1 **Concepts, estimation and interpretation of SNP-based heritability**

2

3 Jian Yang[1,2], Jian Zeng[1], Michael E. Goddard[3,4], Naomi R. Wray[1,2], Peter M. Visscher[1,2]

4

5 [1]Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland 4072,
6 Australia
7 [2]Queensland Brain Institute, The University of Queensland, Brisbane, Queensland 4072,
8 Australia
9 [3]Faculty of Veterinary and Agricultural Science, University of Melbourne, Parkville, Victoria 3010,
10 Australia
11 [4]Biosciences Research Division, Department of Economic Development, Jobs, Transport and
12 Resources, Bundoora, Victoria 3083, Australia

13

14 Correspondence: JY (jian.yang@uq.edu.au) and PMV (peter.visscher@uq.edu.au)

15

16 **Narrow-sense heritability (h^2) is an important genetic parameter that quantifies the**
17 **proportion of phenotypic variance in a trait attributable to the additive genetic variation**
18 **generated by all causal variants. Estimation of h^2 previously relied on closely related**
19 **individuals but recent developments allow the estimation of variance explained by all**
20 **SNPs used in a genome-wide association study (GWAS) in conventionally unrelated**
21 **individuals, i.e. the SNP-based heritability (h^2_{SNP}). In this perspective, we discuss recently**
22 **developed methods to estimate h^2_{SNP} for a complex trait (and genetic correlation between**
23 **traits) using individual- or summary-level GWAS data. We discuss the issues that could**
24 **influence the accuracy of $\hat{h}^2_{\mathrm{SNP}}$, definitions, assumptions and interpretations of the models,**
25 **and pitfalls of misusing the methods and misinterpreting the models and results.**

26

27 Estimation of the variance explained by all SNPs used in a population-based genome-wide
28 association study (GWAS) was initially motivated by the 'missing heritability' problem[1]. The
29 problem was that the estimated variance explained by genome-wide significant (GWS) SNPs
30 discovered in GWAS (denoted $\hat{h}^2_{\mathrm{GWS}}$) was only a fraction of the estimated heritability ($\hat{h}^2$) from
31 family or twin studies[2], where $\hat{h}^2_{\mathrm{GWS}}$ was estimated in a multi-SNP model to account for linkage
32 disequilibrium (LD) among SNPs and in an independent sample to avoid overestimation due to
33 the winner's curse issue[3]. Taking human height as an example, $\hat{h}^2_{\mathrm{GWS}}$ was 5% before 2010 (ref[4]),
34 which is much smaller than a frequently quoted $\hat{h}^2$ of 80% from family or twin studies[5-7]. This
35 raised concerns about the cost-effectiveness of GWAS as an experimental design for gene
36 discovery[8]. Several explanations of the missing heritability were proposed, including a large

37 number of common variants of small effect yet to be discovered, rare variants of large effects not
38 tagged by common SNPs on genotyping arrays, and inflation in pedigree-based $\hat{h}^2$ due to shared
39 environmental effects, non-additive genetic variation, and/or epigenetic factors[2,9]. The missing
40 heritability question also reignited the debate about the 'common-disease common-variant'
41 hypothesis[10], i.e. whether the proportion of heritability for common disease not explained by
42 GWS loci is due to rare variants of large effect not tagged by the current generation of SNP
43 arrays, or undetected common variants of small effect[2,11]. It is therefore important to quantify
44 the proportion of variance attributable to all common SNPs (e.g. minor allele frequency, MAF ≥
45 0.01) used in GWAS. If common SNPs are the major contributor to heritability, then the concern
46 about missing heritability is premature because it depends on experimental sample size of
47 GWAS[12].
48

49 **Estimation of the SNP-based heritability – the GREML approach**

50 SNP-based heritability (or h^2_{SNP}) was initially defined as the proportion of phenotypic variance
51 explained by all SNPs on a genotyping array[13], and is therefore dependent of the density of SNP
52 array. The concept has now been expanded to refer to the variance explained by any set of SNPs,
53 e.g. all genetic variants from in-depth whole-genome sequencing (WGS) or imputed from a
54 reference[14]. Yang et al. used a mixed linear model (MLM) approach to estimate h^2_{SNP} in a GWAS
55 data set of unrelated individuals, and demonstrated that common SNPs on a genotyping array
56 explain a large proportion (45%) of variance in height[13]. Here, "unrelated" means distantly
57 related individuals rather than individuals with no genetic relatedness because even random
58 pairs of individuals drawn from a general population could share distant ancestors. Given the
59 small $\hat{h}^2_{\mathrm{GWS}}$ (5%) and relatively large $\hat{h}^2_{\mathrm{SNP}}$ (45%), it was concluded that for complex traits like
60 height there are likely a large number of common variants with effect sizes too small to pass the
61 stringent GWS threshold (P < 5e-8) in GWAS even with the sample sizes that were considered
62 large at that time (n = 1,000s to early 10,000s before 2010), consistent with a model of polygenic
63 inheritance. It was subsequently predicted that more genetic variants could be discovered with
64 larger sample sizes, whilst keeping the same experimental design of GWAS. This prediction has
65 been realized by recent studies with n > 100,000 for height, BMI, schizophrenia and many other
66 traits and diseases[15-20]. Under a polygenic model, the amount of unexplained heritability by GWS
67 loci depends on sample size[12]. The aforementioned comparison of 5% vs. 80% for height in 2009
68 (ref[4]) became 16% vs. 80% only five years later[15]. Given the nearly linear relationship between
69 the number of GWS loci and logarithm of sample size (i.e. $\log(n)$) observed in published GWAS[12]
70 and the highly polygenic nature of most complex traits[21,22], we predict that the shrinking of the
71 gap between $\hat{h}^2_{\mathrm{GWS}}$ and $\hat{h}^2_{\mathrm{SNP}}$ will be less than linear in $\log(n)$ because the variance explained by
72 SNPs discovered in studies with larger sample sizes tend to be smaller.

73

74 The Yang et al. approach was subsequently termed as Genomic Relatedness-matrix (GRM)
75 Restricted Maximum Likelihood (GREML[23]) and implemented in the GCTA software tool[24] (**Box**
76 **1**). GREML shares features with a pedigree-based analysis (part 1 of the **Supplementary Note**),
77 but GREML is usually applied to a sample of unrelated individuals (note that this is also the usual
78 experimental design for GWAS), and hence $\hat{h}^2_{\mathrm{SNP}}$ is unlikely to be confounded by common
79 environmental effects (**Fig. 1**). For distantly related pairs, the amount of genome shared is small
80 and highly variable, and it is unlikely that those pairs that share slightly more genome than
81 average will also share more common environments in a relatively homogenous population. The
82 use of unrelated individuals also means that $\hat{h}^2_{\mathrm{SNP}}$ is unlikely to be contaminated with
83 contributions from non-additive genetic effects since the shared non-additive genetic effects are
84 tiny (of the order of the square of the shared additive effects), whereas this could be a problem
85 in $\hat{h}^2$ estimated from families depending on the study design. In addition, GREML can be applied
86 to family data but the estimate should be interpreted with caution (part 3 of the **Supplementary**
87 **Note**).
88
89 The GREML estimate quantifies directly the proportion of phenotypic variance explained by all
90 SNPs used in GWAS, and therefore provides the upper limit of $\hat{h}^2_{\mathrm{GWS}}$ given the same
91 experimental design. The information to estimate h^2_{SNP} comes from very small coefficients of
92 genetic relationship between pairs of individuals, but small standard error (SE) of $\hat{h}^2_{\mathrm{SNP}}$ (part 4
93 of the **Supplementary Note**) can be achieved because of the large number of pairwise
94 relationships (e.g., 50 million pairs for a study using 10,000 individuals) although these pairs are
95 not independent. Subsequent work has extended the method to estimate h^2_{SNP} in disease data[25]
96 (part 5 of the **Supplementary Note**) and genetic correlation (r_{g}) between traits[26,27] (part 6 of
97 the **Supplementary Note**). There are several caveats of estimating h^2_{SNP} using data from case-
98 control studies (part 5 of the **Supplementary Note**) and interpreting the estimates on different
99 scales (**Fig. 2**).
100
101 There are multiple terms and notations that have been used to describe the parameter
102 estimated in GREML, e.g. chip heritability, heritability on chip or SNP heritability. We
103 recommend using the term SNP-based heritability and the notation h^2_{SNP}. Unlike h^2, which is a
104 population-level parameter irrespective of experimental design, h^2_{SNP} is a parameter given a set
105 of SNPs. We believe that it is also necessary to use a specific notation $\hat{h}^2_{\mathrm{ped}}$ to represent h^2
106 estimated from pedigrees (including twins) because of the potential biases in pedigree-based $\hat{h}^2$
107 due to confounding factors such as common environmental effects. We have shown above that

108 h^2_{SNP} is per definition smaller than h^2 because not all the causal variants, in particular those with
109 low frequency, can be perfectly tagged by SNPs used in GWAS (**Fig. 3a**; part 1 of the
110 **Supplementary Note**). Here, by causal variant we mean a genetic mutation that causes a
111 cascade of events in biological pathways and thereby a consequence in phenotypic change,
112 rather than an associated variant identified from GWAS. In a particular case where h^2_{SNP} is
113 defined as the variance explained by all causal variants, then $h^2_{\text{SNP}} = h^2$. In reality, however,
114 causal variants are unknown. An unbiased estimate of h^2 might be achieved by estimating h^2_{SNP}
115 from in-depth WGS data assuming that all causal variants have been sequenced and there is no
116 difference in LD between causal and other sequence variants[14] (see below for more discussion).
117

118 **Both GWAS and estimation of h^2_{SNP} by GREML utilize LD**

119 GWAS relies, by design, on genotyped common SNPs tagging unknown causal variants in the
120 same chromosomal region. Estimating how much trait variation is tagged when fitting all SNPs
121 simultaneously also utilise LD between SNPs and unobserved causal variants. A sparse SNP
122 array that does not cover common variation in the genome well is less likely to lead to the
123 discovery of trait-associated variants (even with a large sample size) and fitting those SNPs
124 together in a GREML analysis will result in a smaller proportion of phenotypic variance
125 explained than a denser SNP array (**Fig. 3a**). Since the maximum possible LD correlation
126 between two genetic variants declines as their difference in MAF increases[28], genetic variation at
127 rare variants (i.e. MAF < 0.01) is unlikely to be well tagged by common SNPs on genotyping
128 arrays (**Fig. 3a**). If causal variants are located in genomic regions with a different LD property
129 from the rest of the genome, it can lead to bias in $\hat{h}^2_{\text{SNP}}$ (ref[14,29,30]) (see below for more
130 discussion).
131

132 **Interpretation and misinterpretation of the GREML model**

133 There are several circumstances where the principle of GREML is misinterpreted and the
134 method is misapplied, and this could potentially lead to misleading or confusing inference.
135 GREML is based on a random-effect model (**Box 1**). If the number of SNPs (m) is smaller than
136 sample size (n), this model is similar to a linear regression analysis (fixed-effect model) in terms
137 of estimating h^2_{SNP} (note that the adjusted R^2 from multiple regression is an unbiased estimate of
138 variance explained in a fixed-effect model). Such a hypothetical experiment would not rely on
139 selecting SNPs to be individually genome-wide significant nor would it rely on assumptions
140 about the genetic architecture. In either a linear regression or random-effect model, the effect
141 sizes of SNPs are fitted jointly (therefore accounting for LD among SNPs), i.e. the effect of any
142 SNP is interpreted as the effect size of this SNP conditioning on the joint effects of all other SNPs.
143 In GWAS, m is normally larger than n, in which case there is no unique solution to the fixed-effect

144 model, a well-known over-fitting problem in statistics. In a random-effect model, there is an
145 additional assumption that the joint SNP effects $\mathbf{u} = \{u_1, u_2, \cdots, u_m\}$ follow a normal distribution
146 with mean 0 and variance σ_{u}^2 (see **Box 1** for notations) so that the model parameters are
147 estimable even when *m* is larger than *n*, where σ_{u}^2 is interpreted as per-SNP genetic variance
148 when all SNPs are fitted jointly, hence accounting for LD[31]. Therefore, σ_{u}^2 is not consistent across
149 models having different numbers of SNPs. There is a misunderstanding that GREML does not
150 account for LD because it does not have a covariance matrix for **u** (ref[32]). This is incorrect. In fact,
151 the LD correlations among SNPs have been modeled by fitting the SNP genotype matrix **W**,
152 similar to that in linear regression analysis[31]. Since σ_{u}^2 is the variance of a SNP effect
153 conditioning on the joint effects of all other SNPs and w_{ij} is the standardised SNP genotype, the
154 additive genetic variance captured by all SNPs is $\sigma_{\mathrm{g(SNP)}}^2 = m\sigma_{\mathrm{u}}^2$ (**Box 1**).
155
156 In part 8 of the **Supplementary Note** we list five scenarios where GREML (or the GCTA tool) is
157 misused, resulting in potentially misleading results. In addition, there is often a question about
158 whether the SNPs included in GREML analysis need to be pruned for LD. As discussed above,
159 GREML accounts for LD so that LD pruning is not necessary (but see later for discussion of bias
160 due to the non-random distribution of causal variants with respect to LD). LD pruning using a
161 high r^2 threshold might increase the estimate but the likelihood of the model is not improved
162 (**Fig. 3b**). We need to be cautious about interpreting the GREML estimate from pruned SNPs
163 because of the change in MAF spectrum of SNPs by LD pruning (**Fig. 3c**). Changing the set of
164 SNPs means that the underlying parameter being estimated (i.e. h_{SNP}^2 for a set of LD- pruned
165 SNPs) is different from the original parameter (i.e. h_{SNP}^2 for all SNPs).
166

167 Bias due to non-random distribution of causal variants with respect to LD

168 We have mentioned above that $\hat{h}_{\mathrm{SNP}}^2$ using WGS data could be a biased estimate of h^2 if the LD
169 property of causal variants is different from that of the other variants[14,29,30,34]. The unbiasedness
170 of GREML to estimate h^2 using WGS data depends on the ratio of $\overline{r_{\mathrm{MQ}}^2}$ (mean LD r^2 between
171 causal and non-causal variants) to $\overline{r_{\mathrm{MM}}^2}$ (mean LD r^2 between non-causal variants)[14]. Note that
172 because r^2 is a function of MAF, difference in MAF spectrum between causal and non-causal
173 variants will lead to a difference in LD (i.e. MAF-mediated LD bias), resulting in a bias in $\hat{h}_{\mathrm{SNP}}^2$.
174 One solution is to stratify SNPs by MAF (i.e. MAF-stratified GREML, GREML-MS)[14,33,35], which
175 reduced bias in the estimate due to MAF-mediated LD bias. However, a more general approach is
176 to not rely on a specific model of the interplay between allele frequency, effect size and LD, but
177 instead stratify SNPs by MAF and LD jointly and estimating genetic variance with MAF-LD
178 subsets. This approach, termed GREML-LDMS, appears to provide unbiased estimates of h^2 as

179 well as the contributions of common and rare variants to h^2 in simulations based on WGS data
180 regardless of the underlying genetic architecture and distribution of causal variants with respect
181 to MAF and LD[14,36]. We recommend the use of GREML-LDMS to estimate h^2_{SNP} in imputed data
182 (part 9 of the **Supplementary Note**). The applications of GREML-LDMS to WGS data sets with
183 rich phenotypes in the future will be able to provide nearly unbiased estimates of h^2 in unrelated
184 individuals and quantify the variance explained by all rare variants for a range of complex traits.
185 However, large sample sizes are required to estimate h^2_{SNP} with useful precision because
186 $\mathrm{var}(\hat{h}^2_{\mathrm{SNP}})$ depends on sample size and variant density[37] (part 4 of the **Supplementary Note**),
187 e.g., a sample size of ~33,000 is needed to get an SE of 0.02 for WGS data.
188
189 Speed et al.[29] proposed a method called LDAK to correct for the LD bias. The basic idea is to
190 weight each SNP by a factor inversely proportional to its LD with SNPs nearby. This weighting
191 strategy can introduce MAF bias because it gives more weights to SNPs with lower MAF
192 (supplementary figure 2 of Yang et al.[14]) as LD is a function of MAF[28]. The LDAK model implicitly
193 assumed that variance explained by a rare variant (e.g. $0.001 < \mathrm{MAF} < 0.01$) is more than 10
194 times larger than that by a common variant (e.g. $0.1 < \mathrm{MAF} < 0.5$) (based on the LDAK weights
195 calculated from a sequenced reference set[14]). This an unrealistic model because it predicts that
196 we would have orders of magnitude higher power to detect rare variants than common variants,
197 a prediction not consistent with empirical results, e.g., human height[15,38], schizophrenia[17,39] and
198 type 2 diabetes [40]. The LDAK-induced MAF bias can be substantial especially when there is a
199 large number of rare variants (e.g. in a WGS data set), leading to an inflated estimate of h^2_{SNP}
200 (ref[14]).
201
202 The LDAK model has recently been changed substantially[41]. Two new parameters have been
203 added: one is a weighting according to MAF and the other is a weighting according to imputation
204 accuracy. Although it is not the justification for these two new parameters, both give more
205 weight to common variants than the original LDAK model[41]. The revised LDAK model is now
206 more similar to GREML-LDMS[14], but not identical, since Speed et al.[41] estimate a higher SNP
207 heritability from their empirical analyses on a range of traits. In simulation studies to compare
208 the methods, the results depend on the model used to simulate the data. Unfortunately we
209 cannot be sure which is the correct model for any given trait. GREML-LDMS makes fewer
210 assumptions about the relationship between causal variants, LD and MAF and thereby appears
211 to be more robust than the revised LDAK method[36], although at the expense of estimating more
212 parameters. On balance, we conclude that this topic merits further investigation[36], since the
213 relationship between local LD, locus heterozygosity and additive genetic variance for complex
214 traits has not yet been resolved, and indeed may differ across the genome and between traits.

215

216 **Assumptions about relationship between effect sizes and allele frequencies**

217 Under an evolutionarily neutral model, the proportion of variance in a polygenic trait explained
218 by all variants in a MAF bin is linearly proportional to the width of the MAF bin[14] (variance
219 explained by a rare variant, on average, is tiny but there are a large number of them). Therefore,
220 a significant deviation of the observed variance explained in a MAF bin from the expected value
221 is evidence that the trait has been under natural selection[14,42]. In GCTA-GREML, we standardise
222 the SNP genotypes and assume the effect size per standardised genotype (u_i) follows a normal
223 distribution. This implicitly assumes larger per-allele effect (b_i) for a SNP with lower MAF,
224 consistent with a model of purifying selection where variants with larger effect sizes tend to be
225 under higher selection and therefore are more likely to be at lower frequencies (e.g. MAF < 0.1).
226 There is an option in GCTA to run GREML assuming that effect size is independent of MAF
227 (neutral model). However, the difference between the two models is trivial in GREML-MS
228 analysis[14]. Moreover, GREML-MS allows the data to reveal the relationship between variance
229 explained and MAF. One of the important extensions of GREML in the future is to estimate
230 directly from the data a parameter to quantify the relationship between b_i and allele frequency
231 whilst fitting a mixture distribution to the joint effects of SNPs[43] (part 10 of the **Supplementary**
232 **Note**).

233

234 **Comparison with HE regression**

235 As described in **Box 1**, the GREML analysis is based on an MLM that is equivalent to fitting the
236 additive genetic values of all individuals, i.e. $\mathbf{y} = \mathbf{g} + \mathbf{e}$ with $\text{var}(\boldsymbol{y}) = \mathbf{A}\sigma^2_{\text{g(SNP)}} + \mathbf{I}\sigma^2_{\text{e}}$. The
237 variance components in this model are usually estimated using the REML approach. However,
238 the REML algorithm is computationally intensive (part 11 of the **Supplementary Note**).
239 Alternatively, $\sigma^2_{\text{g(SNP)}}$ can be estimated from Haseman-Elston (HE) regression[37,44], i.e.
240 $y_i y_j = b_0 + b_1 A_{ij} + e_{ij}$, where $b_1 = \sigma^2_{\text{g(SNP)}}$. The performance of GREML has been compared
241 using extensive simulations in Golan et al.[45] in ascertained case-control studies where GREML
242 estimates can be biased especially when m/n is small and disease prevalence is low. We also
243 performed simulation to compare the two methods with an emphasis on the SE under a
244 polygenic model (part 12 of the **Supplementary Note**). HE regression is computationally much
245 more efficient but slightly less powerful than REML as the SE of $\hat{h}^2_{\text{SNP}}$ from HE regression is
246 larger than that from REML (**Supplementary Table 1** and part 12 of the **Supplementary Note**).
247 The small difference in SE between the methods might not be important when the sample size
248 becomes very large. For example, given $\hat{h}^2_{\text{SNP}}$ > 0.1, whether the SE is 0.01 (REML) or 0.015 (HE
249 regression) does not make any difference in making statistical inference whether $h^2_{\text{SNP}} = 0$. HE

250 regression can also be used to estimate multiple genetic components, e.g. multiple sets of SNPs
251 stratified by MAF or chromosomes (**Fig. 4**), or to estimate genetic correlations between traits
252 (**Supplementary Table 2**). These analyses have been implemented in the latest version of GCTA
253 (GCTA-HE) (**URLs**). In addition, phenotype correlation – genotype correlation (PCGC) regression
254 is an implementation of HE regression designed for disease data to attenuate the biases in
255 ascertained case-control studies[22,45] (**URLs**).
256
257 **Non-additive genetic variation**
258 The GREML approach has been extended to estimate dominance genetic variance tagged by
259 SNPs in unrelated individuals based on a classical quantitative genetics model[46]. Similar to the
260 additive GREML method, the dominance GREML model fits the additive and dominance effects of
261 all SNPs as two sets of random effects in an MLM. This is an orthogonal model because the
262 additive and dominance genotype variables and thereby the additive and dominance GRMs are
263 independent. On average across 79 quantitative traits, additive genetic variation explained ~15%
264 of the phenotypic variance and dominance genetic variation explained ~3% of variance[46]. The
265 ratio of additive to dominance variance is consistent with what is expected from theory[47]. The
266 method can be further extended to estimate genetic variance attributable to epistasis[48] based on
267 the classical quantitative genetics model[49], $y = g_A + g_D + g_{AA} + g_{AD} + g_{DD} + e$, where g_A and
268 g_D are the additive and dominance genetic values of an individual, and g_{AA}, g_{AD} and g_{DD} are the
269 additive-by-additive, additive-by-dominance, and dominance-by-dominance epistatic genetic
270 values respectively. However, the sample size will need to be very large to get a precise estimate
271 of epistatic variance because the variance in epistatic genetic relationship between unrelated
272 individuals is very small. For instance, the genetic relationship for g_{AA} is A_{ij}^2 which has a
273 variance of $2[\text{var}(A_{ij})]^2$ (ref[49]). For HapMap3 SNPs, $\text{var}(A_{ij}) \approx 2.0 \times 10^{-5}$ so that the variance
274 in genetic relationship for g_{AA} is $\sim 1.0 \times 10^{-9}$, meaning that we will need over a million
275 unrelated individuals to estimate the variance explained by g_{AA} with an SE < 0.05 (> 4 million
276 unrelated individuals to get SE < 0.01). The variance in dominance genetic relationship is
277 smaller than additive genetic relationship. Therefore, it will be even more difficult to estimate
278 variance for g_{AD} or g_{DD}.
279
280 **Estimating h^2_{SNP} and r_g from GWAS summary data**
281 We have discussed above the MLM-based approaches to estimate h^2_{SNP} using individual-level
282 GWAS data. There are other methods that are able to estimate h^2_{SNP} from GWAS summary data
283 (estimated SNP effects and their standard errors for all SNPs analyzed in a study)[50]. For example,
284 the AVENGEME method that uses maximum likelihood to estimate the genetic variance of a trait,
285 the proportion of genetic variants affecting the trait, and the genetic covariance (and therefore

286 genetic correlation) between traits from the test-statistic for association between phenotype
287 and polygenic risk score (PRS)[51,52]. We can also estimate h^2_{SNP} directly from summary data
288 utilizing the deviation of the observed χ^2 test-statistic for a SNP from its expected value under
289 the null hypothesis of no association[56] (part 13 of the **Supplementary Note**). This is the basic
290 principle of the recently developed LD score regression approach (LDSC)[53]. This approach
291 requires only the summary-level data from GWAS because LD scores can be estimated from a
292 reference sample (e.g. the 1000 Genomes). LDSC has been extended to estimate r_{g} between traits
293 using summary data[54], which allows the traits measured on different samples regardless
294 whether there is an overlap between samples (and the proportion of sample overlap is
295 estimated), and to partition $\hat{h}^2_{\mathrm{SNP}}$ by functional annotation[55]. This method provides great
296 flexibility for researchers to estimate r_{g} between any two GWAS data sets. Both GREML and
297 LDSC aim at estimating the variance explained by all SNPs used in GWAS. However, there are
298 distinct differences between the two methods. LDSC is orders of magnitude faster than GREML,
299 and the computing time for LDSC does not scale up with sample size. LDSC only requires
300 summary-level data, which allows the re-analysis of summary data available from published
301 meta-analyses. There are also limitations for LDSC. LDSC is not applicable to estimate the
302 variance explained by rare variants (e.g. MAF < 0.01) using either imputed or WGS data[36] nor the
303 variance explained by SNPs in small genomic regions (although the latter has been overcome by
304 the HESS method developed recently[56]), and is more sensitive to genetic architecture of the trait
305 (**Supplementary Table 3**). Result from a previous study shows that $\hat{h}^2_{\mathrm{SNP}}$ from LDSC are
306 consistently smaller than those from GREML in the same data set[57], which is likely due to the
307 errors in LD scores estimated from the reference (by default LDSC uses LD score from HapMap3
308 SNPs in 1000 Genomes). We therefore advise using LD scores from the data used to generate the
309 GWAS summary statistics. While this may not be possible for published summary statistics, it
310 should be possible for large cohorts such as the UK Biobank. It is noteworthy that LDSC will
311 suffer bias in a similar way as GREML if causal variants are non-randomly distributed with
312 respect to LD. The estimate of r_{g} from bivariate LDSC is consistent with that from bivariate
313 GREML but the Jackknife SE of $\hat{r}_{\mathrm{g}}$ from LDSC is larger than that expected from the approximation
314 theory[37,54,57].
315
316 **Summary**
317 We have provided a perspective of the methods for estimating SNP-based heritability in
318 unrelated individuals using GWAS data. We emphasized that the GREML approach accounts for
319 LD when estimating h^2_{SNP} and actually utilizes LD to tag causal variants if they are not observed.
320 We discussed the concepts and assumptions of the methods and scenarios under which the
321 estimates could be biased, the methods could be misused and the results could be

322 misinterpreted. We further discussed the extensions and applications of the methods in large
323 data sets in the future (**Box 2**). These future directions could expand our understanding of the
324 genetic architecture for human complex traits and inform the design of future experiments to
325 fully dissect genetic variation and genetic correlations.
326
327 **URLs**
328 GCTA: http://cnsgenomics.com/software/gcta/
329 GCTA-HE: http://cnsgenomics.com/software/gcta/he.html
330 PCGC: https://www.hsph.harvard.edu/alkes-price/software/
331 LDSC: https://github.com/bulik/ldsc
332
333 **Acknowledgments**
334 We thank Alkes Price for his constructive and helpful comments on an earlier version of the
335 manuscript. This research was supported by the Australian National Health and Medical
336 Research Council (1078901, 1078037 and 1113400), the Australian Research Council
337 (DP160101343), the US National Institutes of Health (GM099568 and MH100141-01), and the
338 Sylvia & Charles Viertel Charitable Foundation (Senior Medical Research Fellowship). This
339 research has been conducted using data from dbGaP (accession numbers: phs000090.v1.p1 and
340 phs000091.v2.p1), UK10K project and UK Biobank Resource (application number: 12514). A full
341 list of acknowledgements to these data sets can be found in **Supplementary Note**.
342
343 **Author contributions**
344 All authors conceived and designed the project. J.Y., J.Z. and N.R.W. performed the analyses. All
345 authors wrote the manuscript.
346
347 **References**
348 1. Maher, B. Personal genomes: The case of the missing heritability. *Nature* **456**, 18-21
349 (2008).
350 2. Manolio, T.A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747-
351 753 (2009).
352 3. Xiao, R. & Boehnke, M. Quantifying and correcting for the winner's curse in genetic
353 association studies. *Genet Epidemiol* **33**, 453-62 (2009).
354 4. Visscher, P.M. Sizing up human height variation. *Nat. Genet.* **40**, 489-90 (2008).
355 5. Fisher, R.A. The correlation between relatives on the supposition of Mendelian
356 inheritance. *Trans. Roy. Soc. Edin.* **52**, 399-433 (1918).
357 6. Silventoinen, K. *et al.* Heritability of adult body height: a comparative study of twin
358 cohorts in eight countries. *Twin Res* **6**, 399-408 (2003).
359 7. Macgregor, S., Cornes, B.K., Martin, N.G. & Visscher, P.M. Bias, precision and heritability of
360 self-reported and clinically measured height in Australian twins. *Hum. Genet.* **120**, 571-
361 80 (2006).

362 8. Goldstein, D.B. Common genetic variation and human traits. *N Engl J Med* **360**, 1696-8
363 (2009).
364 9. Eichler, E.E. *et al.* Missing heritability and strategies for finding the underlying causes of
365 complex disease. *Nat Rev Genet* **11**, 446-50 (2010).
366 10. Schork, N.J., Murray, S.S., Frazer, K.A. & Topol, E.J. Common vs. rare allele hypotheses for
367 complex diseases. *Curr Opin Genet Dev* **19**, 212-9 (2009).
368 11. Gibson, G. Rare and common variants: twenty arguments. *Nat Rev Genet* **13**, 135-45
369 (2012).
370 12. Visscher, P.M., Brown, M.A., McCarthy, M.I. & Yang, J. Five years of GWAS discovery. *Am. J.*
371 *Hum. Genet.* **90**, 7-24 (2012).
372 13. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human
373 height. *Nat. Genet.* **42**, 565-569 (2010).
374 14. Yang, J. *et al.* Genetic variance estimation with imputed variants finds negligible missing
375 heritability for human height and body mass index. *Nat. Genet.* **47**, 1114-1120 (2015).
376 15. Wood, A.R. *et al.* Defining the role of common variation in the genomic and biological
377 architecture of adult human height. *Nat. Genet.* **46**, 1173-1186 (2014).
378 16. Locke, A.E. *et al.* Genetic studies of body mass index yield new insights for obesity
379 biology. *Nature* **518**, 197-206 (2015).
380 17. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological
381 insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-7 (2014).
382 18. Okada, Y. *et al.* Genetics of rheumatoid arthritis contributes to biology and drug
383 discovery. *Nature* **506**, 376-81 (2014).
384 19. Liu, J.Z. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel
385 disease and highlight shared genetic risk across populations. *Nat Genet* **47**, 979-86
386 (2015).
387 20. Liu, C. *et al.* Meta-analysis identifies common and rare variants influencing blood
388 pressure and overlapping with metabolic trait loci. *Nat Genet* **48**, 1162-70 (2016).
389 21. Yang, J. *et al.* Ubiquitous polygenicity of human complex traits: genome-wide analysis of
390 49 traits in Koreans. *PLoS Genet.* **9**, e1003355 (2013).
391 22. Loh, P.R. *et al.* Contrasting genetic architectures of schizophrenia and other complex
392 diseases using fast variance-components analysis. *Nat Genet* **47**, 1385-92 (2015).
393 23. Benjamin, D.J. *et al.* The genetic architecture of economic and political preferences. *Proc*
394 *Natl Acad Sci U S A* **109**, 8026-31 (2012).
395 24. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex
396 trait analysis. *Am. J. Hum. Genet.* **88**, 76-82 (2011).
397 25. Lee, S.H., Wray, N.R., Goddard, M.E. & Visscher, P.M. Estimating missing heritability for
398 disease from genome-wide association studies. *Am. J. Hum. Genet.* **88**, 294-305 (2011).
399 26. Lee, S.H., Yang, J., Goddard, M.E., Visscher, P.M. & Wray, N.R. Estimation of pleiotropy
400 between complex diseases using single-nucleotide polymorphism-derived genomic
401 relationships and restricted maximum likelihood. *Bioinformatics* **28**, 2540-2542 (2012).
402 27. Lee, S.H. *et al.* Genetic relationship between five psychiatric disorders estimated from
403 genome-wide SNPs. *Nat Genet* **45**, 984-994 (2013).
404 28. Wray, N.R. Allele frequencies and the r2 measure of linkage disequilibrium: impact on
405 design and interpretation of association studies. *Twin Res Hum Genet* **8**, 87-94 (2005).
406 29. Speed, D., Hemani, G., Johnson, M.R. & Balding, D.J. Improved heritability estimation from
407 genome-wide SNPs. *Am. J. Hum. Genet.* **91**, 1011-21 (2012).
408 30. Gusev, A. *et al.* Quantifying missing heritability at known GWAS loci. *PLoS Genet.* **9**,
409 e1003993 (2013).
410 31. Yang, J., Lee, S.H., Wray, N.R., Goddard, M.E. & Visscher, P.M. GCTA-GREML accounts for
411 linkage disequilibrium when estimating genetic variance from genome-wide SNPs. *Proc*
412 *Natl Acad Sci U S A* **113**, E4579-80 (2016).
413 32. Krishna Kumar, S., Feldman, M.W., Rehkopf, D.H. & Tuljapurkar, S. Limitations of GCTA as
414 a solution to the missing heritability problem. *Proc Natl Acad Sci U S A* **113**, E61-70
415 (2016).

416 33. Lee, S.H. *et al.* Estimation of SNP heritability from dense genotype data. *Am. J. Hum. Genet.*
417 **93**, 1151-5 (2013).
418 34. Gusev, A. *et al.* Partitioning heritability of regulatory and cell-type-specific variants
419 across 11 common diseases. *Am. J. Hum. Genet.* **95**, 535-52 (2014).
420 35. Lee, S.H. *et al.* Estimating the proportion of variation in susceptibility to schizophrenia
421 captured by common SNPs. *Nat. Genet.* **44**, 247-250 (2012).
422 36. Evans, L. *et al.* Comparison of methods that use whole genome data to estimate the
423 heritability and genetic architecture of complex traits. *bioRxiv* (2017).
424 37. Visscher, P.M. *et al.* Statistical power to detect genetic (co)variance of complex traits
425 using SNP data in unrelated samples. *PLoS Genet* **10**, e1004269 (2014).
426 38. Marouli, E. *et al.* Rare and low-frequency coding variants alter human adult height.
427 *Nature* **542**, 186-190 (2017).
428 39. Purcell, S.M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia.
429 *Nature* **506**, 185-90 (2014).
430 40. Fuchsberger, C. *et al.* The genetic architecture of type 2 diabetes. *Nature* **536**, 41-7
431 (2016).
432 41. Speed, D. *et al.* Reevaluation of SNP heritability in complex human traits. *Nat Genet*
433 (2017).
434 42. Gazal, S. *et al.* Linkage disequilibrium dependent architecture of human complex traits
435 reveals action of negative selection. *bioRxiv* (2016).
436 43. Zeng, J. *et al.* Widespread signatures of negative selection in the genetic architecture of
437 human complex traits. *bioRxiv* (2017).
438 44. Haseman, J.K. & Elston, R.C. The investigation of linkage between a quantitative trait and
439 a marker locus. *Behav. Genet.* **2**, 2-19 (1972).
440 45. Golan, D., Lander, E.S. & Rosset, S. Measuring missing heritability: inferring the
441 contribution of common variants. *Proc Natl Acad Sci U S A* **111**, E5272-81 (2014).
442 46. Zhu, Z. *et al.* Dominance genetic variation contributes little to the missing heritability for
443 human complex traits. *Am J Hum Genet* **96**, 377-85 (2015).
444 47. Hill, W.G., Goddard, M.E. & Visscher, P.M. Data and theory point to mainly additive
445 genetic variance for complex traits. *PLoS Genet* **4**, e1000008 (2008).
446 48. Ronnegard, L., Pong-Wong, R. & Carlborg, O. Defining the assumptions underlying
447 modeling of epistatic QTL using variance component methods. *J Hered* **99**, 421-5 (2008).
448 49. Lynch, M. & Walsh, B. *Genetics and analysis of quantitative traits*, (Sunderland, MA:
449 Sinauer Associates, 1998).
450 50. Pasaniuc, B. & Price, A.L. Dissecting the genetics of complex traits using summary
451 association statistics. *Nat Rev Genet* (2016).
452 51. Palla, L. & Dudbridge, F. A Fast Method that Uses Polygenic Scores to Estimate the
453 Variance Explained by Genome-wide Marker Panels and the Proportion of Variants
454 Affecting a Trait. *Am J Hum Genet* **97**, 250-9 (2015).
455 52. Dudbridge, F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet* **9**,
456 e1003348 (2013).
457 53. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from
458 polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291-295 (2015).
459 54. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits.
460 *Nat. Genet.* **47**(2015).
461 55. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide
462 association summary statistics. *Nat Genet* **47**, 1228-35 (2015).
463 56. Shi, H., Kichaev, G. & Pasaniuc, B. Contrasting the Genetic Architecture of 30 Complex
464 Traits from Summary Association Data. *Am J Hum Genet* **99**, 139-53 (2016).
465 57. Yang, J. *et al.* Genome-wide genetic homogeneity between sexes and populations for
466 human height and body mass index. *Hum Mol Genet* **24**, 7445-9 (2015).
467 58. Lynch, M. & Ritland, K. Estimation of pairwise relatedness with molecular markers.
468 *Genetics* **152**, 1753-66 (1999).

469 59. Hayes, B.J., Visscher, P.M. & Goddard, M.E. Increased accuracy of artificial selection by
470 using the realized relationship matrix. *Genet. Res.* **91**, 47-60 (2009).
471 60. Browning, B.L. & Browning, S.R. A fast, powerful method for detecting identity by
472 descent. *Am J Hum Genet* **88**, 173-82 (2011).
473 61. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide
474 range of complex diseases of middle and old age. *PLoS Med* **12**, e1001779 (2015).
475

476

477 **Box 1. Statistical model used in the GREML approach to estimate h^2_{SNP}**

478 The statistical model used by GREML can be described in its simplest form as

479 $\mathbf{y} = \mathbf{Wu} + \mathbf{e}$

480 where $\mathbf{y}$ is an $n \times 1$ vector of standardised phenotypes with n the sample size, $\mathbf{W} = \{w_{ij}\}$ is an

481 $n \times m$ standardised SNP genotype matrix with m the number of SNPs, $\mathbf{u} = \{u_i\}$ is an $m \times 1$

482 vector of the additive effects of all variants when fitted jointly in the model, $\mathbf{u} \sim N(0, \mathbf{I}\sigma^2_{\mathrm{u}})$ with $\mathbf{I}$

483 an identity matrix, and $\mathbf{e}$ is a vector of residuals, $\mathbf{e} \sim N(0, \mathbf{I}\sigma^2_{\mathrm{e}})$. An equivalent model is

484 $\mathbf{y} = \mathbf{g} + \mathbf{e}$

485 where $\mathbf{g} \sim N(0, \mathbf{A}\sigma^2_{\mathrm{g(SNP)}})$ with $\sigma^2_{\mathrm{g(SNP)}}$ the additive genetic variance captured by SNPs,

486 $\sigma^2_{\mathrm{g(SNP)}} = m\sigma^2_{\mathrm{u}}$, $\mathbf{A} = \mathbf{WW}'/m$, and $h^2_{\mathrm{SNP}} = \sigma^2_{\mathrm{g(SNP)}}/[\sigma^2_{\mathrm{g(SNP)}} + \sigma^2_{\mathrm{e}}]$. The parameters to be estimated

487 are $\sigma^2_{\mathrm{g(SNP)}}$ (or σ^2_{u}) and σ^2_{e}. The matrix $\mathbf{A}$ describes the variance-covariance structure of the

488 random effects $\mathbf{g}$, and is assumed to be known in the estimation process. In practice, $\mathbf{A}$ is called

489 the SNP-derived genetic (or genomic) relationship matrix (GRM) and is estimated from the SNP

490 data. The estimate of $\sigma^2_{\mathrm{g(SNP)}}$ from GREML can be described as the estimated variance explained

491 by all the SNPs ($m\hat{\sigma}^2_{\mathrm{u}}$) or equivalently as the estimated genetic variance by contrasting the

492 phenotypic similarity between unrelated individuals to their SNP-derived genetic

493 similarity[13,58,59].

494

495

496 **Box 2. Future applications of SNP-based heritability to large datasets**

497 The methods for estimating h^2_{SNP} can be extended and applied to large data sets in the future.
498 These future directions include

499 i) Applications of GREML-LDMS or similar approaches (that account for bias in $\hat{h}^2_{\mathrm{SNP}}$ due to LD
500 bias) to in-depth WGS data to obtain nearly unbiased estimates of h^2 for a range of complex
501 traits and quantify the variance attributable to all rare variants;

502 ii) Methods that provide an unbiased estimate of h^2 from identity-by-descent information
503 inferred from SNP array data[60];

504 iii) Methods to estimate $\hat{h}^2_{\mathrm{ped}}$ from pedigree data accounting for common environmental effects
505 and assortative mating;

506 iv) Fast Bayesian MLM approaches based on flexible models that are applicable to WGS data to
507 estimate the distribution of effect sizes of all variants;

508 v) Methods to estimate h^2_{SNP} free of assumptions about the relationship between per-allele effect
509 and allele frequency[43].

510

511

512 **Figure Legends**

513

514 **Figure 1. Interpretation of estimated genetic variance depends on ascertainment of the**
515 **sample.** Shown in red are pedigree-based heritability estimate ($\hat{h}^2_{\text{ped}}$) for height from 2,824
516 pairs of full siblings in the UK Biobank (UKB) data [61] ("5k related" on the left; sibling correlation
517 = 0.520), $\hat{h}^2_{\text{SNP}}$ from a GREML analysis of 35,000 unrelated UKB individuals using all the
518 genotyped SNPs ("35k unrelated" on the right), and the estimates in between from GREML
519 analyses in a mixed sample of unrelated and close relatives (part 2 of the **Supplementary Note**).
520 The difference between $\hat{h}^2_{\text{ped}}$ and $\hat{h}^2_{\text{SNP}}$ demonstrates the genetic variation (due to rare variants
521 in particular) not tagged by genotyped SNPs and/or confounding in $\hat{h}^2_{\text{ped}}$ from common
522 environmental effects and non-additive genetic variation. Shown in green are the results from
523 the same analyses as above for a simulated phenotype based on a common environmental model
524 without genetic effect (part 2 of the **Supplementary Note**). Error bars indicate the standard
525 errors of the estimates.

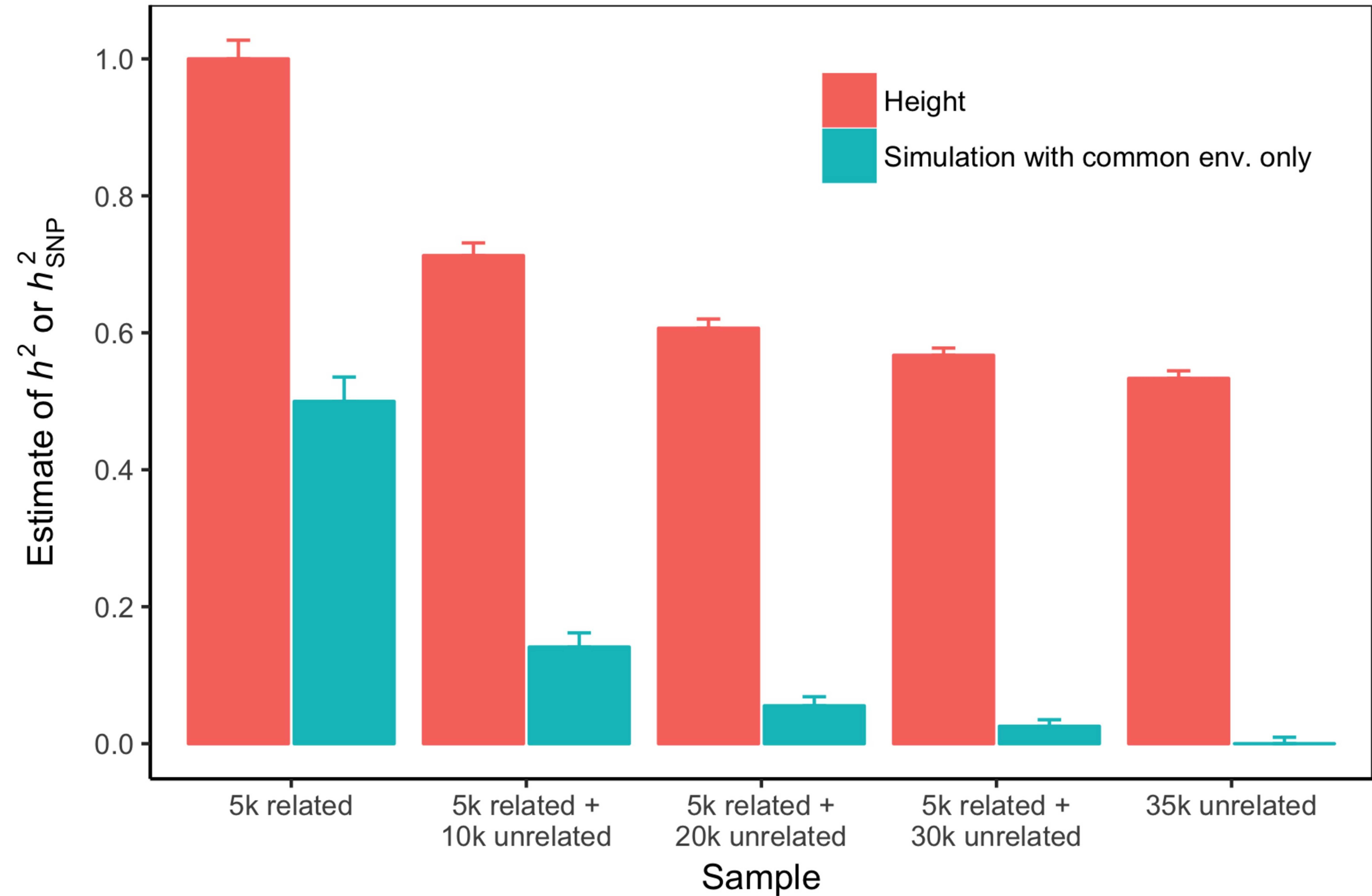
526

527 **Figure 2. Relationship between SNP-heritability on the liability scale ($h^2_{SNP(l)}$) and SNP-**
528 **heritability estimated from case-control samples.** The figures show that the same estimate of
529 $h^2_{\text{SNP}(l)}$ a) 0.1, b) 0.2, c) 0.4, d) 0.6 on the liability scale can correspond to a wide range of SNP
530 heritability estimates from case-control samples on the observed 0-1 scale (part 5 of the
531 **Supplementary Note**), depending on the proportion of cases in the sample (P) and the assumed
532 lifetime risk of disease (K) used to transform the estimates to the liability scale. For each plotted
533 line the minimum value assumes a population sample with $P = K$. In real-application we advise
534 investigating the sensitivity of estimates of $h^2_{SNP(l)}$ to choice of K, but we find that the impact is
535 small when $K < 0.05$. As shown in the bottom panels of the figure, for a rare disease with high
536 $h^2_{\text{SNP}(l)}$, $h^2_{\text{SNP}(O)}$ is expected to be larger than 1 because of the non-linear relationship between
537 genetic variance and phenotypic variance on the observed 0-1 scale.

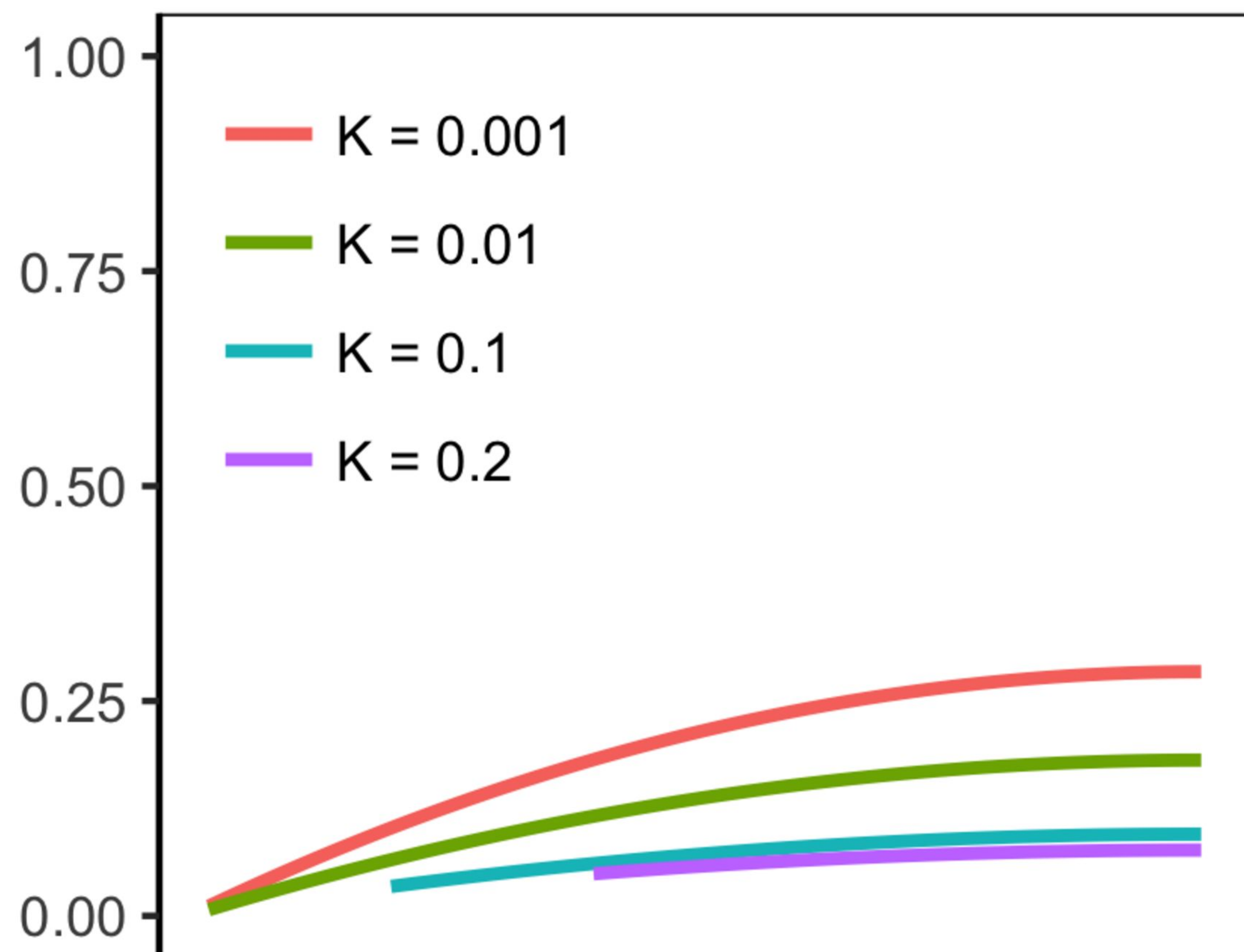
538

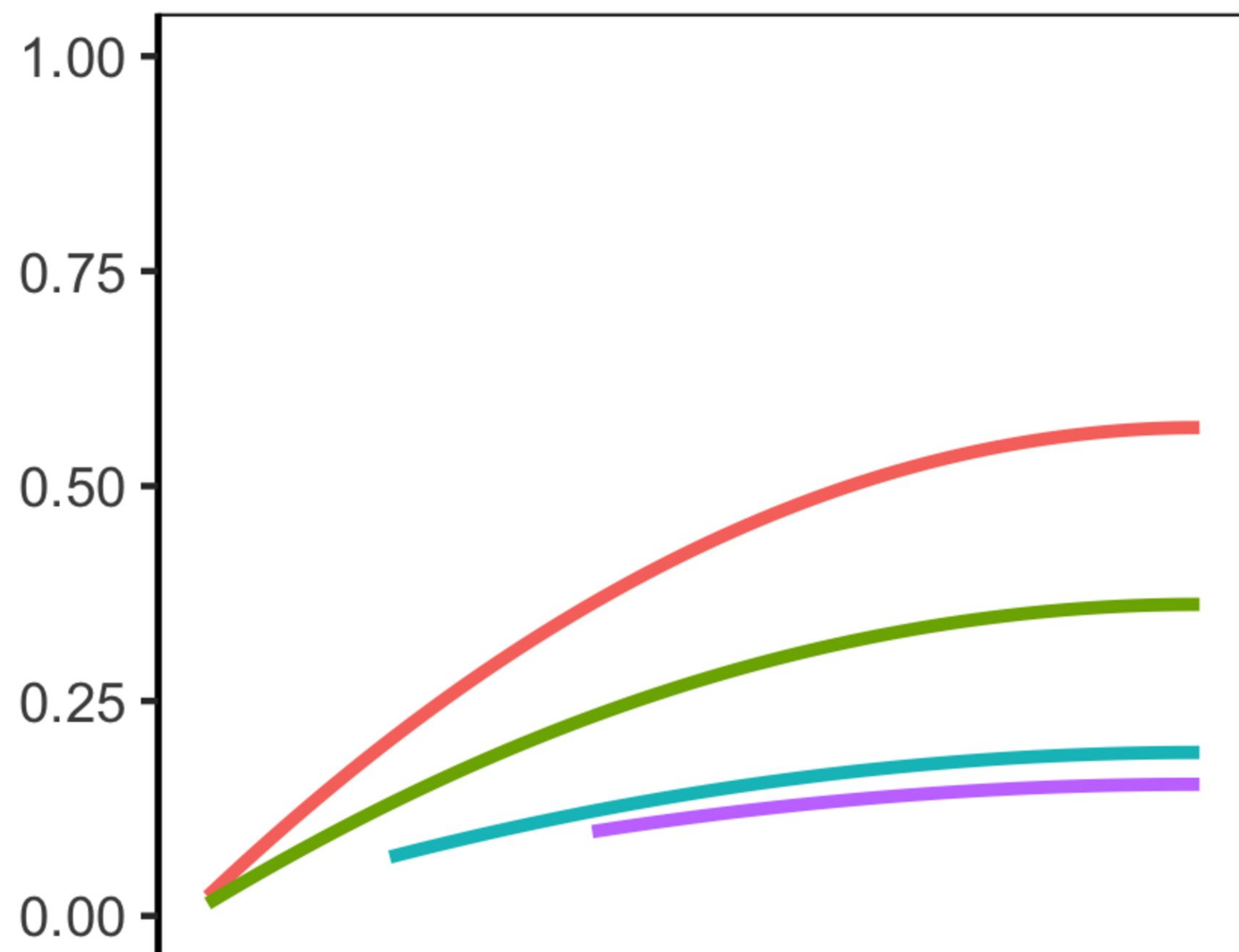
539 **Figure 3. Estimation of genetic variance depends on ascertainment of SNPs and genetic**
540 **architecture.** In panel (a), shown are the estimates of h^2_{SNP} using SNPs on six different SNP
541 panels for a simulated traits under two scenarios: 1) causal variants are random with both
542 common and rare variants (red) and 2) causal variants are rare (green) (see part 7 of the
543 **Supplementary Note** for details of the simulation). The six SNP panels are Affymetrix 6.0 array
544 (affy6), Affymetrix Axiom array (affyAxiom), HapMap 3 project (HM3), Illumina OmniExpress
545 (illu1M), Illumina Omni2.5 (illu2M) and Illumina CoreExome (illuCoreE). In panel (b), we show
546 the effect of LD pruning on $\hat{h}^2_{\text{SNP}}$ and the likelihood ratio test (LRT) statistic. LD pruning was

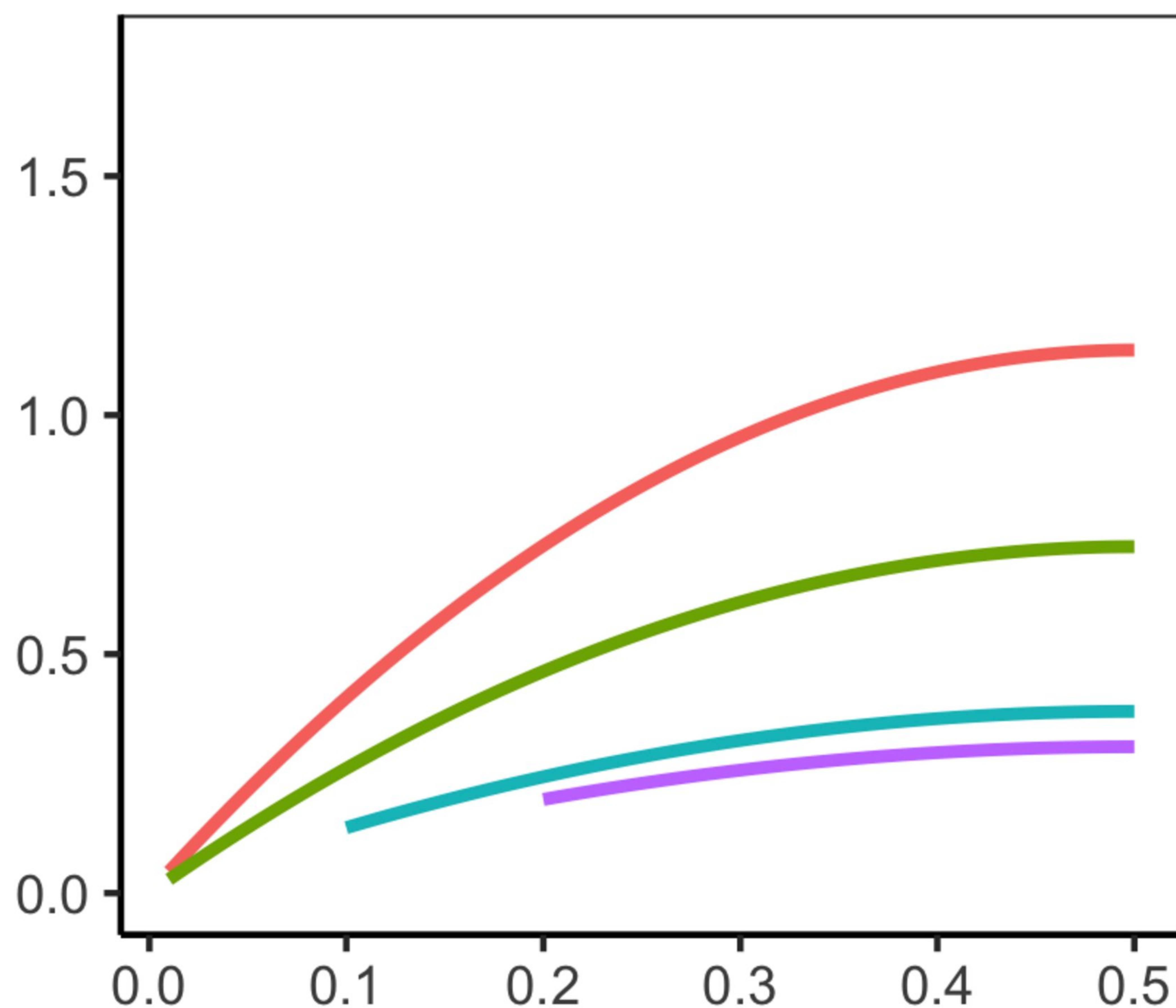
547 performed based on HM3 SNPs in PLINK (--indep-pairwise 50 5 r^2) with the LD r^2 threshold
548 shown on the x-axis. The last column with r^2 threshold of 1 represents the result without LD
549 pruning (i.e. all HM3 SNPs). The GREML analyses were performed using common SNPs on HM3.
550 Shown in panel (c) is the distribution of MAF of HM3 variants after LD pruning with different r^2
551 thresholds (no pruning for r^2 threshold of 1.0).
552
553 **Figure 4. Multiple component GREML or HE regression for sets of SNPs stratified by MAF.**
554 Results are $\hat{h}^2_{\mathrm{SNP}}$ with its SE (error bar) in each MAF group averaged over 200 simulation
555 replicates using ~11,500 unrelated individuals (SNP-based relatedness < 0.05) and ~550,000
556 genotyped SNPs after standard quality controls. In each simulation replicate, 1,000 SNPs were
557 selected at random as causal variants with their effects sampled from a standard normal
558 distribution with mean 0 and variance 1. The true heritability was 0.5 (roughly 0.1 per MAF bin).
559 The SE of the estimate from HE regression was calculated using the Jackknife approach where
560 one individual was left out at a time.

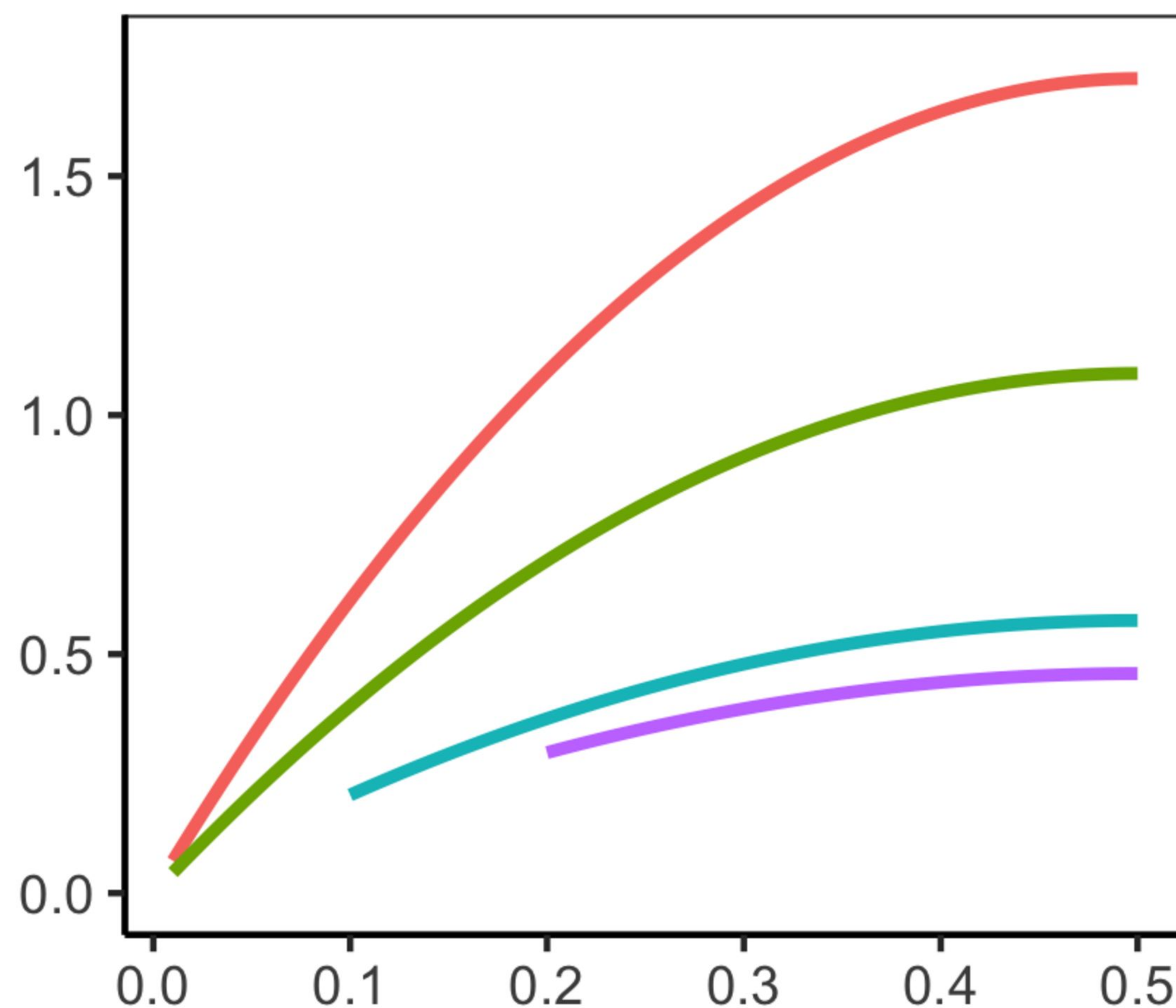


Height
Simulation with common env. only
Estimate of h^2 or h^2_{SNP}
1.0
0.8
0.6
0.4
0.2
0.0
5k related
5k related + 10k unrelated
5k related + 20k unrelated
5k related + 30k unrelated
35k unrelated
Sample

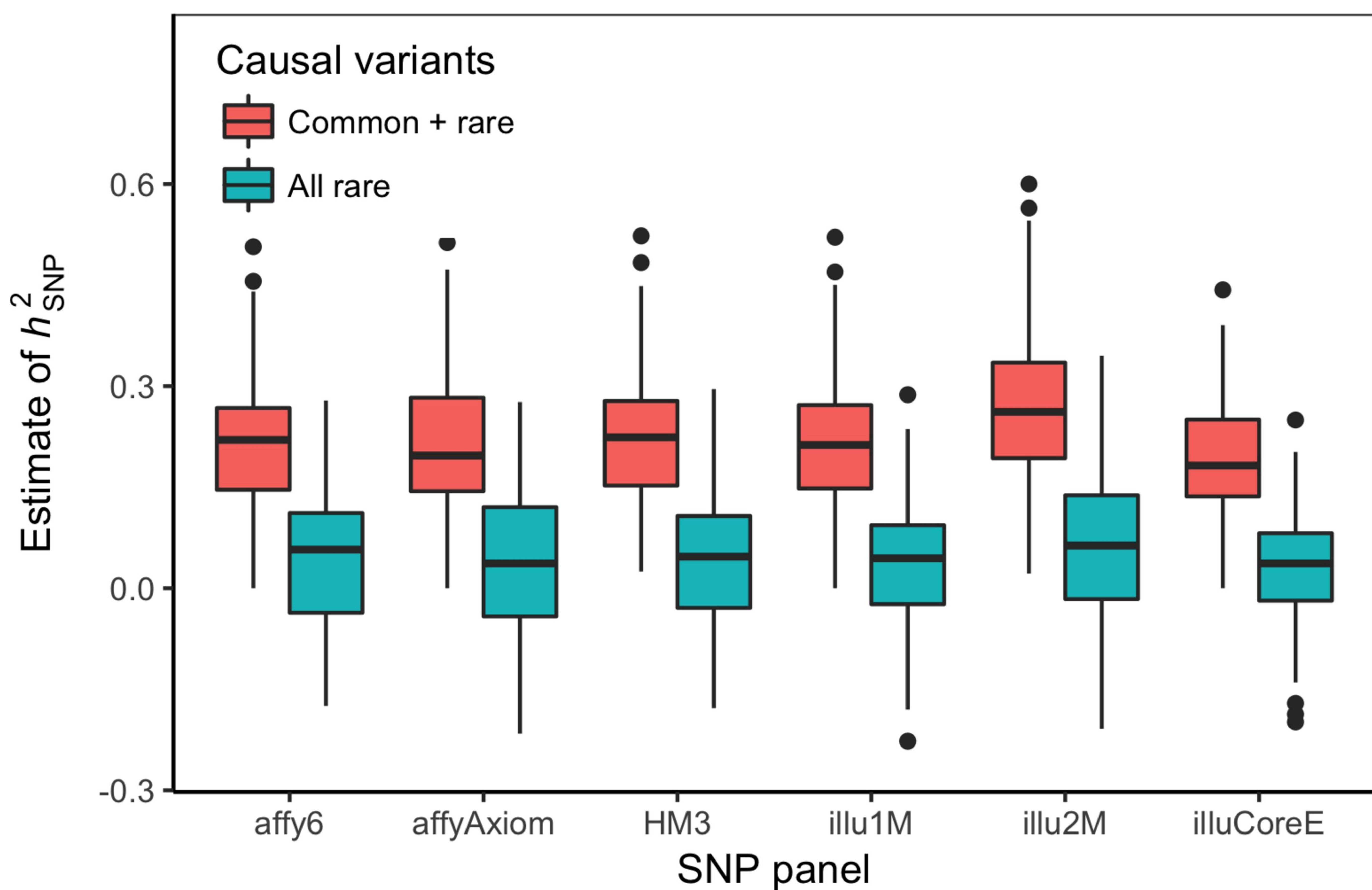

$h^2_{SNP(l)} = 0.1$
1.00
0.75
0.50
0.25
0.00
K = 0.001
K = 0.01
K = 0.1
K = 0.2


$h^2_{SNP(l)} = 0.2$
1.00
0.75
0.50
0.25
0.00

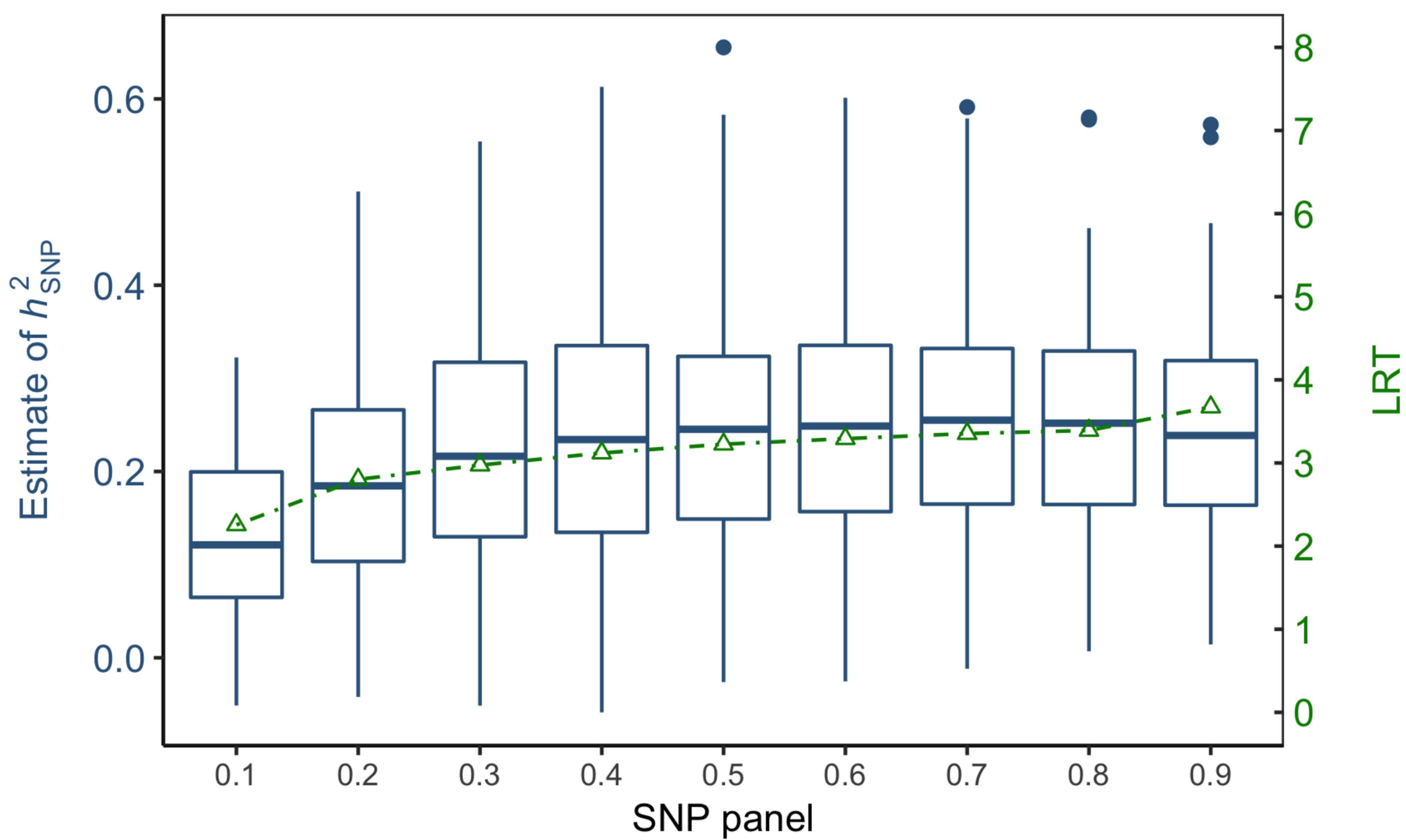

$h^2_{SNP(l)} = 0.4$
1.5
1.0
0.5
0.0
0.0
0.1
0.2
0.3
0.4
0.5


$h^2_{SNP(l)} = 0.6$
1.5
1.0
0.5
0.0
0.0
0.1
0.2
0.3
0.4
0.5
h^2_{SNP} estimated on 0/1 case-control scale
Proportion of sample that are cases

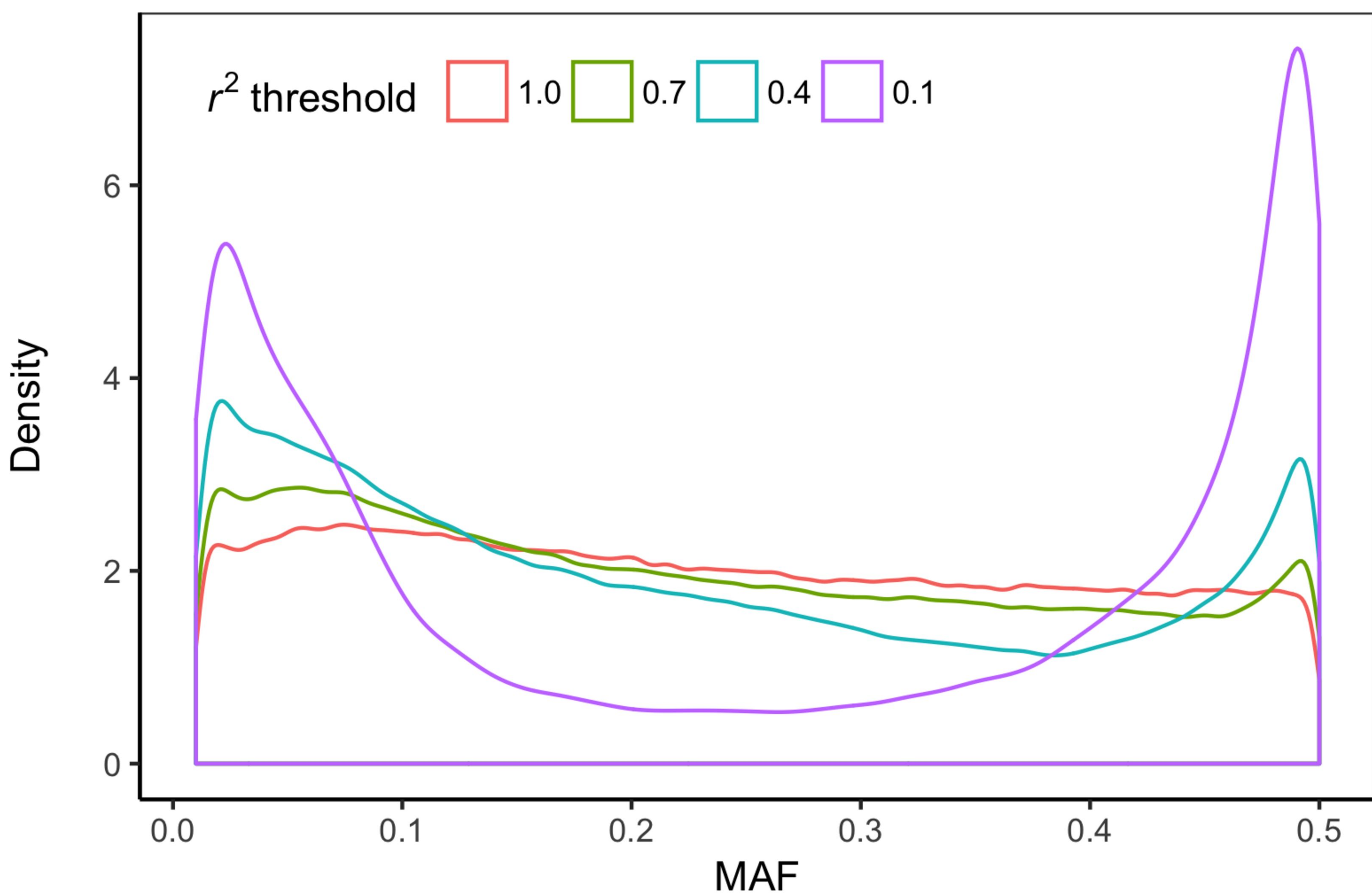
a

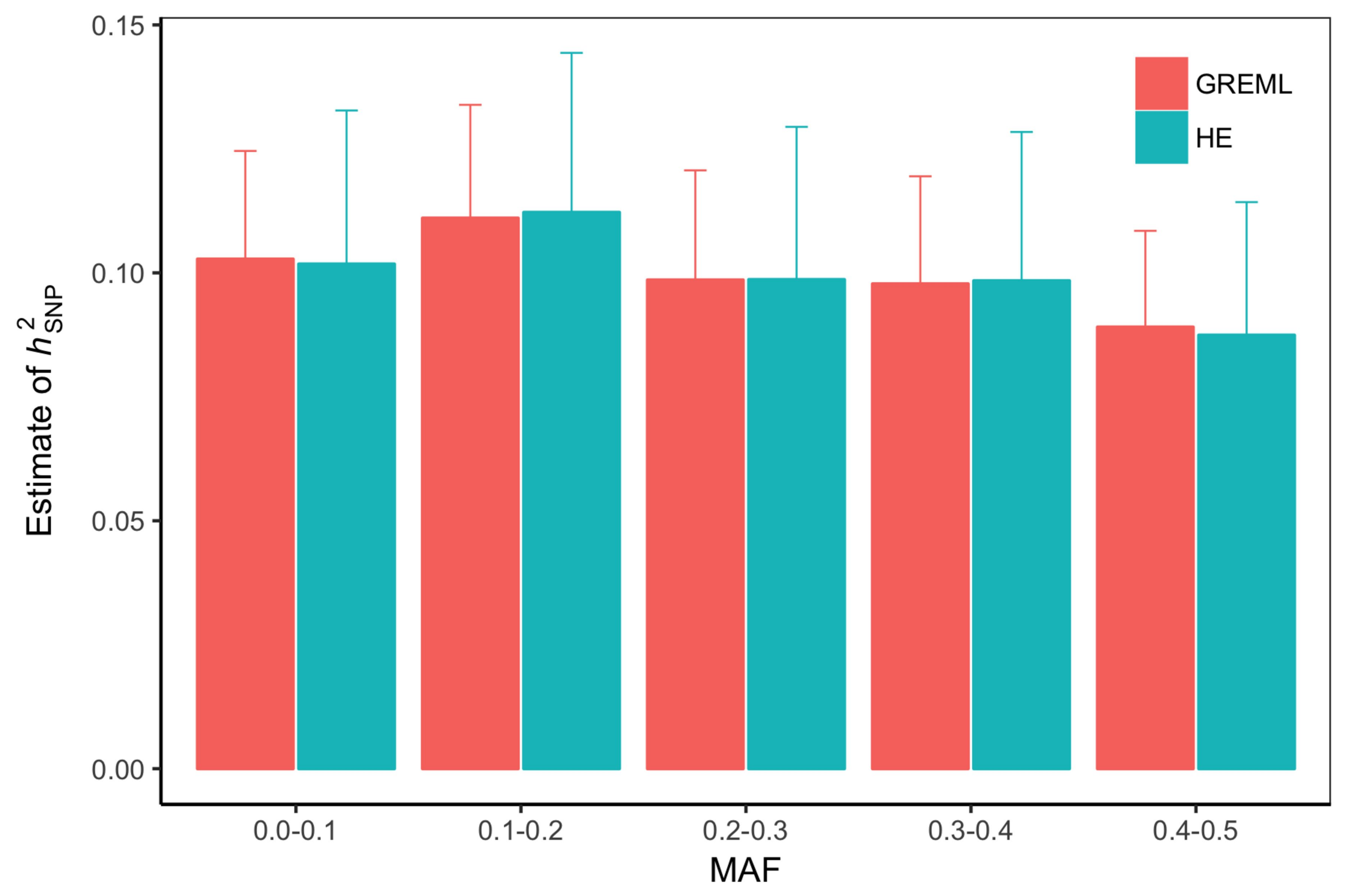


b



c




0.15
0.10
0.05
0.00
Estimate of h^2_{SNP}
GREML
HE
0.0-0.1
0.1-0.2
0.2-0.3
0.3-0.4
0.4-0.5
MAF