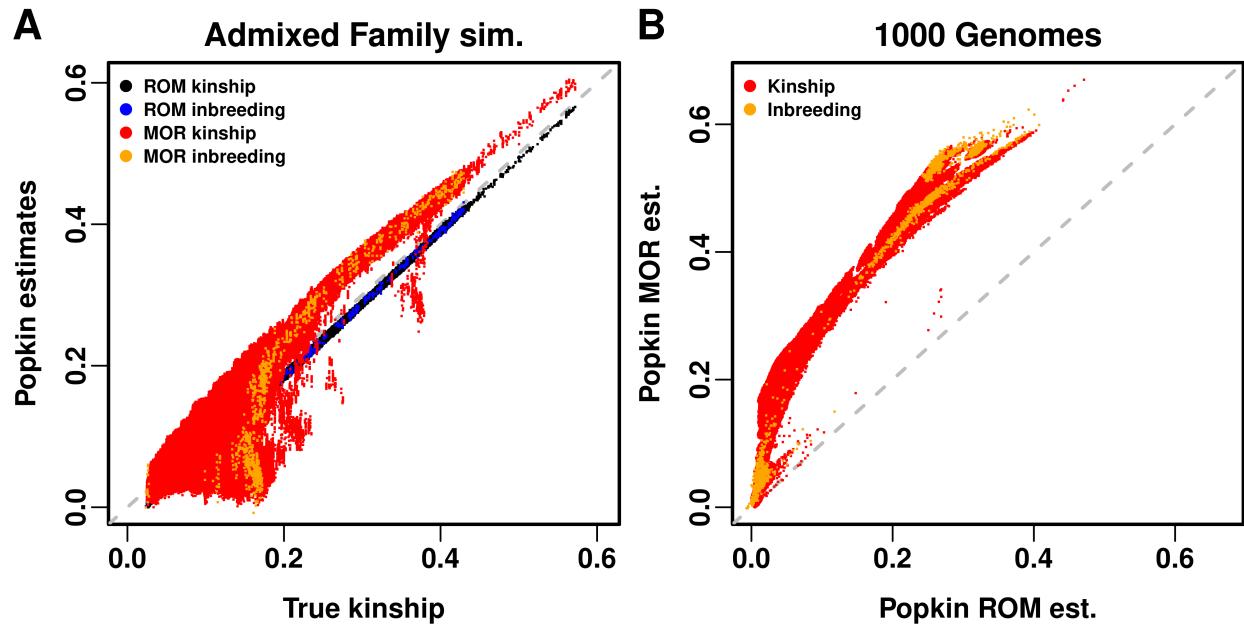


Figure S1: **Comparison of popkin ROM and MOR estimates.** Kinship (off-diagonal of matrix) and inbreeding (transformed diagonal) are plotted in different colors, which shows that their biases (if any) overlap. **A.** In admixed family simulation, both estimates are compared against true kinship. Popkin 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.

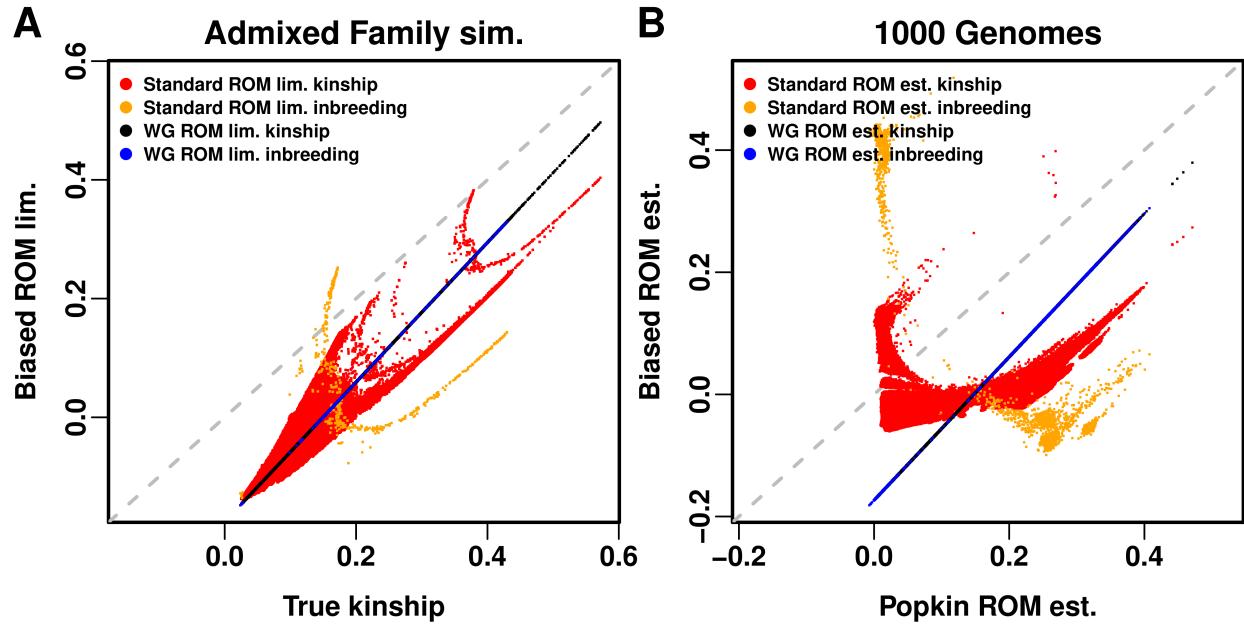


Figure S2: **Comparison of biased kinship estimates.** Kinship (off-diagonal of matrix) and inbreeding (transformed diagonal) are plotted in different colors, to show their biases separately: for WG they are the same, whereas for Standard estimates they differ. In both cases WG is uniformly downwardly biased, whereas Standard biases vary for different pairs of individuals, and are even somewhat negatively correlated with true or unbiased estimates. **A.** In admixed family simulation, both estimates are compared against true kinship. **B.** In 1000 Genomes, since true kinship is unknown, popkin ROM takes its place.

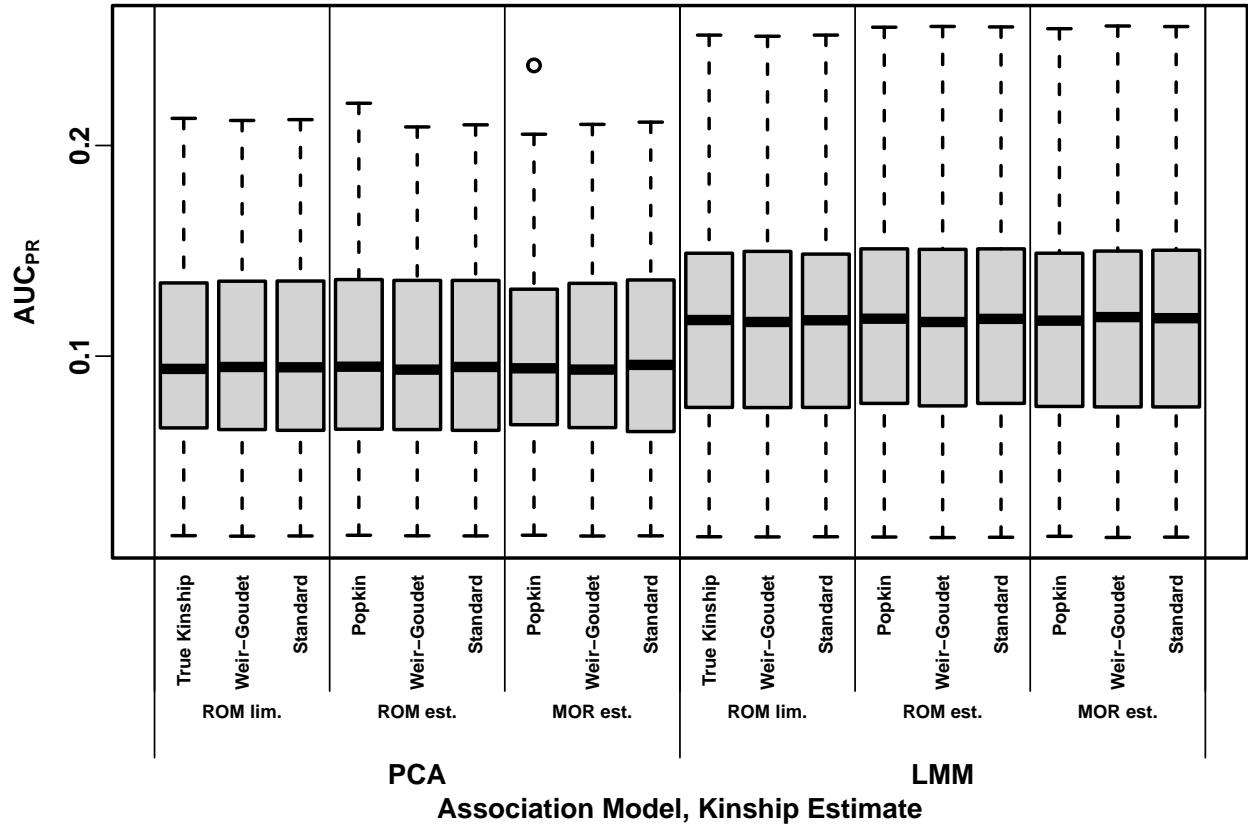


Figure S3: Distributions of Area Under the Precision-Recall Curve (AUC<sub>PR</sub>) on the admixed family simulation with  $h^2 = 0.3$ . Like Fig. 2 except simulated with lower heritability.

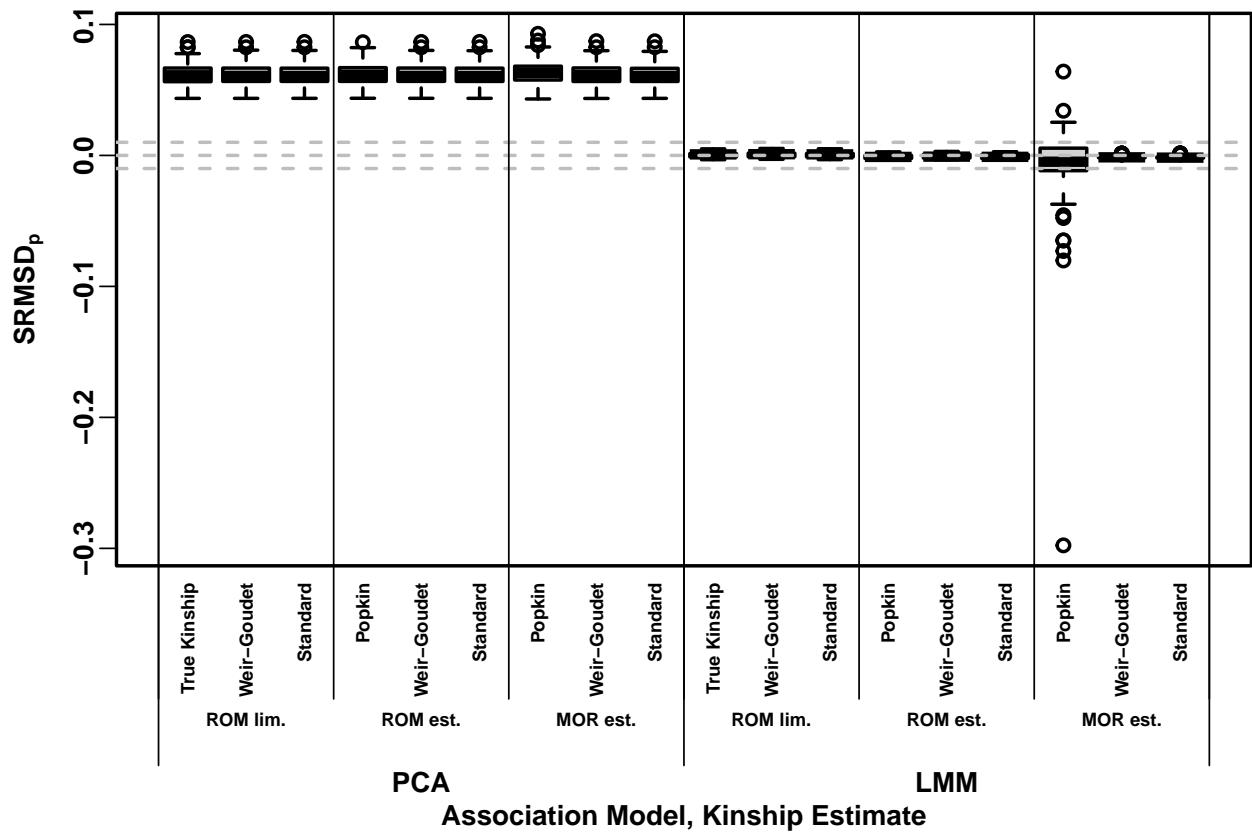


Figure S4: Signed Root Mean Square Deviation of null p-values ( $\text{SRMSD}_p$ ) on the admixed family simulation with  $h^2 = 0.8$ . Same methods and simulation as Fig. 2, see that for more information.  $|\text{SRMSD}_p| < 0.01$  (area between gray dashed lines) is considered calibrated. All PCA runs are miscalibrated by similar amounts, whereas most LMM runs are calibrated with few exceptions.

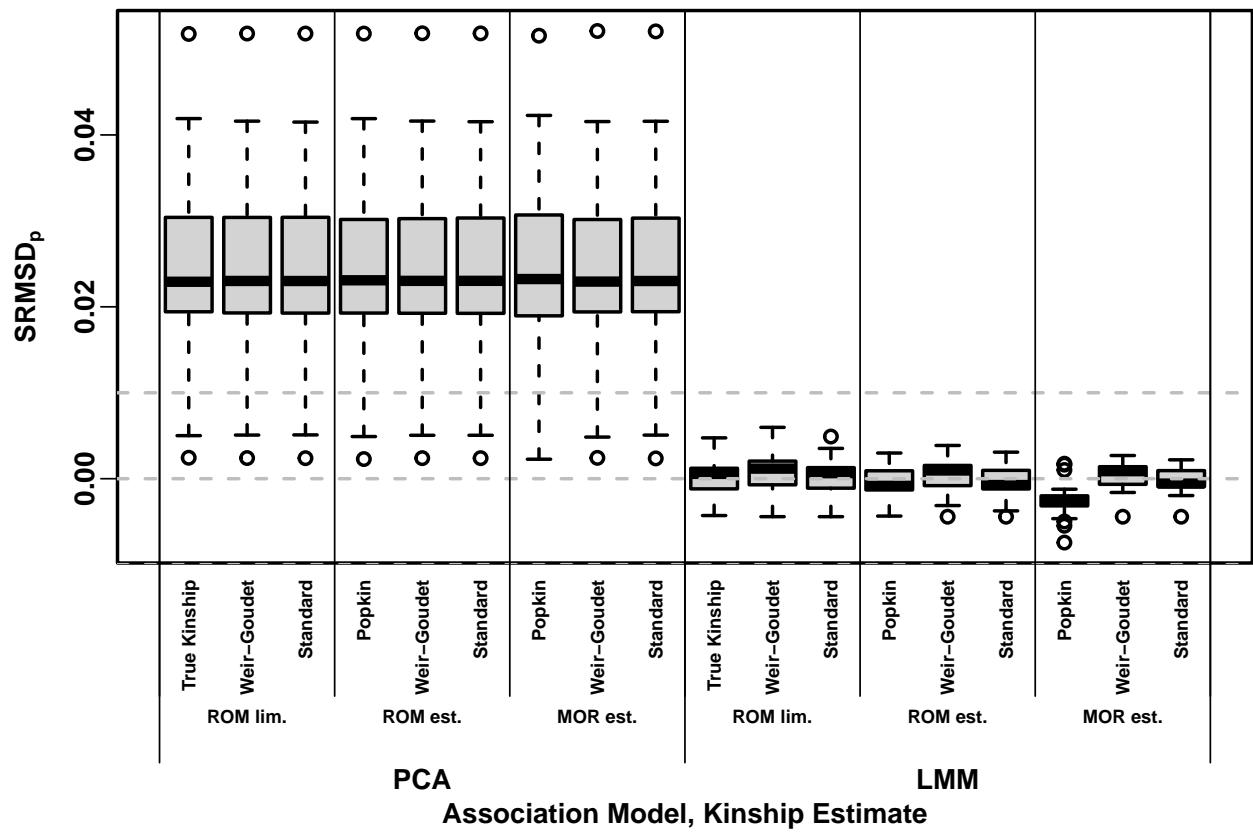


Figure S5: Signed Root Mean Square Deviation of null p-values ( $\text{SRMSD}_p$ ) on the admixed family simulation with  $h^2 = 0.3$ . Like Fig. S4 except simulated with lower heritability.

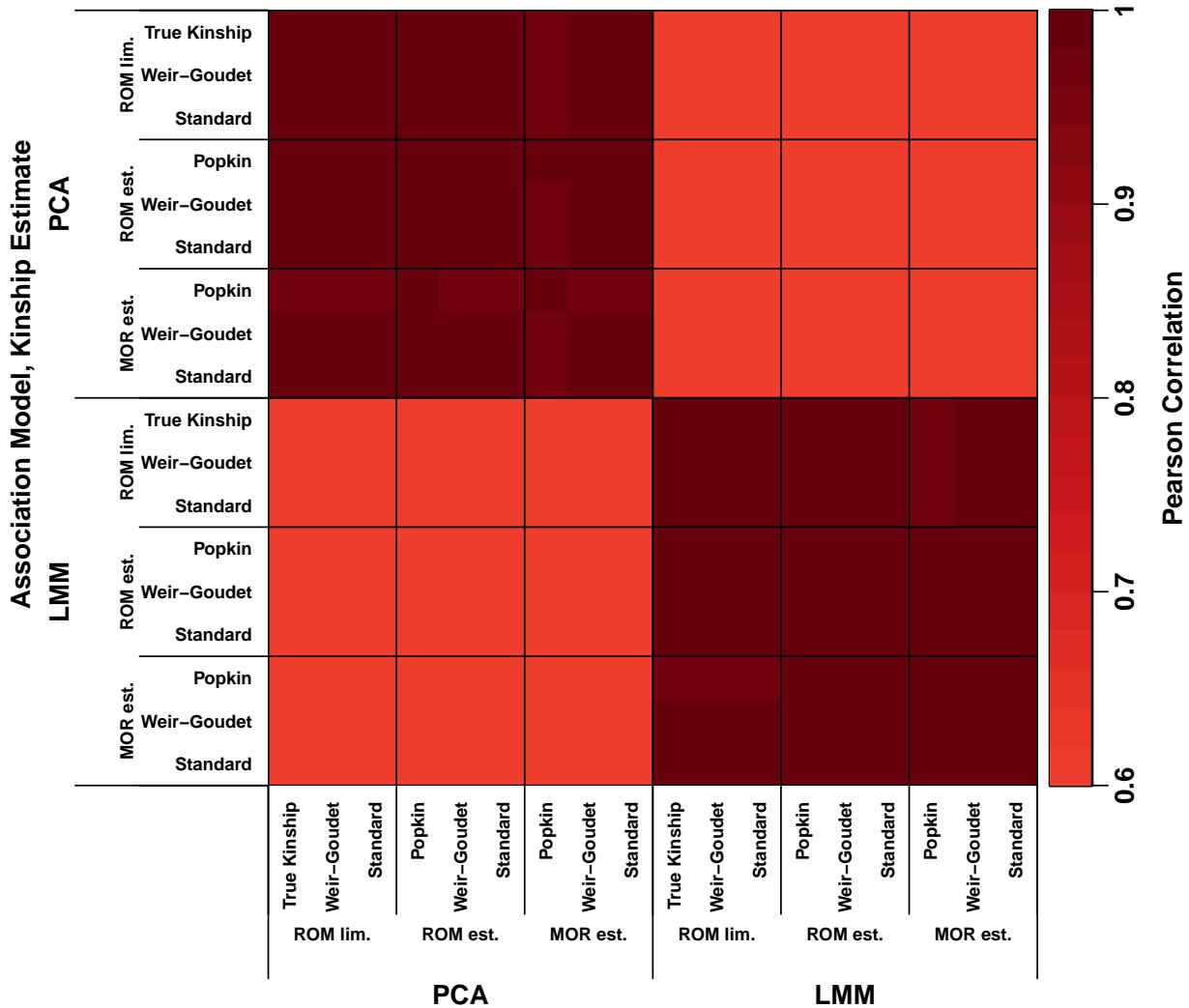


Figure S6: Correlation between p-values on the admixed family simulation with  $h^2 = 0.8$ . Like Fig. 3 but measuring correlation instead of agreement.

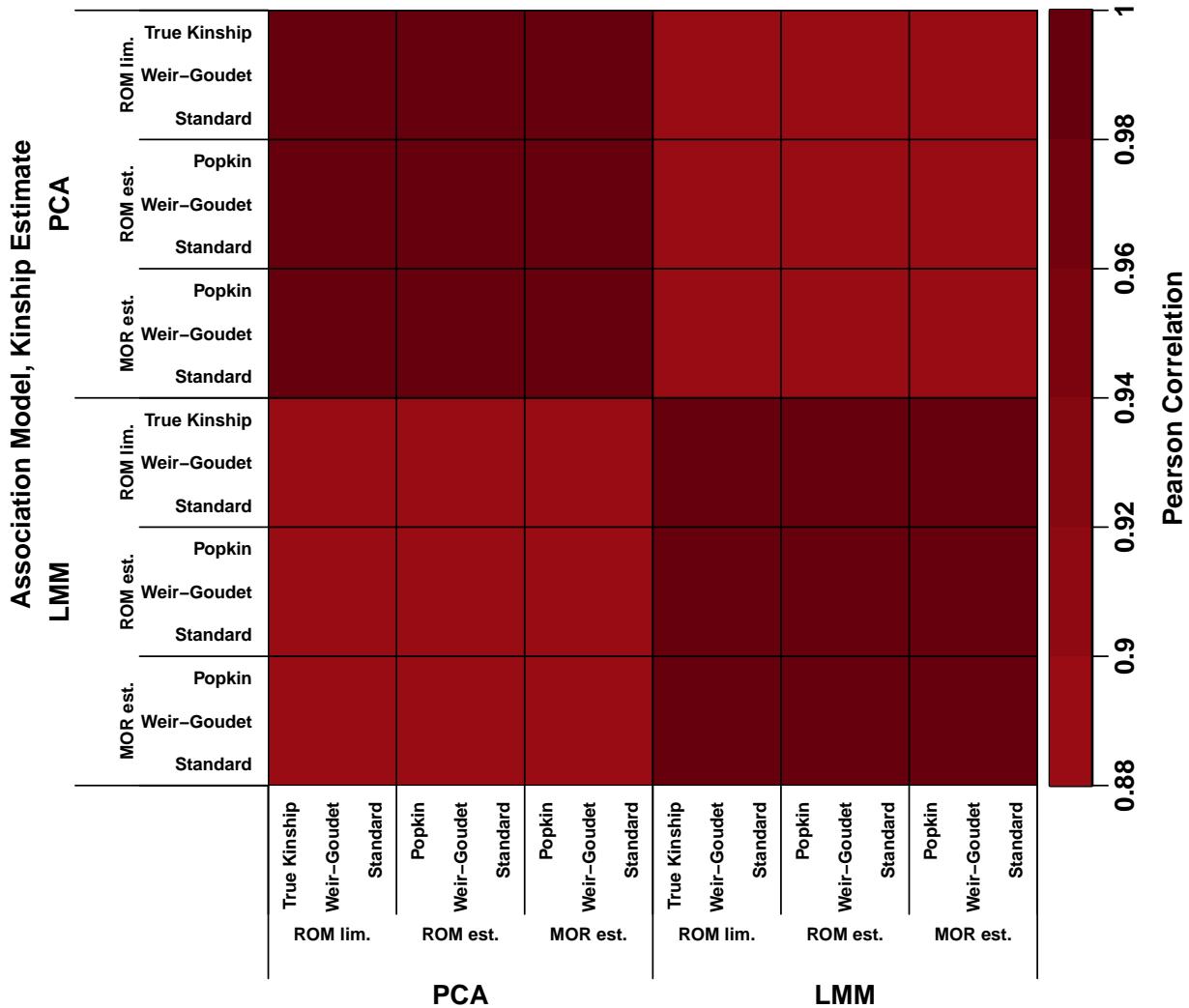


Figure S7: Correlation between p-values on the admixed family simulation with  $h^2 = 0.3$ . Like Fig. S6 except simulated with lower heritability.

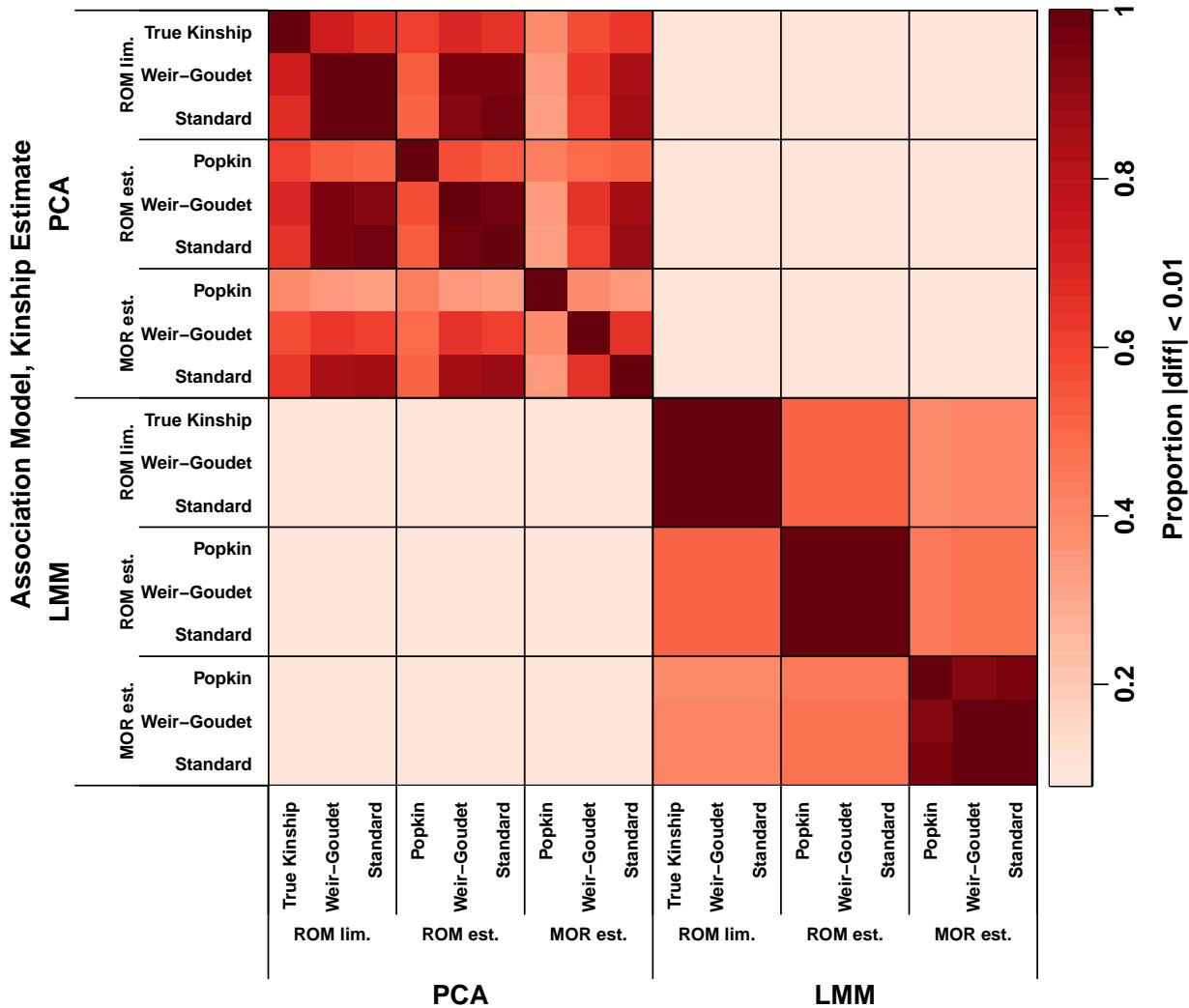


Figure S8: **Agreement between p-values on the admixed family simulation with  $h^2 = 0.3$ .**  
Like Fig. 3 except simulated with lower heritability.

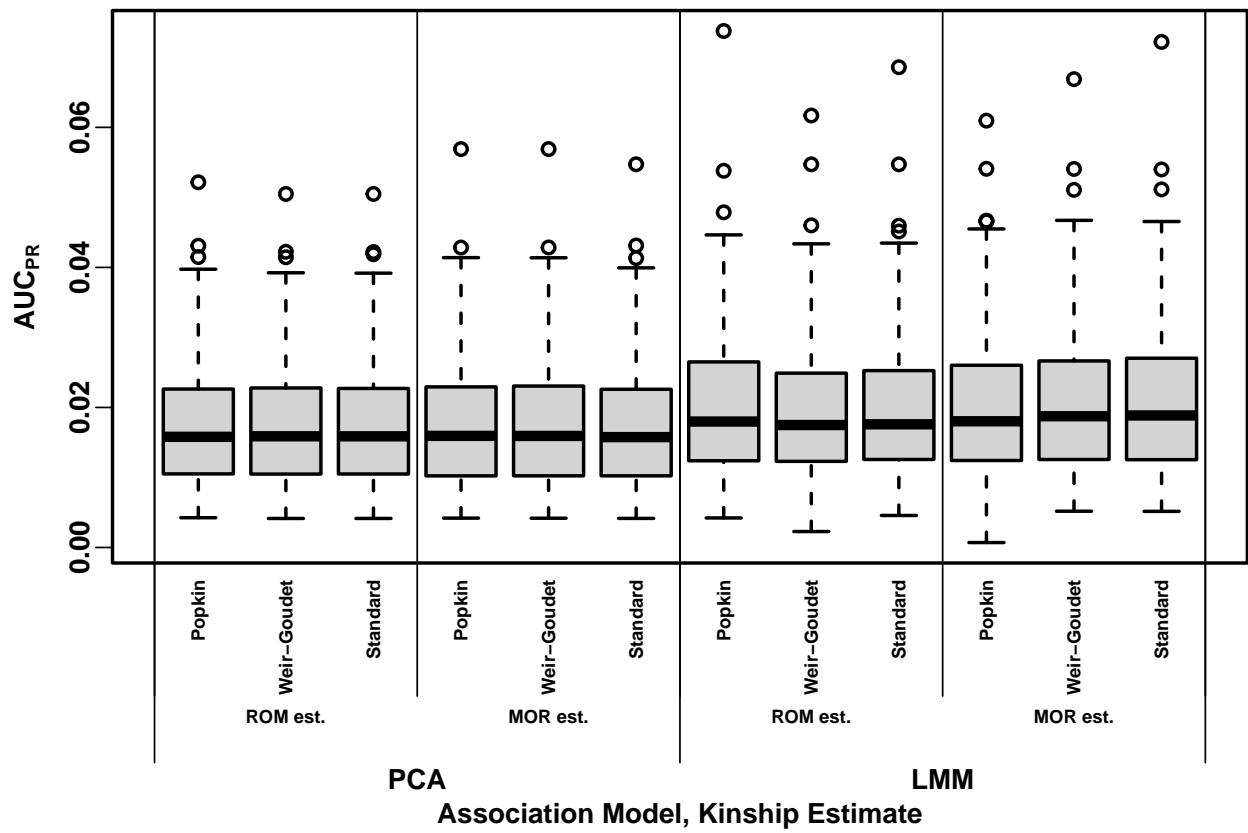


Figure S9: Distributions of Area Under the Precision-Recall Curve ( $AUC_{PR}$ ) on 1000 Genomes with  $h^2 = 0.3$ . Like Fig. 5 except simulated with lower heritability.

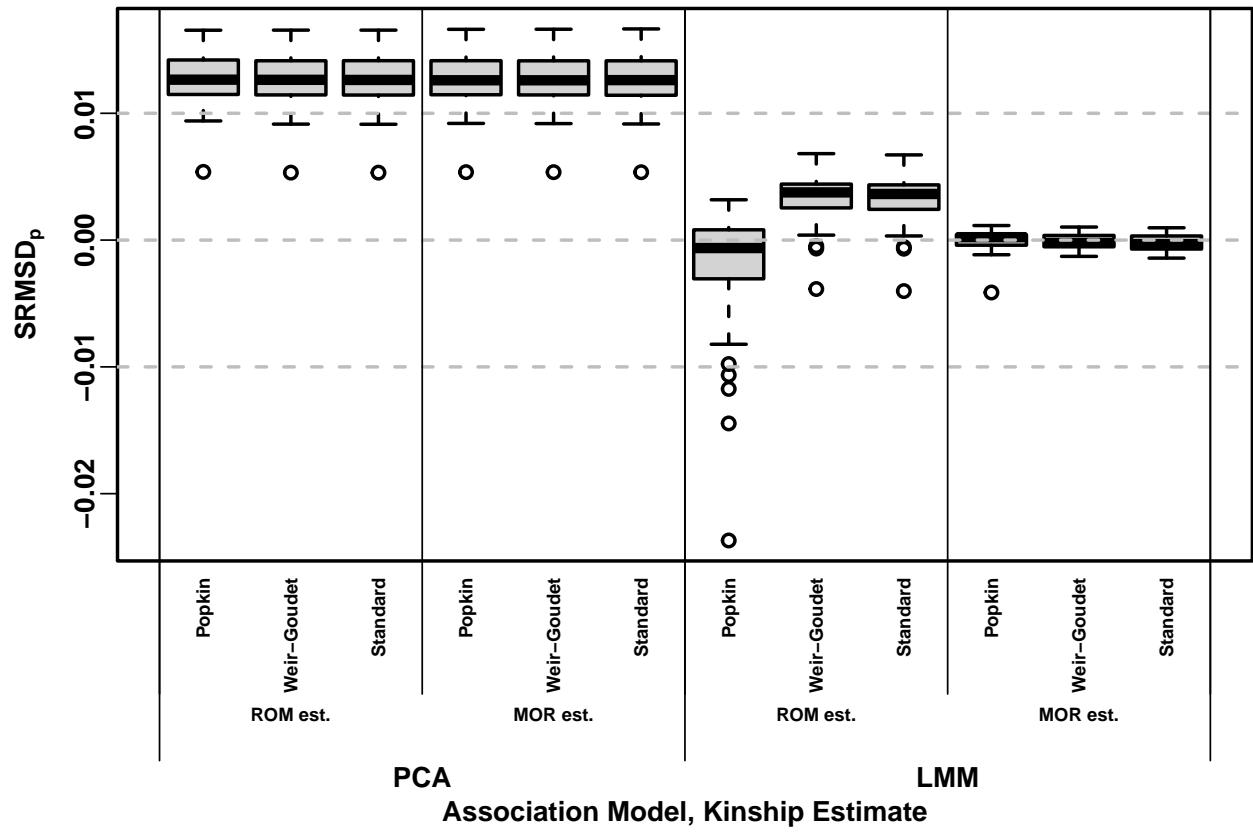


Figure S10: Signed Root Mean Square Deviation of null p-values ( $\text{SRMSD}_p$ ) on 1000 Genomes with  $h^2 = 0.8$ . Same methods and simulation as Fig. 5, and y-axis statistic and conclusions of Fig. S4, see those for more information.

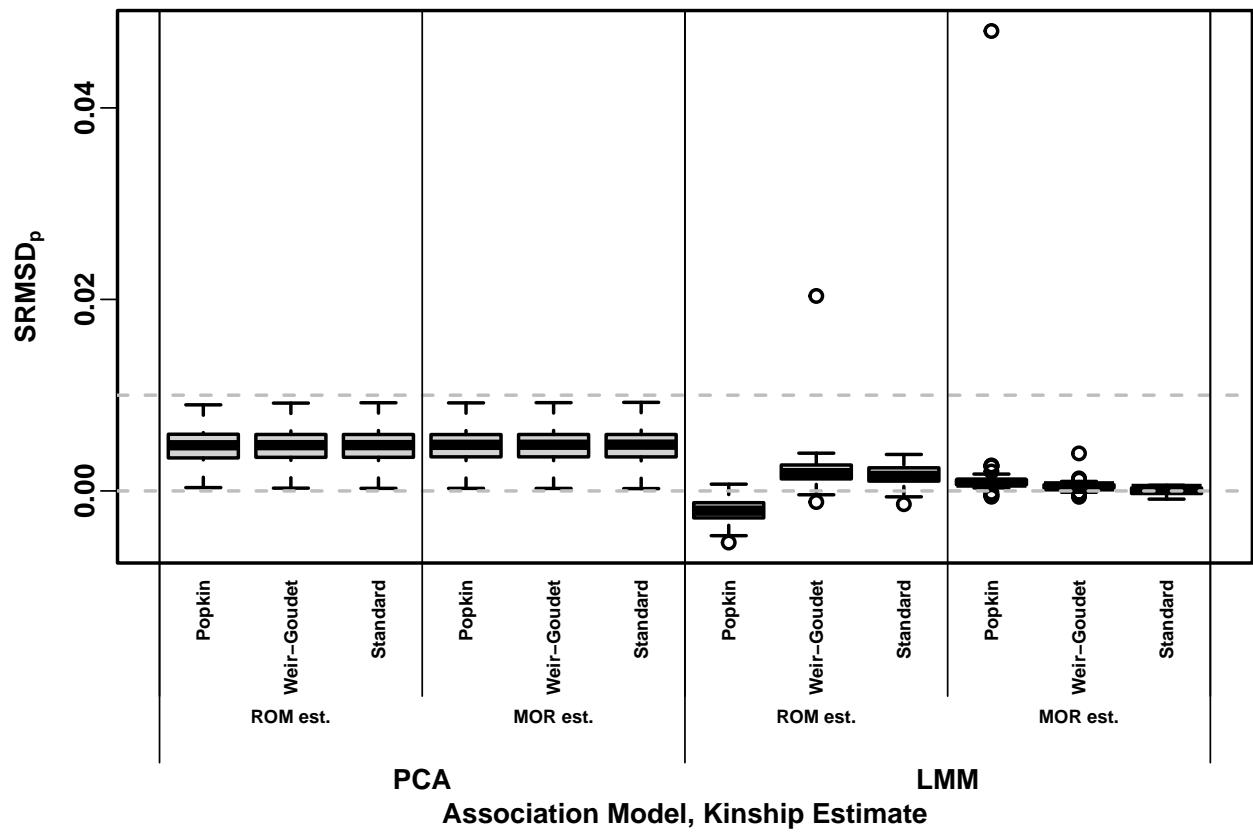


Figure S11: Signed Root Mean Square Deviation of null p-values ( $\text{SRMSD}_p$ ) on 1000 Genomes with  $h^2 = 0.3$ . Like Fig. S10 except simulated with lower heritability.

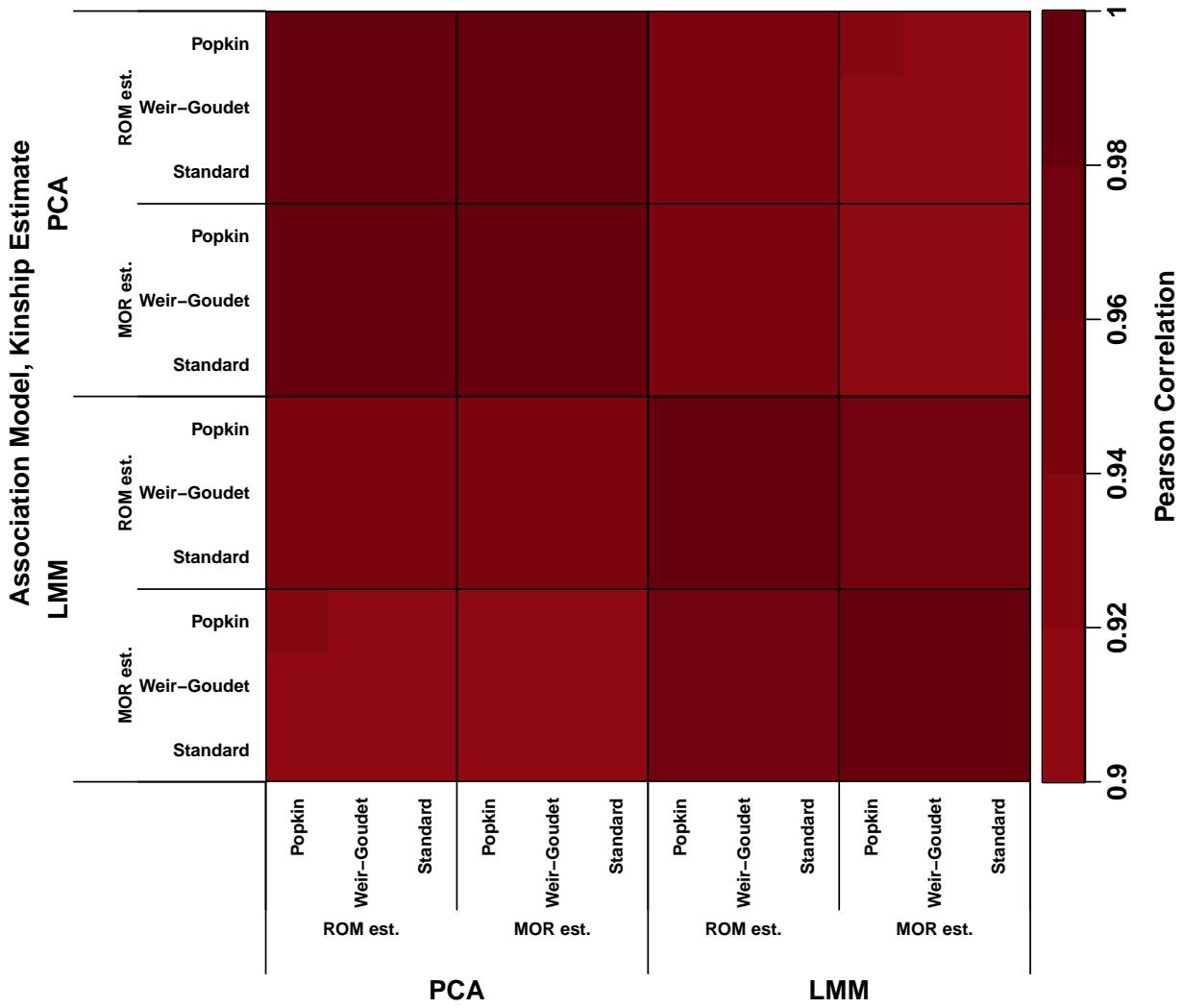


Figure S12: Correlation between p-values on 1000 Genomes with  $h^2 = 0.8$ . See Fig. S6 for more details.

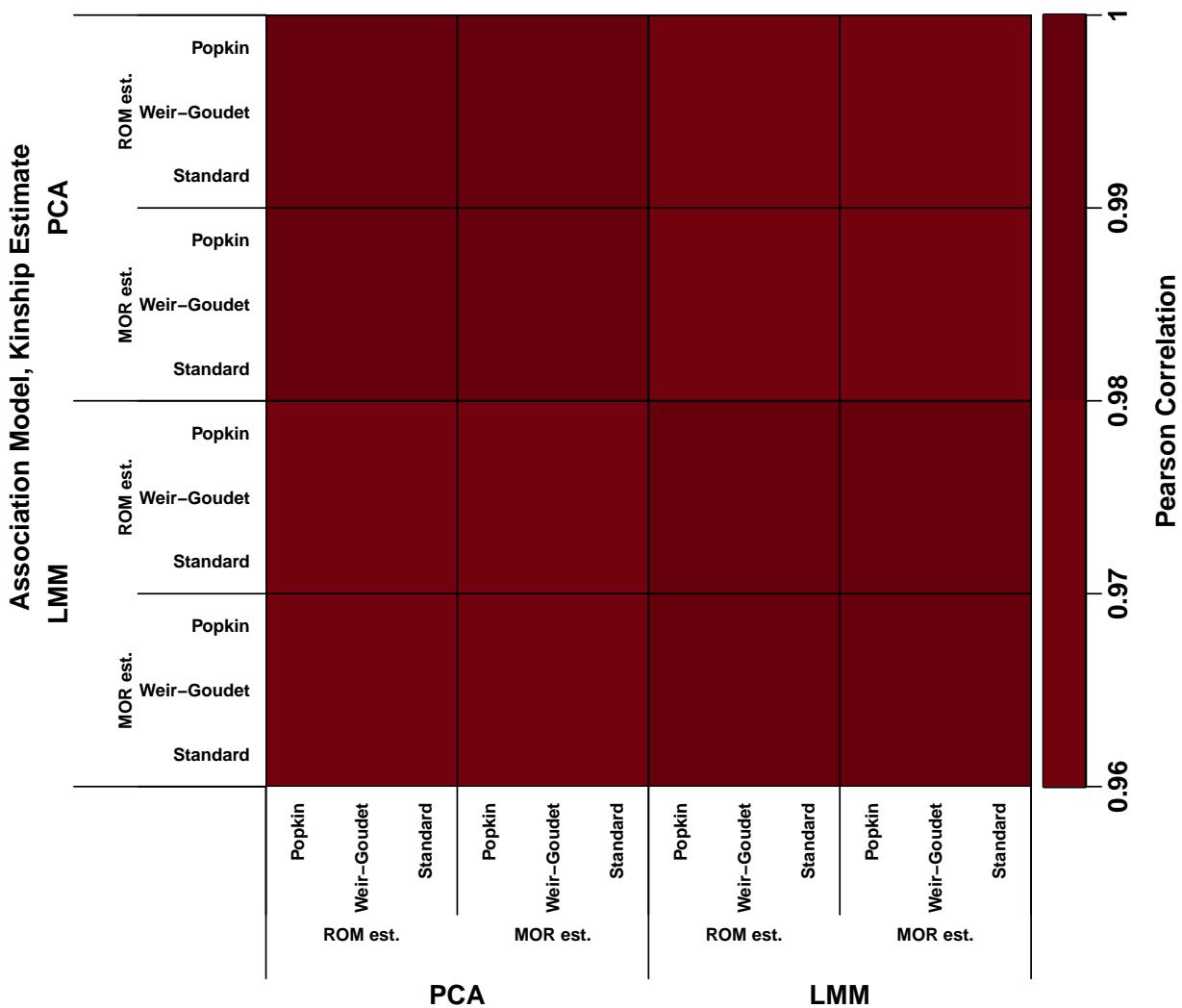


Figure S13: Correlation between p-values on 1000 Genomes with  $h^2 = 0.3$ . Like Fig. S12 except simulated with lower heritability.

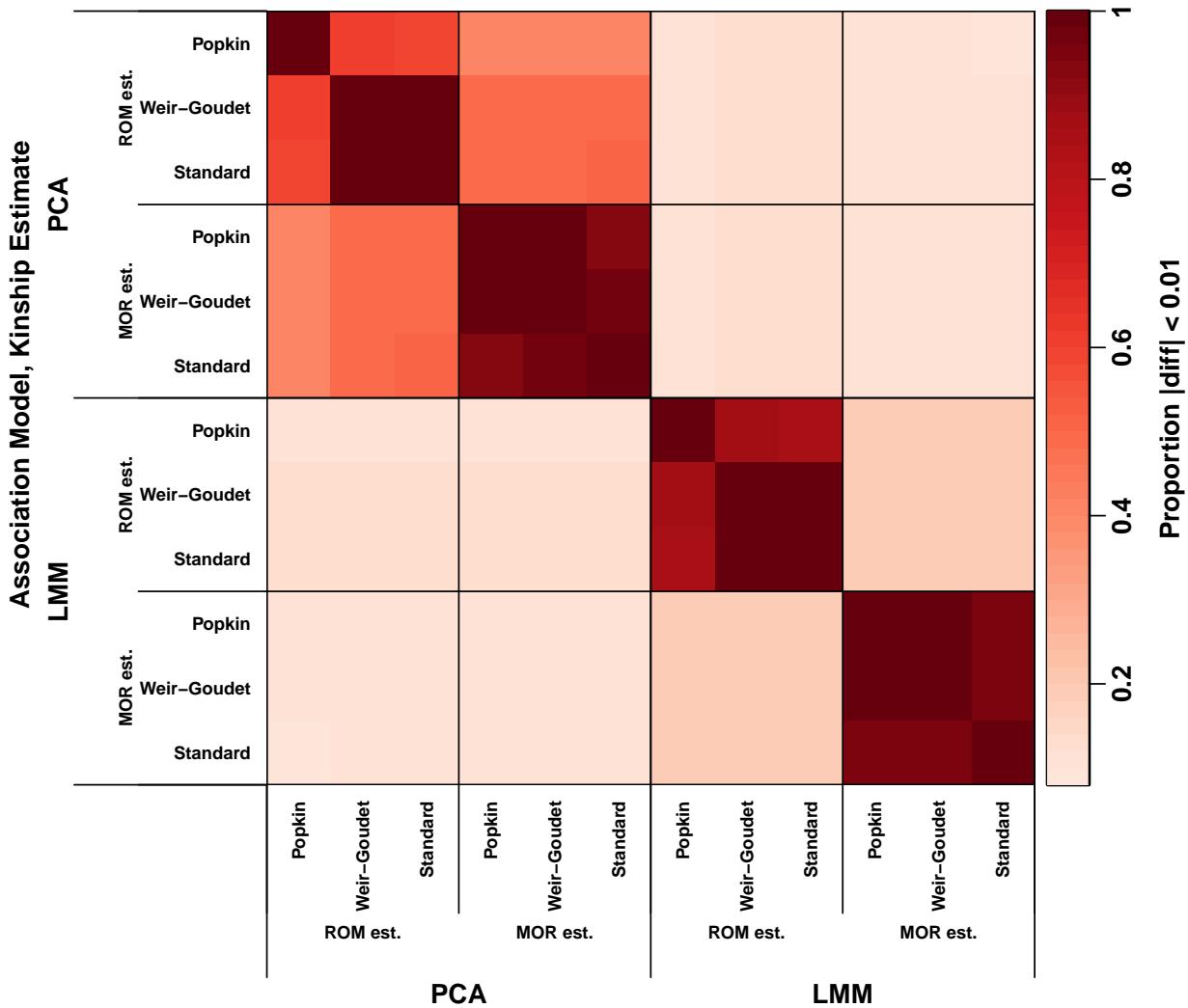


Figure S14: Agreement between p-values on 1000 Genomes with  $h^2 = 0.8$ . See Fig. 3 for more details.

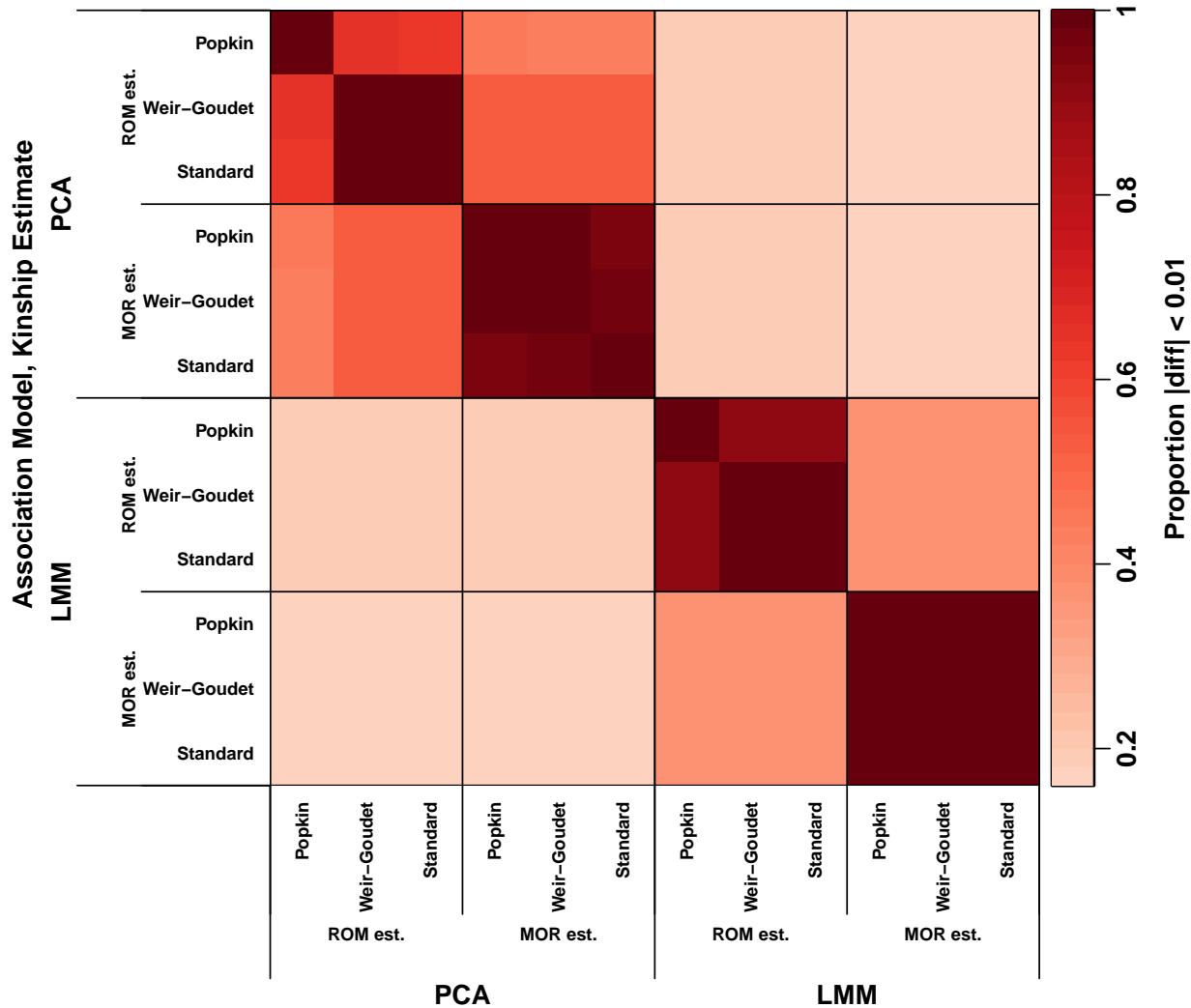


Figure S15: **Agreement between p-values on 1000 Genomes with  $h^2 = 0.3$ .** Like Fig. S14 except simulated with lower heritability.

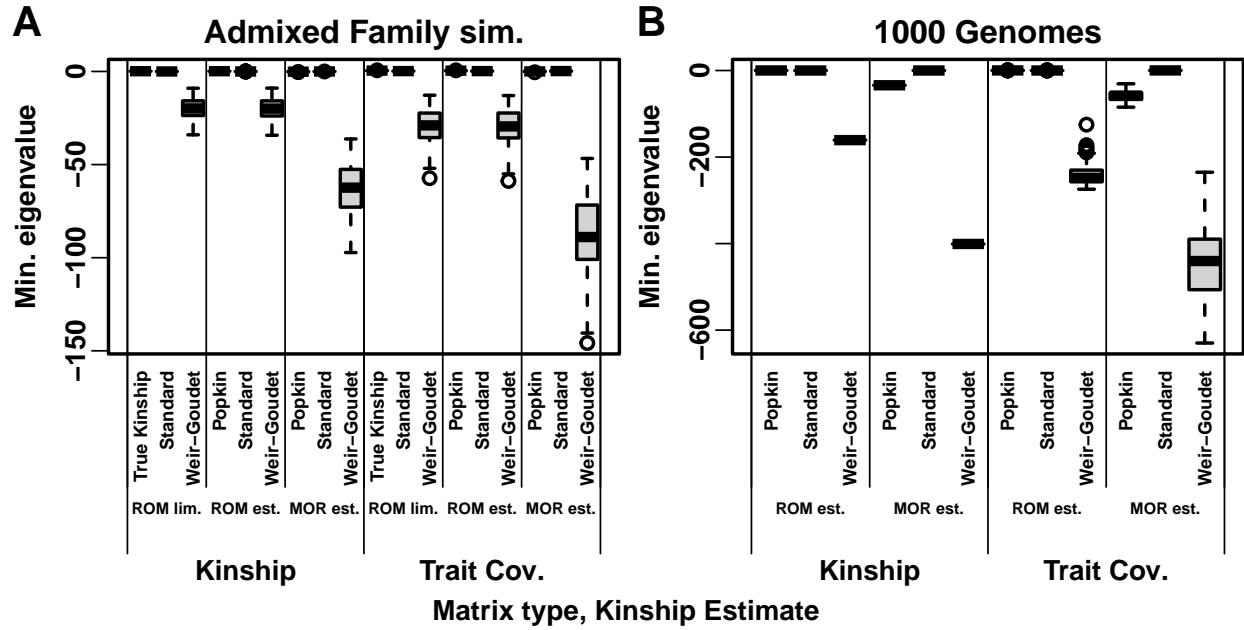


Figure S16: Minimum eigenvalue of kinship and trait covariance ( $\mathbf{V}$ ) matrices with  $h^2 = 0.8$ . Each distribution is over 100 replicates (1000 Genomes kinship has one value since genotypes are fixed, but  $\mathbf{V}$  varies per replicate). All WG matrices have very large negative eigenvalues, and Popkin MOR has negative eigenvalues as well; in these cases  $\mathbf{V}$  always has negative eigenvalues too.

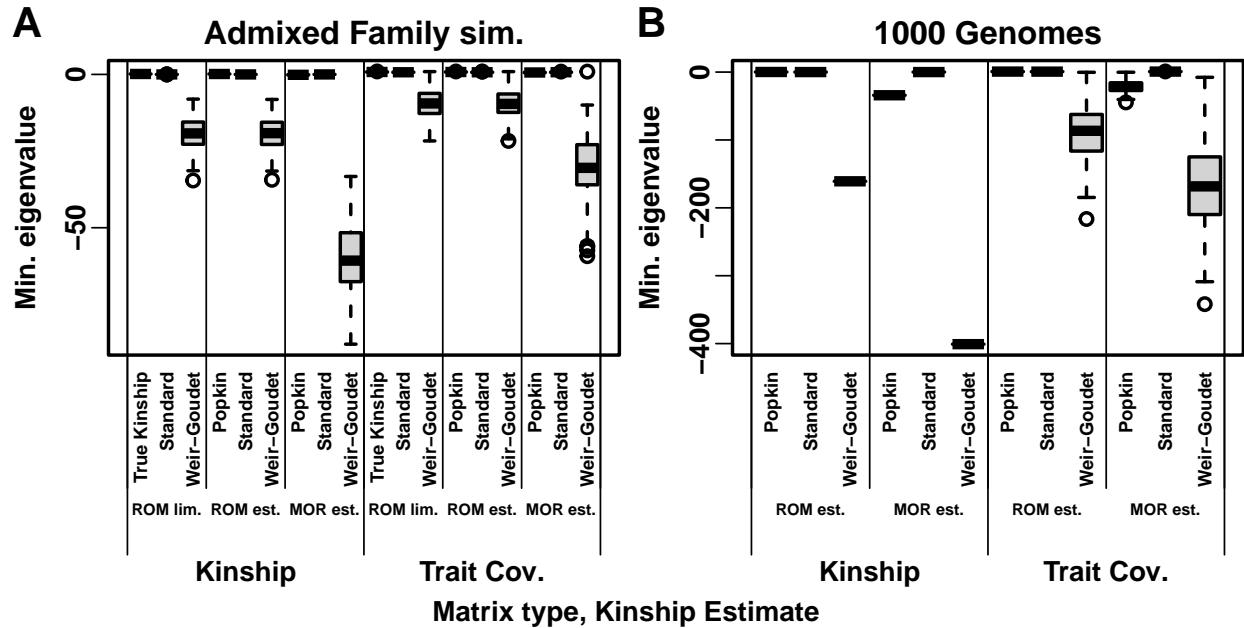


Figure S17: Minimum eigenvalue of kinship and trait covariance ( $\mathbf{V}$ ) matrices with  $h^2 = 0.3$ . Like Fig. S16 except simulated with lower heritability.

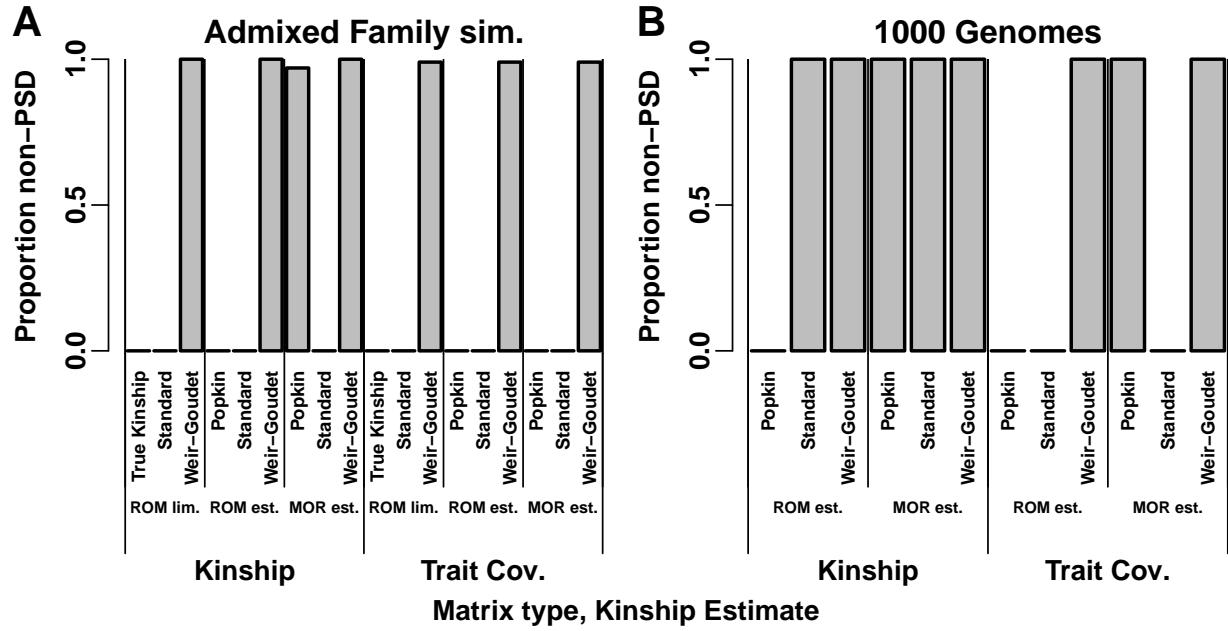


Figure S18: Proportion of kinship and trait covariance ( $\mathbf{V}$ ) matrices with  $h^2 = 0.3$  that are not positive semidefinite (PSD). Like Fig. 7 except simulated with lower heritability.

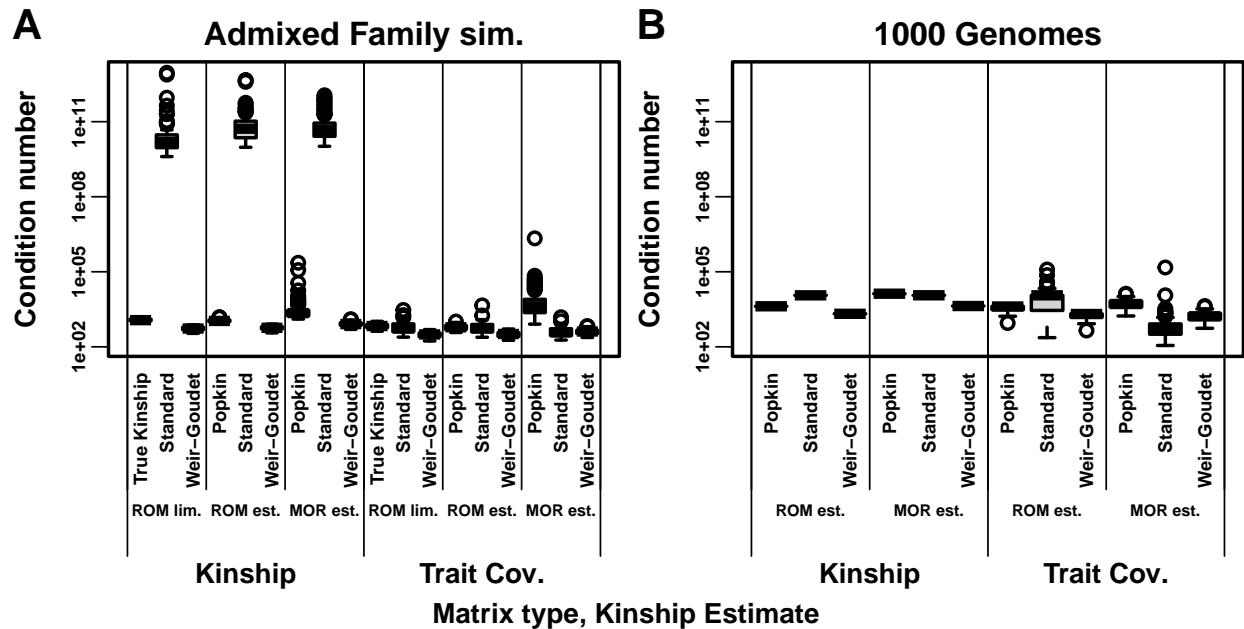


Figure S19: Condition numbers of kinship and trait covariance ( $\mathbf{V}$ ) matrices with  $h^2 = 0.8$ . Larger condition numbers reflect ill-conditioned problems such as near singularity. Each distribution is over 100 replicates (1000 Genomes kinship has one value since genotypes are fixed, but  $\mathbf{V}$  varies per replicate).

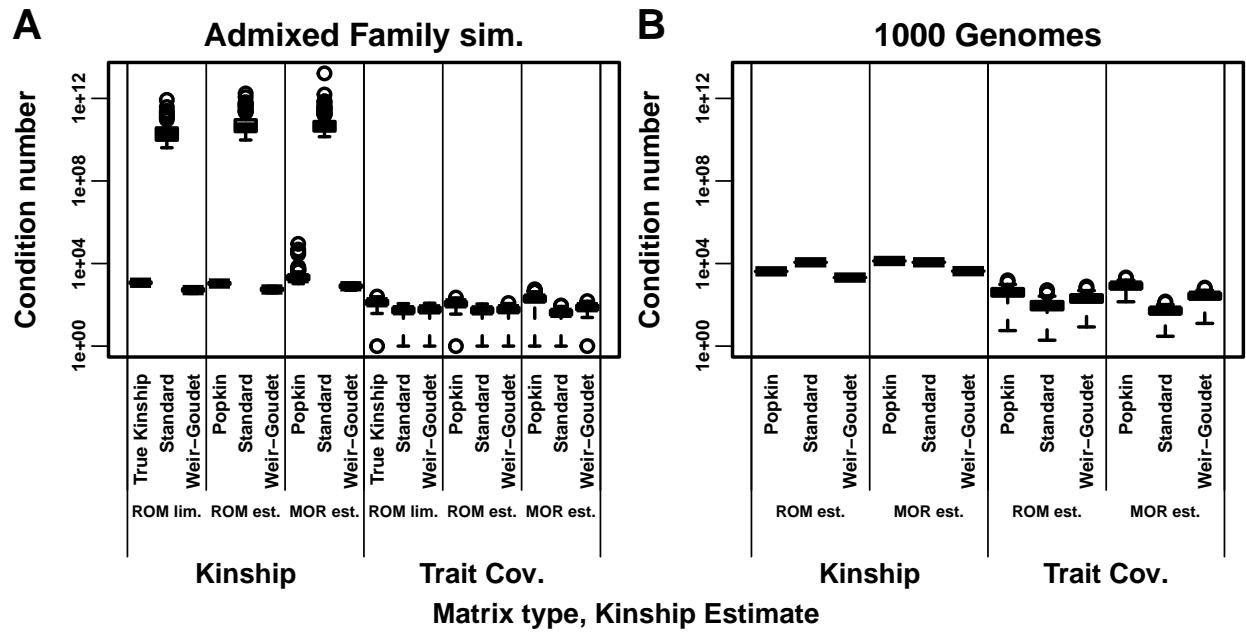


Figure S20: Condition numbers of kinship and trait covariance ( $V$ ) matrices with  $h^2 = 0.3$ . Like Fig. S19 except simulated with lower heritability.

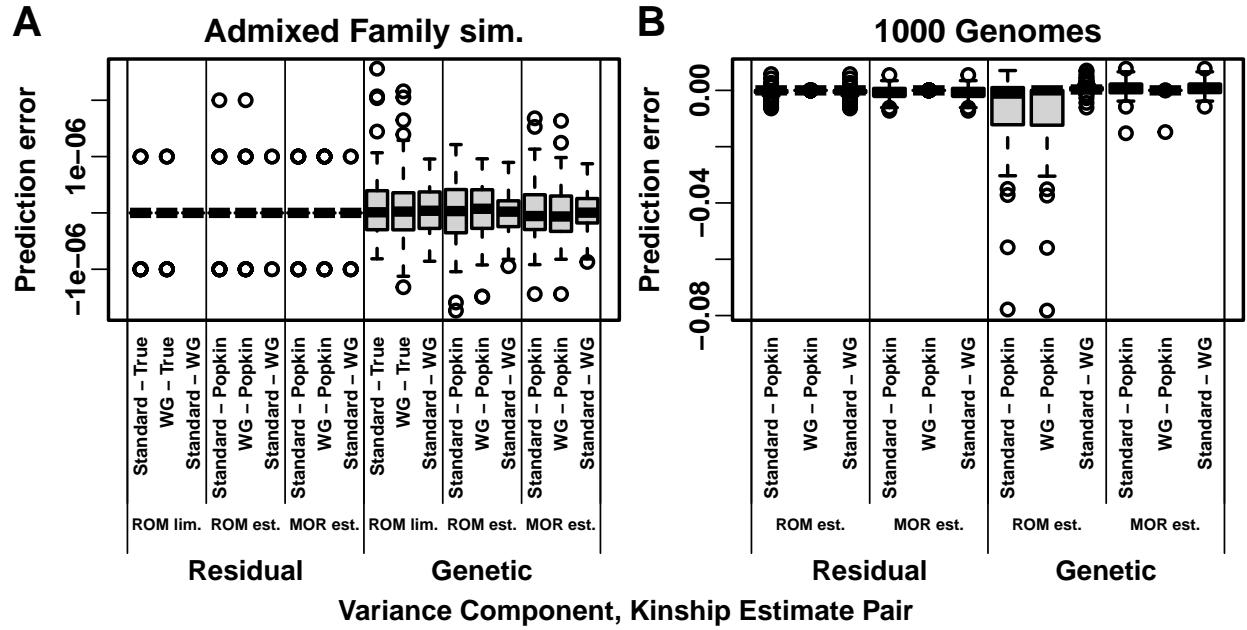


Figure S21: **Variance component prediction errors across evaluations with  $h^2 = 0.8$ .** Here we test the predictions that  $\sigma_{\epsilon}^{2l} = \sigma_{\epsilon}^2$  and  $\sigma^{2l} = c\sigma^2$  in Eq. (16). For Residual, prediction error (y-axis) is  $\sigma_{\epsilon}^{2l} - \sigma_{\epsilon}^2$  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.

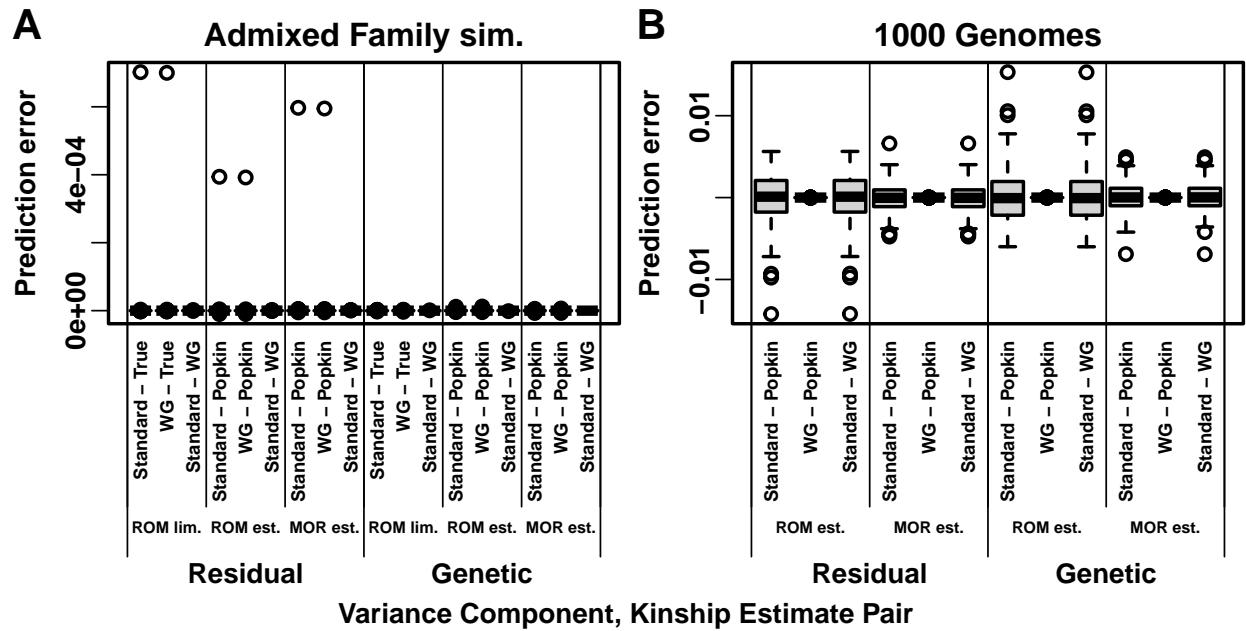


Figure S22: **Variance component prediction errors across evaluations with  $h^2 = 0.3$ .** Like Fig. S21 except simulated with lower heritability. All errors are zero within the reported precision.

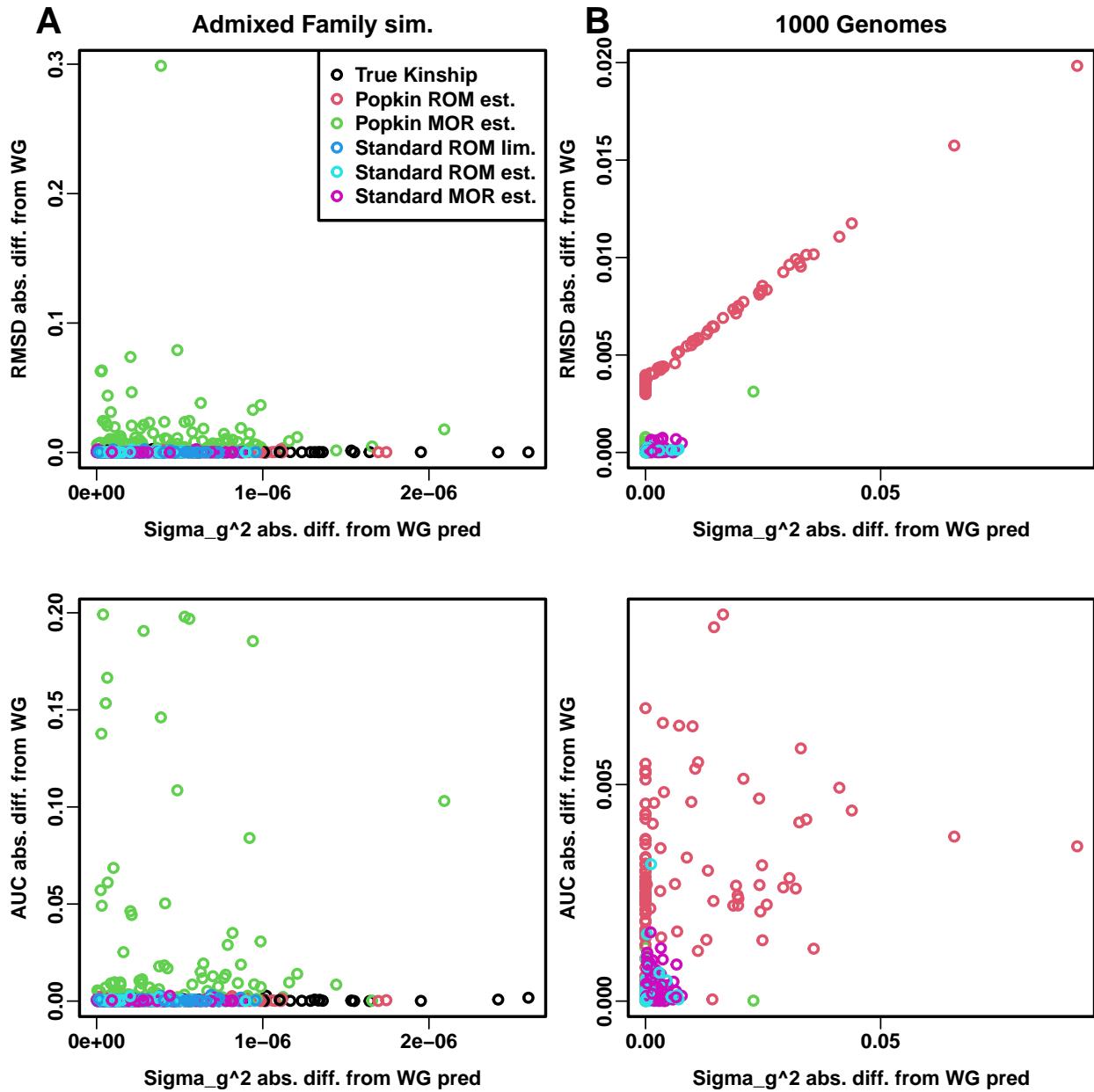
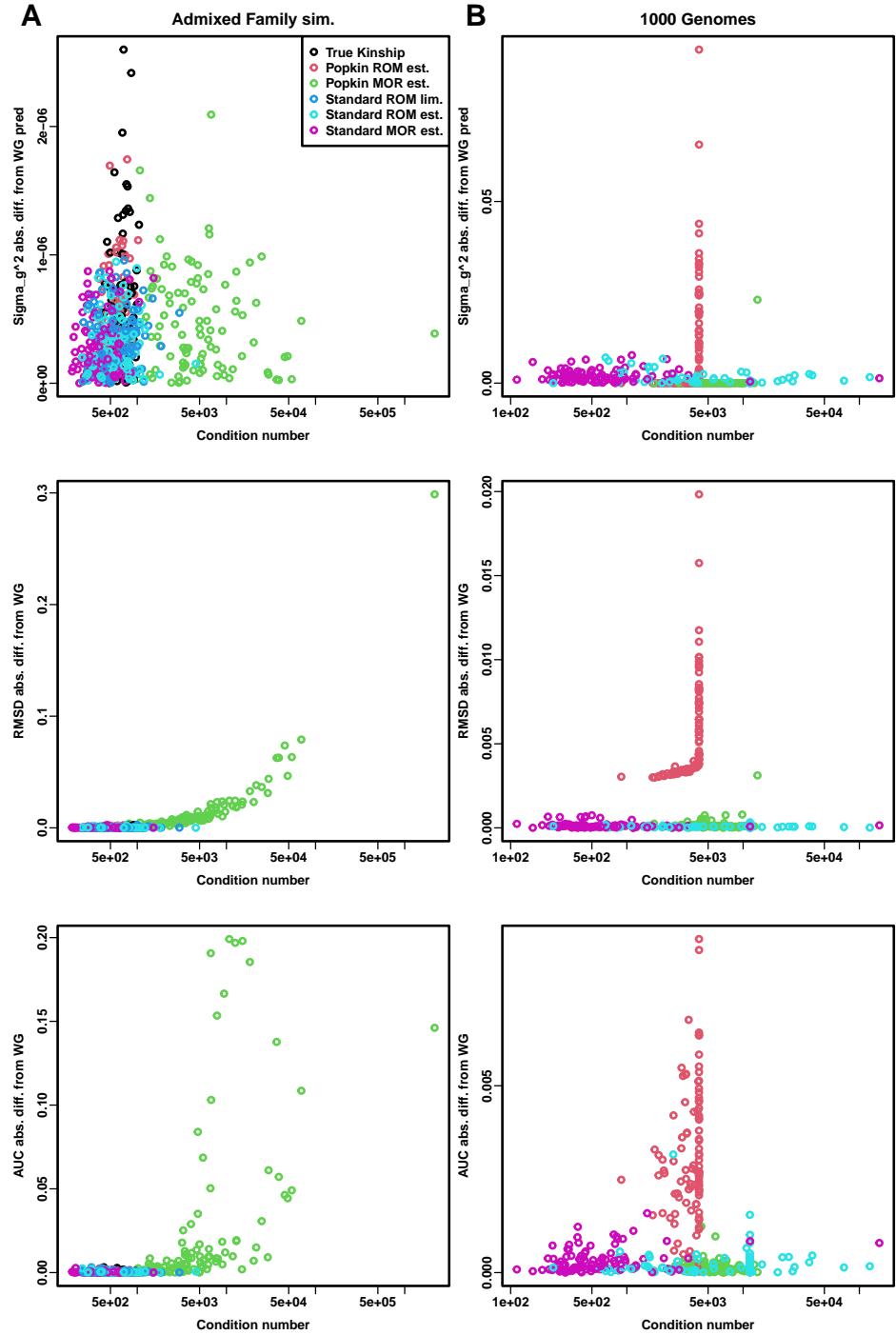


Figure S23: **AUC<sub>PR</sub>** and SRMSD<sub>p</sub> prediction errors explained by variance component errors. Evaluations with  $h^2 = 0.8$  only. Genetic variance component ( $\sigma^2$ ) absolute error is calculated with the formulas in Fig. S21 using WG as reference since its  $\mathbf{V}$  had the lowest condition numbers (Fig. S19). 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 S24:  $AUC_{PR}$  and  $SRMSD_p$  prediction errors explained by the condition number of  $\mathbf{V}$ .** Evaluations with  $h^2 = 0.8$  only.  $AUC_{PR}$  and  $SRMSD_p$  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. S19). **A.** The large popkin MOR prediction errors ( $AUC_{PR}$ ,  $SRMSD_p$ , 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  $\mathbf{V}$ .