

A powerful replicability analysis of genome-wide association studies

Yan Li¹, Haochen Lei², Xiaoquan Wen³ and Hongyuan Cao^{2*}

Abstract

Replicability is the cornerstone of modern scientific research. Reliable identifications of genotype-phenotype associations that are significant in multiple genome-wide association studies (GWASs) provide stronger evidence for the findings. Current replicability analysis relies on the independence assumption among single nucleotide polymorphisms (SNPs) and ignores the linkage disequilibrium (LD) structure. We show that such a strategy may produce either overly liberal or overly conservative results in practice. We develop an efficient method, ReAD, to detect replicable SNPs associated with the phenotype from two GWASs accounting for the LD structure. The local dependence structure of SNPs across two heterogeneous studies is captured by a four-state hidden Markov model (HMM) built on two sequences of p -values. By incorporating information from adjacent locations via the HMM, our approach provides more accurate SNP significance rankings. ReAD is scalable, platform independent and more powerful than existing replicability analysis methods with effective false discovery rate (FDR) control. Through analysis of datasets from two asthma GWASs and two ulcerative colitis

¹School of Mathematics, Jilin University, Changchun, Jilin 130012, China.

²Department of Statistics, Florida State University, Tallahassee, FL 32306, USA.

³Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA.

*Corresponding author: hongyuancao@gmail.com

19 GWASs, we show that ReAD can identify replicable genetic loci that existing methods
20 might otherwise miss.

21 **1 Introduction**

22 Genome-wide association studies (GWASs) allow for simultaneous study of millions of single
23 nucleotide polymorphisms (SNPs). Numerous genetic risk variants associated with various
24 phenotypes and complex diseases have been reported over the past couple of decades (Mc-
25 Carthy et al., 2008; MacArthur et al., 2017). These associations provide insights into the archi-
26 tecture of disease susceptibility. Despite these progresses, many reported genotype-phenotype
27 associations fail to replicate in other studies (Ioannidis et al., 2001; Chanock et al., 2007).
28 An analysis of past studies indicates that the cumulative prevalence of irreproducible preclini-
29 cal research (including GWAS) exceeds 50% (Ioannidis, 2005; Prinz et al., 2011; Begley and
30 Ellis, 2012; Freedman et al., 2015). Approximately 28 billion annually is spent on preclinical
31 research that is not replicable in the United States alone (Freedman et al., 2015). Irrepli-
32 cable and/or inconsistent between-study associations might be spurious findings caused by
33 confounding factors, such as population stratification, misclassification of phenotypes, geno-
34 typing errors, or technical biases, among others. Replicability is now considered a *sine qua non*
35 for establishing credible genotype-phenotype associations in the era of GWAS (Moonesinghe
36 et al., 2008; Huffman, 2018). We study conceptual replicability where consistent results are
37 obtained using different processes and populations that target the same scientific question.
38 For GWASs, replicability analysis aims to detect genetic risk loci that are significantly as-
39 sociated with the same phenotype across different studies (Heller and Yekutieli, 2014; Heller
40 et al., 2014; Bogomolov and Heller, 2022). By eliminating genetic associations that can not
41 be generalized across studies, replicability analysis provides stronger support for genuine sci-
42 entific findings, avoids wasted resources, and improves efficiency of drug development. This

43 helps the translation of bench discoveries to bedside therapies.

44 In GWASs, millions of SNPs are tested simultaneously, requiring multiple testing adjust-
45 ment. False discovery rate (FDR), defined as the expectation of the proportion of false discov-
46 eries over total discoveries, is a commonly used metric for type I error control (Benjamini and
47 Hochberg, 1995). A central characteristic of GWAS data is the linkage disequilibrium (LD)
48 among SNPs, with which alleles at nearby sites can co-occur on the same haplotype more
49 often than by chance alone (Pritchard and Przeworski, 2001; Wall and Pritchard, 2003). As
50 a result, it is common to observe that phenotype-associated SNPs form clusters and exhibit
51 high correlations within clusters (Wei et al., 2009). An effective approach to account for the
52 LD structure among SNPs is through the hidden Markov model (HMM) (Churchill, 1992).
53 Existing GWAS literature (Sun and Cai, 2009; Wei et al., 2009) using HMM for a single study
54 is not applicable to replicability analysis of multiple studies. Furthermore, their approaches
55 cannot be generalized to more than one study due to the heterogeneity of LD across different
56 studies (Lonjou et al., 2003). Replicability analysis of GWASs explicitly accounting for the
57 LD structure has not been studied before to the best of our knowledge.

58 To claim replicability, an *ad hoc* approach is to implement an FDR control method, such
59 as the Benjamini and Hochberg (BH) procedure (Benjamini and Hochberg, 1995), for each
60 study and intersect significant results from all studies as replicable findings. This approach
61 does not control the FDR and moreover has low power as it does not borrow information
62 from different studies. The maximum of p -values across studies (P_{\max}) is a straightforward
63 significance measure for replicability (Benjamini et al., 2009). After summarizing data from
64 multiple studies by P_{\max} , classic FDR control procedures such as BH are used for replicability
65 analysis. This procedure is overly conservative as it guards against the worst scenario and does
66 not incorporate the composite null structure of replicability analysis. For independent features
67 from high-throughput experiments, various methods were proposed for replicability analysis.
68 These methods are not robust to heterogeneity of different studies (Li et al., 2011; Philtron

69 et al., 2018), require tuning parameters (Zhao et al., 2020), impose parametric assumptions
70 on the p -values (Heller et al., 2014) or demand access to full datasets which can be prohibitive
71 due to privacy concerns or logistics (McGuire et al., 2021).

72 We address the limitations of existing methods by developing an efficient method, ReAD
73 (Replicability Analysis accounting for Dependence) to detect replicable genotype-phenotype
74 associations across two GWASs by incorporating the LD structure. We use GWAS summary
75 statistics such as p -values, treating multiple studies symmetrically. Our approach models the
76 clustered signals from two studies with a four-dimensional HMM accounting for the hetero-
77 geneity of LD structures in different studies. Conditional on the HMM, we model the two
78 p -value sequences as a four-group mixture of SNPs (Efron, 2012; Chung et al., 2014). The
79 replicability null hypothesis consists of three components: zero effects in both studies, zero
80 effect in one study and non-zero effect in another study and vice versa. ReAD calculates the
81 posterior probability of replicability null given data. Compared to other replicability analysis
82 methods, ReAD is robust as it is non-parametric, jointly models the signal and non-signal
83 from different studies, and accounts for the heterogeneity of different studies. ReAD pro-
84 vides more efficient rankings of importance for replicable SNPs by pooling information from
85 two p -value sequences via the forward and backward probabilities (Rabiner and Juang, 1986;
86 Murphy, 2012). ReAD applies a step-up procedure to identify clusters of genotype-phenotype
87 associated signals, improving the power of replicability analysis while effectively controlling
88 the FDR. ReAD is computationally scalable to whole genome with tens of millions of SNPs.
89 Its implementation combines the non-parametric expectation-maximization (EM) algorithm
90 (Dempster et al., 1977) and the pool-adjacent-violator algorithm (PAVA) in shape constraint
91 inference (Robertson et al., 1988; Busing, 2022), without any tuning parameters. We conduct
92 extensive simulation studies to evaluate the performance of our approach across a wide range
93 of scenarios. By applying our procedure to summary statistics of two asthma GWASs and
94 two ulcerative colitis GWASs, we show that ReAD identifies more replicable genetic loci that

95 otherwise might be missed using existing methods that do not account for the LD structure.
 96 These identified association signals pinpoint potential new loci on metabolisms and immunity.

97 2 Results

98 2.1 Method overview

99 ReAD takes p -values from two independent GWASs with the same phenotype as input. Sup-
 100 pose we have J SNPs with corresponding p -values $(p_{1j}, p_{2j}), j = 1, \dots, J$. We aim to identify
 101 replicable SNPs associated with the phenotype in both studies. Our method can handle SNPs
 102 in the whole genome where J is in the order of millions. We use θ_{ij} to represent the inferred
 103 association status of SNP j in study i . For each SNP, we consider its association analysis

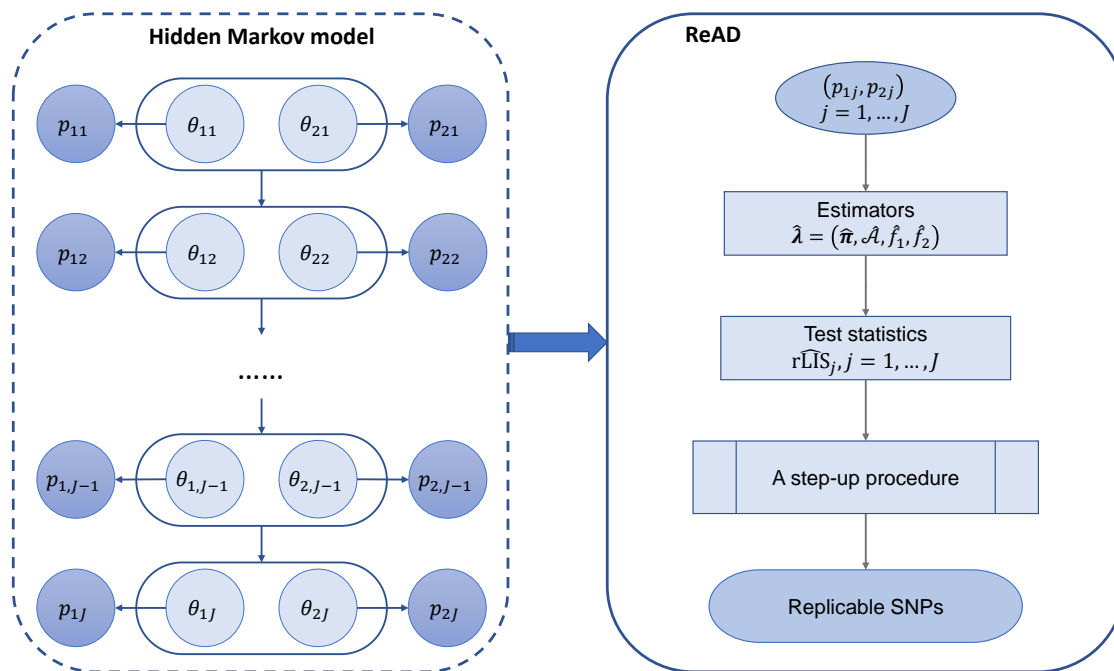


Figure 1: Schematic of ReAD. θ_{ij} represents the inferred association status of SNP j ($j = 1, \dots, J$) in study i ($i = 1, 2$). For each SNP j , we consider its association analysis results are replicable if $\theta_{1j} = \theta_{2j} = 1$. The dependence structure among SNPs across two studies can be modeled with a HMM.

104 results are replicable if its corresponding θ values are consistently 1. The correlations between
105 θ 's within a study are caused by LD among tested SNPs, and we model their dependence
106 structure using a Markov chain. Following Li and Stephens (2003), this is an effective way to
107 model the correlations between observed p -values. Given the observed p -values are from both
108 studies, the overall model structure can be represented by a HMM.

109 We present our schematic in Figure 1. We use $s_j \in \{0, 1, 2, 3\}$ to denote the joint inferred
110 association status for SNP j , where $s_j = 0$ if $\theta_{1j} = \theta_{2j} = 0$, $s_j = 1$ if $\theta_{1j} = 0$ and $\theta_{2j} = 1$,
111 $s_j = 2$ if $\theta_{1j} = 1$ and $\theta_{2j} = 0$, and $s_j = 3$ if $\theta_{1j} = \theta_{2j} = 1$. The composite null for replicability
112 analysis corresponds to $s_j \in \{0, 1, 2\}$. To capture the local dependence of LD structure among
113 SNPs, we impose a four-state HMM on $\mathbf{s} = (s_1, \dots, s_J)$. The transition matrix is denoted
114 as $\mathcal{A} = \{a_{kl} : k, l = 0, 1, 2, 3\}$, where the transition probability from $s_j = k$ to $s_{j+1} = l$ is
115 given by a_{kl} , and $\sum_{l=0}^3 a_{kl} = 1$ for all k . An efficient EM algorithm in combination with
116 the forward-backward procedure and PAVA is developed to estimate the unknown parameters
117 and functions. We use the posterior probability of being replicability null, $\text{rLIS}_j, j = 1, \dots, J$,
118 as the test statistic and obtain $\widehat{\text{rLIS}}_j$ for all SNPs. By applying a step-up procedure on
119 $\widehat{\text{rLIS}}_j, j = 1, \dots, J$, we get powerful testing results while controlling the FDR. More details of
120 ReAD can be found in the Methods Section and the Supplemental Note A.

121 2.2 Simulation study

122 2.2.1 Simulation I

123 In simulation I, we evaluated the FDR and statistical power of ReAD based on the rLIS
124 statistic across two studies. Here power is defined as the averaged proportion of true discoveries
125 among the total number of non-null hypotheses. We compare the FDR and power of ReAD
126 with several replicability analysis methods developed under independence, including the *ad*
127 *hoc* BH method, the MaxP method based on P_{\max} (Benjamini et al., 2009) and the STAREG

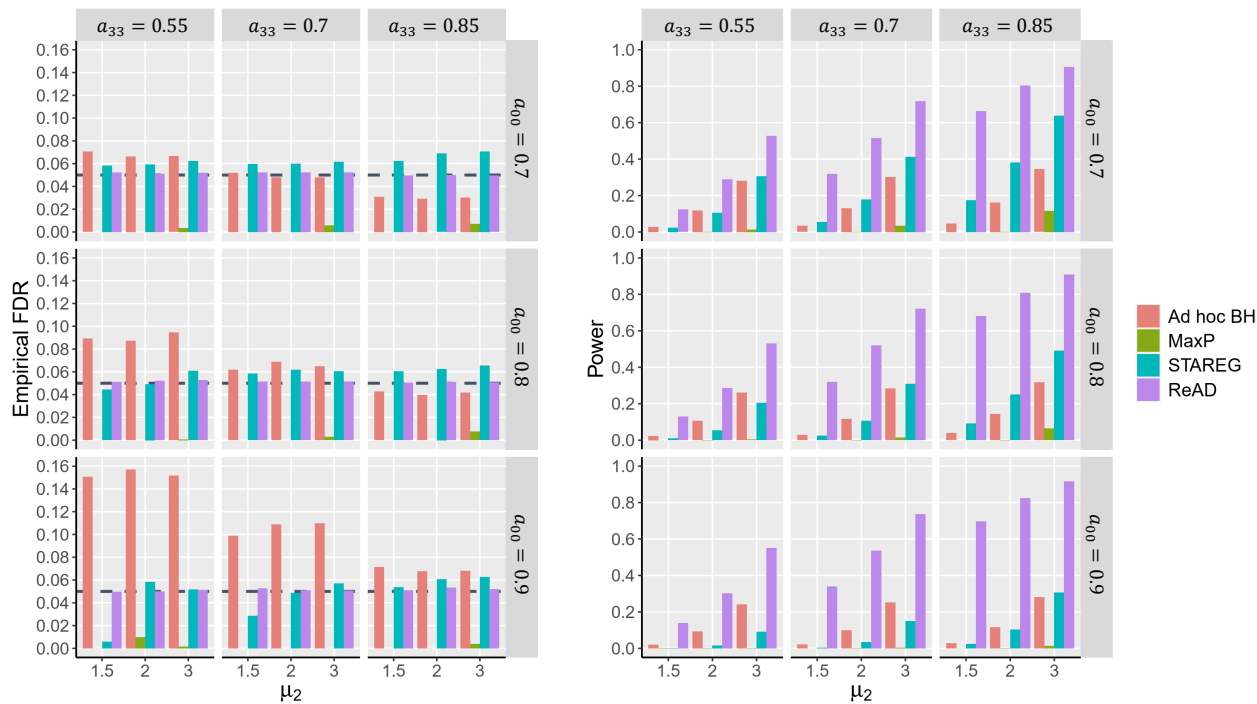


Figure 2: FDR control and power comparison of different methods.

128 method based on the local false discovery rate (Lfd_r) (Li et al., 2023). Details of these methods
 129 can be found in the Supplemental Note B. An extensive comparison with more replicability
 130 analysis methods can be found in Supplemental Note C.

131 In each simulation, the hidden states of 10,000 SNPs were generated from a four-state
 132 Markov chain. A detailed description of the data generating process is provided in Sup-
 133 plemental Note C. In all simulations, we fix the initial distribution of four states as $\boldsymbol{\pi} =$
 134 $(0.9, 0.025, 0.025, 0.05)$. The signals from two studies are generated from normal distributions
 135 with mean μ_i and variance $\sigma_i^2, i = 1, 2$. We vary the transition matrix $\mathcal{A} = \{a_{kl} : k, l =$
 136 $0, 1, 2, 3\}$ and μ_2 while fix $\mu_1 = 2$, and $\sigma_1 = \sigma_2 = 1$. Empirical FDR and power are calculated
 137 from 100 replications for each setting. The results are summarized in Figure 2 (left: FDR;
 138 right: power). In Figure 2, each row corresponds to a different a_{00} , and each column corre-
 139 sponds to a different a_{33} . In each panel, we set μ_2 to 1.5, 2, or 3. At FDR level 0.05, we see
 140 that the *ad hoc* BH fails to control the FDR. MaxP is overly conservative across all settings.

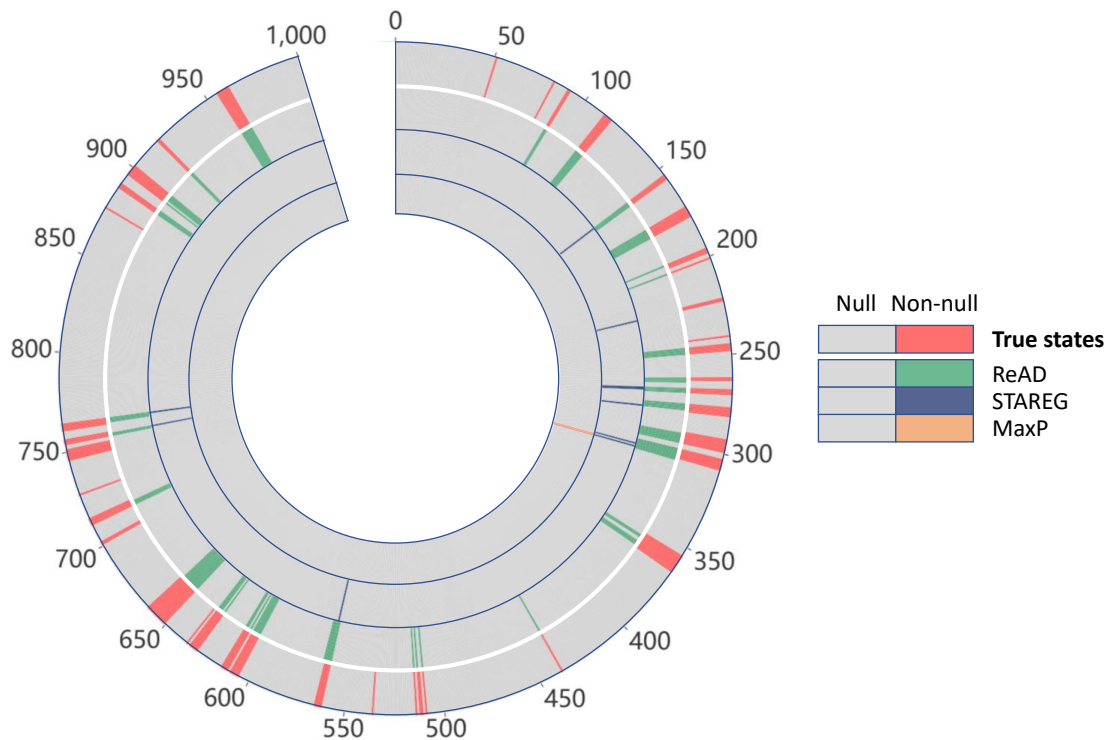


Figure 3: Methods comparison for cluster identification. Circles range from 1 (the outermost circle) to 4 (the innermost circle). The outermost circle represents true states; circle 2 represents ReAD, circle 3 represents STAREG and circle 4 represents MaxP.

141 STAREG has a slight FDR inflation in some settings. By accounting for the local dependence
142 structure via the rLIS statistic, ReAD properly controls the FDR and has substantial power
143 gain compared to competing methods. The powers of all methods increase as μ_2 increases.

144 The forward-backward procedure of HMM implies that a small rLIS does not occur alone,
145 but in clusters. Therefore, ReAD tends to identify the entire cluster of genotype-phenotype
146 associations. Such clusters are unlikely to occur by chance and are more plausible biological
147 signals. To illustrate this, p -values for two studies are generated following the above strategy
148 by setting $a_{00} = 0.9$, $a_{33} = 0.7$, and $\mu_2 = 2$. We compare three methods for testing the
149 composite replicability null hypotheses across two studies: the MaxP method (Benjamini et al.,
150 2009), the STAREG method (Li et al., 2023), and the ReAD method. Figure 3 presents results
151 of different methods in one replication. It can be seen that MaxP is extremely conservative,

152 which only identifies one single signal; STAREG rejects individual hypotheses with very small
153 p -values in both studies; whereas ReAD can identify clusters of replicable signals.

154 2.2.2 Simulation II

155 By incorporating the LD structure in GWASs through HMM, the rLIS statistic integrates
156 information from adjacent locations. Therefore, the rankings of SNPs based on rLIS are
157 different from the rankings from MaxP (based on P_{\max}) and STAREG (based on Lfdr). We
158 perform simulation studies to demonstrate different rankings in GWASs with realistic LD
159 patterns among SNPs. Data for two studies are generated based on two SNP matrices from the
160 Genetic European Variation in Disease project (Lappalainen et al., 2013). The CEU genotype
161 data are collected from 78 Utah residents with Northern and Western European ancestry,
162 and the FIN genotype data are measured from 89 Finnish in Finland. CEU and FIN are
163 both sub-populations of the European Ancestry population, therefore they may have similar

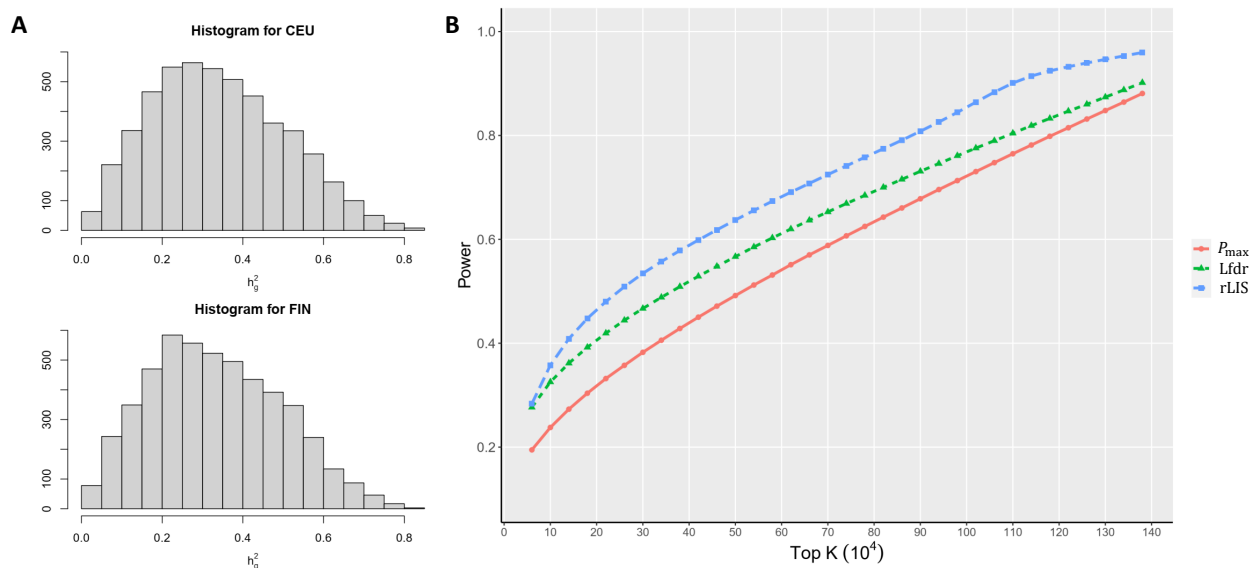


Figure 4: (A) Histogram of estimated h_g^2 for simulated CEU data and FIN data from 5,000 runs. (B) Power of top K SNPs identified by different methods. The power is calculated as the averaged percentages of true positives selected by the top K SNPs ranked based on rLIS, Lfdr produced by STAREG, and P_{\max} used in MaxP. The results are calculated from 100 runs.

164 LD structures. A detailed description of the strategy to generate continuous phenotypes
165 corresponding to SNPs in a single gene across two studies is provided in the Methods Section.
166 To make realistic simulations, we adjust the signal-to-noise ratio in two studies so that the
167 SNP heritability h_g^2 is centered between 0.2 and 0.3. Figure 4(A) presents the histogram of
168 h_g^2 for CEU and FIN studies generated from 5,000 replications. We repeat the above data
169 generating process to simulate GWAS data for 100 genes, resulting in 1,676,400 pairs of
170 p -values for corresponding SNPs. As in Wei et al. (2009), we define five adjacent SNPs on
171 each side of the 400 causal SNPs as relevant SNPs and evaluate the performance of different
172 replicability analysis methods by calculating the percentages of selecting relevant SNPs.

173 We average the percentages of true positives selected by the top K hits from 100 runs
174 as our evaluation criterion. The power curves of different K values based on SNP rankings
175 produced by P_{\max} , Lfdr and rLIS are depicted in Figure 4(B). We see that rLIS shows higher
176 power than P_{\max} and Lfdr, indicating that the rankings based on rLIS are more efficient than
177 the rankings based on P_{\max} and Lfdr in replicability analysis of GWAS data by incorporating
178 the LD block structure through HMM.

179 **2.3 Data analysis**

180 **2.3.1 Replicability analysis of asthma GWASs**

181 Asthma is a complex bronchial disease characterized by chronic inflammation and narrowing
182 of the airways, which is caused by a combination of environmental and genetic factors. The
183 prevalence of asthma varies across different populations and ethnicities. We implement ReAD
184 to conduct replicability analysis of asthma GWASs from the Trans-National Asthma Genetic
185 Consortium (TAGC) and UK Biobank. The results are compared with competing meth-
186 ods. Demenais et al. (2018) conducted ancestry-specific meta-analyses from ethnically-diverse
187 populations and deposited the HapMap2-imputed data in the TAGC consortium. The TAGC

188 asthma GWAS data with high-density genotyped and imputed SNP based on the European-
189 ancestry comprises 8,843,303 genetic variants for 19,954 asthma cases and 107,715 controls.
190 UK Biobank is a large-scale prospective cohort study with over half a million participants
191 aged 40-69 years from the United Kingdom between 2006 and 2010 (Sudlow et al., 2015). The
192 imputed asthma GWAS from UK Biobank contains summary statistics for 8,856,162 genetic
193 variants measured on 39,049 self-reported asthma cases and 298,070 controls. We filter out
194 SNPs with minor allele frequency (MAF) smaller than 0.05, resulting in 6,234,241 SNPs
195 in the TAGC study and 6,242,120 SNPs in UK Biobank. After taking the intersection of
196 SNPs in the two studies, we obtain paired p -values of 6,222,195 SNPs to conduct replicability
197 analysis.

198 As the *ad hoc* BH does not control FDR, we apply MaxP and STAREG on the paired p -
199 values for comparison. The GWAS Catalog (Welter et al., 2014) reported cytogenetic regions
200 (loci) associated with asthma. To assess the replicability of GWAS loci, we state that if at least
201 one of the identified SNPs falls into one of the regions, the locus is identified as replicable. If a
202 locus contains multiple significant SNPs, the SNP with the strongest association is considered
203 as the lead SNP. For instance, if we use STAREG with Lfdr as the test statistic, the SNP
204 with the smallest Lfdr is the lead SNP.

205 At FDR level 5×10^{-8} , MaxP identifies 2,853 significant SNPs in 10 loci, which are also
206 identified by STAREG and ReAD. Compared to MaxP, STAREG identifies 909 additional
207 significant SNPs in 3 loci. By capturing the local LD structure through HMM, ReAD identifies
208 10,084 significant SNPs in 28 genetic loci with replicable asthma associations, of which 15
209 loci are not detected by MaxP or STAREG. Figure 5 presents the Manhattan plots of MaxP,
210 STAREG, and ReAD. In Figure 5, the vertical axis are $-\log_{10}$ transformations of test statistics
211 for replicability analysis, i.e., P_{\max} for MaxP, Lfdr for STAREG, and rLIS for ReAD.

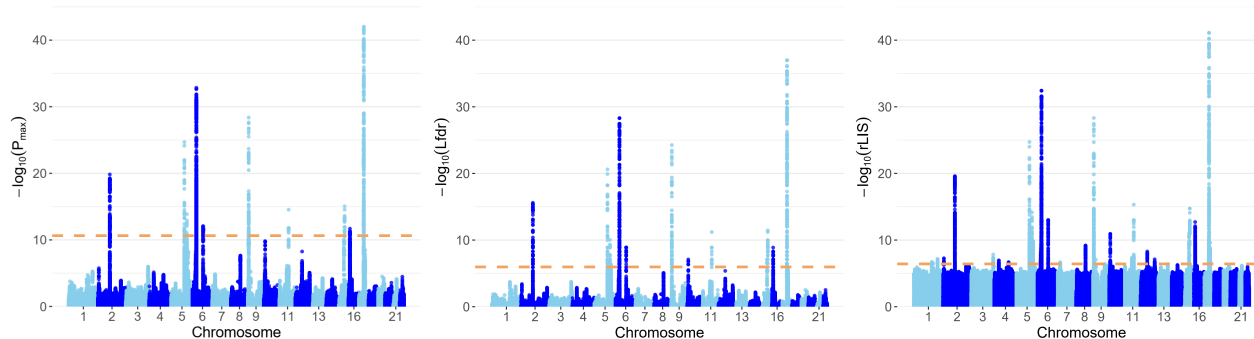


Figure 5: The Manhattan plots based on P_{\max} , Lfdr and rLIS. The dashed horizontal lines denote the FDR cutoffs of 5×10^{-8} produced by MaxP, STAREG, and ReAD, respectively.

Table 1: Main characteristics of the 28 loci associated with asthma in the European-ancestry TAGC and UK Biobank GWASs identified by ReAD. The SNP with the strongest association within each locus is called Lead SNP. The mapped gene denotes genes overlapping or closest to the lead SNP in the identified locus.

| Locus | Lead SNP | Location of lead SNP | Mapped gene | P_{\max} | Lfdr | rLIS |
|---|------------|----------------------|--------------------------|------------|---------|---------|
| Replicable asthma loci identified by all methods | | | | | | |
| 2q12.1 | rs3771180 | chr2:102,337,157 | <i>IL18R1,IL1RL1</i> | 1.5e-20 | 2.5e-16 | 2.5e-20 |
| 5q22.1 | rs10455025 | chr5:111,069,301 | <i>BCLAF1P1,TSLP</i> | 2.0e-25 | 2.4e-21 | 1.9e-25 |
| 5q31.1 | rs20541 | chr5:132,660,272 | <i>IL13,TH2LCRR</i> | 1.4e-14 | 9.1e-11 | 7.0e-15 |
| 6p21.32 | rs17843604 | chr6:32,652,506 | <i>HLA-DQA1,HLA-DQB1</i> | 2.2e-33 | 5.0e-29 | 3.8e-33 |
| 6p21.33 | rs2596465 | chr6:31,445,171 | <i>LINC01149</i> | 1.2e-14 | 4.1e-11 | 3.1e-15 |
| 6q15 | rs2325291 | chr6:90,276,967 | <i>BACH2</i> | 8.6e-13 | 1.3e-09 | 1.0e-13 |
| 9p24.1 | rs992969 | chr9:6,209,697 | <i>GTF3AP1,IL33</i> | 4.3e-29 | 5.5e-25 | 4.8e-29 |
| 11q13.5 | rs2155219 | chr11:76,588,150 | <i>LINC02757,EMSY</i> | 2.9e-15 | 6.3e-12 | 4.8e-16 |
| 15q22.33 | rs17228058 | chr15:67,157,967 | <i>SMAD3</i> | 2.9e-15 | 6.3e-12 | 1.8e-15 |
| 16p13.13 | rs12935657 | chr16:11,125,184 | <i>CLEC16A</i> | 2.1e-12 | 1.3e-09 | 2.0e-13 |
| Replicable asthma loci identified by ReAD and STAREG but not by MaxP | | | | | | |
| 5q31.3 | rs2338822 | chr5:142,123,494 | <i>NDFIP1</i> | 6.0e-09 | 1.2e-06 | 1.5e-10 |
| 10p14 | rs962993 | chr10:9,011,169 | <i>LINC02676</i> | 1.9e-10 | 1.5e-07 | 1.2e-11 |
| 15q22.2 | rs11071558 | chr15:60,777,222 | <i>RORA</i> | 8.3e-11 | 8.1e-08 | 9.9e-12 |

Replicable asthma loci only identified by ReAD

| | | | | | | |
|----------|------------|-------------------|--------------------------|---------|---------|---------|
| 1q32.1 | rs7555556 | chr1:203,121,848 | <i>ADORA1</i> | 5.5e-06 | 9.3e-04 | 6.9e-08 |
| 1q21.3 | rs4845623 | chr1:154,443,301 | <i>IL6R</i> | 1.5e-04 | 5.1e-04 | 2.7e-07 |
| 1q24.2 | rs864537 | chr1:167,442,147 | <i>CD247</i> | 3.2e-05 | 3.0e-03 | 3.2e-07 |
| 2p25.1 | rs10174949 | chr2:8,302,118 | <i>LINC00299</i> | 3.0e-06 | 5.7e-04 | 5.3e-08 |
| 3q28 | rs2889896 | chr3:188,384,928 | <i>LPP</i> | 1.0e-06 | 1.9e-04 | 1.5e-08 |
| 4p14 | rs6815814 | chr4:38,814,717 | <i>TLR1</i> | 1.2e-05 | 1.7e-03 | 1.3e-07 |
| 4q27 | rs1904522 | chr4:122,415,763 | <i>ADAD1</i> | 1.7e-05 | 2.5e-03 | 2.3e-07 |
| 6p22.1 | rs2523716 | chr6:30,202,748 | <i>TRIM26</i> | 1.4e-08 | 4.6e-06 | 3.5e-10 |
| 8q21.13 | rs10957979 | chr8:80,377,552 | <i>RNU6-1213P</i> | 2.3e-08 | 8.6e-06 | 6.5e-10 |
| 11q12.2 | rs174541 | chr11:61,798,436 | <i>FADS2</i> | 2.3e-06 | 5.6e-04 | 4.4e-08 |
| 12q13.3 | rs324014 | chr12:57,116,526 | <i>STAT6</i> | 2.7e-07 | 6.0e-05 | 5.1e-09 |
| 12q24.31 | rs625228 | chr12:120,840,463 | <i>SPPL3</i> | 9.0e-06 | 5.8e-04 | 7.7e-08 |
| 17q21.33 | rs17637472 | chr17:49,384,071 | <i>ZNF652,PHB</i> | 3.3e-09 | 3.3e-06 | 2.5e-10 |
| 17q21.32 | rs12949836 | chr17:49,271,490 | <i>FLJ40194</i> | 1.6e-07 | 6.0e-05 | 4.6e-09 |
| 17q21.2 | rs34349578 | chr17:42,446,111 | <i>ATP6V0A1,RNU7-97P</i> | 5.6e-05 | 1.0e-03 | 1.9e-07 |

212 Table 1 displays main characteristics of the 28 cytogenetic regions identified by ReAD.
213 The mapped gene denotes genes overlapping or closest to the lead SNP in the identified locus.
214 The 15 loci only identified by ReAD harbor signals closely related to asthma. For example,
215 the lead SNP in locus 2p25.1, rs10174949, is in the intron of gene *LINC00299* and plays an
216 important role in atopic dermatitis, including asthma, hay fever and eczema in European and
217 UK populations (Zhu et al., 2020, 2018; Ferreira et al., 2017). The 8q21.13 region is reported
218 to be associated with asthma and hay fever in a European-ancestry study (Ferreira et al.,
219 2014). The lead SNP rs6473226 lies between gene *MIR5708* (chr8:80,241,389–80,241,473)
220 and gene *RNU6-1213P* (chr8:80,405,516–80,405,609), and its association with asthma has
221 been observed in several European-ancestry studies (Demenais et al., 2018; Olafsdottir et al.,
222 2020).

223 2.3.2 Replicability analysis of ulcerative colitis GWASs

224 Inflammatory bowel disease is a chronic, relapsing intestinal inflammatory disease. It has the
225 highest age-standardized prevalence rate in the US followed by the UK (Alatab et al., 2020)
226 with increasing prevalence in Asia and developing countries (Molodecky et al., 2012). Ulcer-
227 ative colitis (UC) is one of the two main forms of inflammatory bowel disease. We conduct
228 replicability analysis of GWASs from the International Inflammatory Bowel Disease Genet-
229 ics Consortium (IIBDGC) and the UK Biobank. The IIBDGC GWAS analyses 8,857,076
230 SNPs from 6,968 UC cases and 20,464 population controls of European descent (Liu et al.,
231 2015). The imputed UK Biobank GWAS data contain summary statistics of 8,856,162 SNPs
232 genotyped on 1,795 self-reported UC cases and 335,324 controls from the United Kingdom.
233 We filter out SNPs with MAF smaller than 0.05, resulting in 6,243,744 SNPs in the IIBDGC
234 study and 6,242,120 SNPs in the UK Biobank. We use the paired p -values of 6,232,147 SNPs
235 common to both studies as input for replicability analysis.

236 We apply MaxP, STAREG, and ReAD on the paired p -values. At FDR level 5×10^{-8} ,
237 MaxP identifies 1,239 significant SNPs in 1 locus. STAREG identifies 1,542 significant SNPs
238 in 2 loci, one of which is also detected by MaxP. ReAD identifies 3,307 significant SNPs in 7
239 genetic loci, including 5 loci that are not detected by MaxP or STAREG. Figure 6 presents
240 the Manhattan plots of MaxP, STAREG and ReAD.

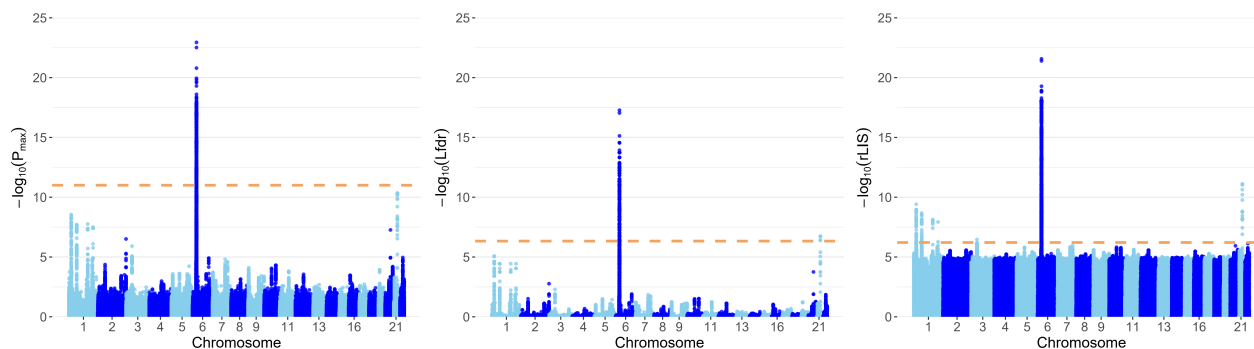


Figure 6: The Manhattan plots based on P_{\max} , Lfdr and rLIS. The dashed horizontal lines denote the FDR cutoffs of 5×10^{-8} produced by MaxP, STAREG, and ReAD, respectively.

241 We assess the replicability of genetic loci identified by different methods in GWAS Catalog
242 (Welter et al., 2014). Table 2 presents the main characteristics of the 7 replicable genetic loci
243 identified by ReAD. UC associations of these loci in cohorts of European descent have been
244 reported in the literature. For instance, the lead SNP of loci 6p21.32, rs6927022, is in the
245 intron of gene *HLA-DQA1*, and the HLA complex is associated with multiple risk alleles
246 for inflammatory bowel disease, including UC (Nowak et al., 2021; Reinshagen et al., 1996;
247 Ashton et al., 2019). The lead SNP harbored in loci 1q23.3, rs1801274, is only identified by
248 ReAD, and has confirmed associations with UC in several European-ancestry studies (Liu
249 et al., 2015; De Lange et al., 2017; Anderson et al., 2011). We have additional validations in
250 DisGeNET, a versatile platform that contains a comprehensive catalog of genes and variants
251 associated with human diseases (Piñero et al., 2015). Many mapped genes of the lead SNP
252 only identified by ReAD have been reported to be associated with UC, such as gene *FCGR2A*
253 in locus 1q23.3, gene *IL23R* in locus 1p31.3, gene *IL10* in locus 1q32.1, and gene *MST1* in
254 locus 3p21.31.

255 **3 Discussion**

256 In this paper, we present ReAD, an efficient method accounting for the LD structure to
257 identify replicable associations from two GWASs datasets. We conduct extensive simulation
258 studies and analyze two GWAS datasets. Compared to conventional approaches that impose
259 independence assumption among SNPs, ReAD provides effective FDR control. It has a sub-
260 stantial power gain in identifying genuine and replicable genetic loci. It is computationally
261 scalable to hundreds of millions of SNPs and has no tuning parameters.

262 In this paper, our discussion mainly focuses on assessing the replicability of each SNP
263 within a genomic locus. We acknowledge that, in the applications of genetic association anal-
264 ysis, varying LD patterns between studies can lead to inconsistent significant findings at the

Table 2: Main characteristics of the 7 loci associated with UC in the European-ancestry IIB-DGC and UK Biobank GWASs identified by ReAD. The SNP with the strongest association within each locus is called the lead SNP. The mapped gene denotes genes overlapping or closest to the lead SNP in the identified locus.

| Locus | Lead SNP | Location of lead SNP | Mapped gene | P_{\max} | Lfdr | rLIS |
|---|-----------|----------------------|------------------------|------------|---------|---------|
| Replicable asthma loci identified by all methods | | | | | | |
| 6p21.32 | rs6927022 | chr6:32,644,620 | <i>HLA-DQA1</i> | 1.1e-20 | 2.8e-15 | 1.2e-19 |
| Replicable asthma loci identified by ReAD and STAREG but not by MaxP | | | | | | |
| 21q22.2 | rs2836882 | chr21:39,094,644 | <i>RPL23AP12</i> | 4.5e-11 | 1.9e-07 | 8.1e-12 |
| Replicable asthma loci only identified by ReAD | | | | | | |
| 1p36.13 | rs4654903 | chr1:19,874,497 | <i>RNF186, OTUD3</i> | 1.3e-08 | 3.6e-05 | 6.7e-09 |
| 1q23.3 | rs1801274 | chr1:161,509,955 | <i>FCGR2A</i> | 1.7e-08 | 3.6e-05 | 7.7e-09 |
| 1p31.3 | rs2201841 | chr1:67,228,519 | <i>C1orf141, IL23R</i> | 6.1e-08 | 1.8e-04 | 1.8e-08 |
| 1q32.1 | rs3024505 | chr1:206,766,559 | <i>Y_RNA, IL10</i> | 4.1e-08 | 8.9e-05 | 5.4e-07 |
| 3p21.31 | rs3197999 | chr3:49,684,099 | <i>MST1</i> | 1.2e-06 | 5.2e-03 | 3.4e-07 |

265 SNP level. Hence scientifically, a more relevant question should be the consistency of under-
 266 lying association signals within each interrogated locus across original and replication studies.
 267 To this end, we apply a simple and practical strategy requiring at least one SNP-level findings
 268 replicable. With the potential varying LD structures fully accounted for by the proposed
 269 HMM, we find this strategy intuitive and effective when applied to genomic loci with proper
 270 resolutions (as illustrated by our simulations and real data examples). Nevertheless, this
 271 locus-level criterion may be considered overly lenient. We will continue to explore alternative
 272 locus-level replicability assessment criteria in our future work.

273 In this work, we use repeated significance to assess replicability. We note that applying
 274 such a replicability criterion is debatable in the scientific community. While acknowledging
 275 its drawbacks, especially its conservativeness, we note the following context-specific factors.
 276 First, despite continued efforts to include more informative statistics summarizing GWAS
 277 findings, a large body of historical GWAS findings are *only* reported in p -values (See GWAS

278 catalog (Welter et al., 2014)), which fundamentally limits applying alternative replicability
279 criteria. Second, because complicated unknown confoundings, e.g., population stratification
280 and unobserved batch effects in genotyping experiments, often cause false positives in genetic
281 association analysis, the genetics community has consistently advocated conservative replica-
282 bility criteria to ensure the reliability of GWAS findings (Skol et al., 2006; McGuire et al.,
283 2021). Third, we emphasize that our main statistical contribution is to account for the corre-
284 lation structure between genetic variants, and our work can be naturally extended to applying
285 other alternative replicability criteria.

286 On a related point, although we exclusively assume that GWAS results are reported in the
287 form of single-SNP testing p -values throughout this paper, the proposed statistical method-
288 ology can be extended to other forms of summary statistics. For example, probabilistic fine-
289 mapping analysis of genetic association signals has become increasingly popular, thanks to
290 the availability of efficient variable selection algorithms (Benner et al., 2016; Wang et al.,
291 2020; Wen et al., 2016). The fine-mapping result is typically given as a posterior inclusion
292 probability (PIP) at the individual SNP level. With the ability to construct a Bayesian cred-
293 ible set for each underlying signal within a genomic locus, the PIPs have many advantages
294 over single-SNP p -values. Theoretically, our work can be straightforwardly extended to this
295 setting by noting the connection that $1 - \text{PIP}$ is equivalent to the local fdr in the Bayesian
296 perspective. We will leave this extension to our future work.

297 4 Methods

298 4.1 The hidden Markov model for replicability analysis

Suppose there are J SNPs in two independent GWASs. We are interested in testing whether the j th SNP is associated with the phenotype in both studies. Let θ_{ij} denote the inferred

association status of SNP j in study i , where $\theta_{ij} = 1$ indicates the j th SNP ($j = 1, \dots, J$) is inferred associated with the phenotype in study i ($i = 1, 2$) and $\theta_{ij} = 0$ otherwise. We use s_j ($j = 1, \dots, J$) to denote the joint status.

$$s_j = \begin{cases} 0, & (\theta_{1j}, \theta_{2j}) = (0, 0), \\ 1, & (\theta_{1j}, \theta_{2j}) = (0, 1), \\ 2, & (\theta_{1j}, \theta_{2j}) = (1, 0), \\ 3, & (\theta_{1j}, \theta_{2j}) = (1, 1). \end{cases}$$

299 The replicability null hypotheses is

$$H_{0j} : s_j \in \{0, 1, 2\}, j = 1, \dots, J. \quad (1)$$

300 Let $\mathbf{p}_i = (p_{ij})_{j=1}^J$ denote p -values of J SNPs in study i . We use mixture models for the
301 conditional distributions of p -values given θ values. Specifically,

$$\begin{aligned} p_{1j} | \theta_{1j} &\sim (1 - \theta_{1j})f_0 + \theta_{1j}f_1, \\ p_{2j} | \theta_{2j} &\sim (1 - \theta_{2j})f_0 + \theta_{2j}f_2, \end{aligned} \quad (2)$$

302 where f_0 is the probability density function of p -values when $\theta_{1j} = \theta_{2j} = 0$, and f_1 and f_2 are
303 the p -value density functions under non-null in study 1 and study 2, respectively. We assume
304 f_0 follows the standard uniform distribution and impose the following monotone likelihood
305 ratio condition (Sun and Cai, 2007; Cao et al., 2013, 2022).

$$f_1(x)/f_0(x) \text{ and } f_2(x)/f_0(x) \text{ are monotonically non-increasing in } x. \quad (3)$$

306 This condition naturally arises as small p -values indicate evidence against the null. To capture

307 the LD structure among SNPs, we assume that $\mathbf{s} = (s_1, \dots, s_J)$ follows a four-state stationary,
 308 irreducible, and aperiodic hidden Markov model (HMM). The transition probabilities

$$a_{kl} = \mathbb{P}(s_{j+1} = l | s_j = k) \quad (4)$$

for $k, l = 0, 1, 2, 3$ with constraint $\sum_{l=0}^3 a_{kl} = 1$. The stationary distribution of each state s_j is $\mathbb{P}(s_j = k) = \pi_k$ for $k = 0, 1, 2, 3$ and $\sum_{k=0}^3 \pi_k = 1$. The paired p -values for the j th SNP are assumed to be conditionally independent satisfying

$$f(p_{1j}, p_{2j} | \theta_{1j}, \theta_{2j}) = f(p_{1j} | \theta_{1j}) f(p_{2j} | \theta_{2j}).$$

Based on the mixture model (2), we have

$$f^{(s_j)}(p_{1j}, p_{2j}) = \begin{cases} f_0(p_{1j}) f_0(p_{2j}), & s_j = 0, \\ f_0(p_{1j}) f_2(p_{2j}), & s_j = 1, \\ f_1(p_{1j}) f_0(p_{2j}), & s_j = 2, \\ f_1(p_{1j}) f_2(p_{2j}), & s_j = 3. \end{cases}$$

Denote by $\mathcal{A} = \{a_{kl} : k, l = 0, 1, 2, 3\}$ the transition matrix, $\boldsymbol{\pi} = (\pi_0, \pi_1, \pi_2, \pi_3)$ the vector of stationary distribution, and $\mathcal{F} = (f^{(0)}, f^{(1)}, f^{(2)}, f^{(3)})$ the probability density functions of the bivariate observations (p_{1j}, p_{2j}) . The convergence theorem of a Markov chain (Theorem 5.5.1 in Durrett (2019)) implies that

$$\frac{1}{J} \sum_{j=1}^J I(s_j = k) \rightarrow \pi_k$$

309 almost surely for $k = 0, 1, 2, 3$ as $J \rightarrow \infty$. As f_0 is assumed to follow a standard uniform
 310 distribution, we use $\boldsymbol{\lambda} = (\boldsymbol{\pi}, \mathcal{A}, f_1, f_2)$ to denote the collection of unknown parameters and
 311 functions in the HMM. Our goal is to separate the replicable SNPs ($s_j = 3$) from the non-

312 replicable SNPs ($s_j \in \{0, 1, 2\}$) based on the observed bivariate p -values.

313 4.2 FDR control for replicability analysis accounting for LD

314 4.2.1 The rLIS statistic for replicability analysis across two studies

Consider the ideal setup that an oracle knows $\boldsymbol{\lambda} = (\boldsymbol{\pi}, \mathcal{A}, f_1, f_2)$. We define the replicability local index of significance (rLIS) as the posterior probability of being null. Specifically,

$$\text{rLIS}_j := \mathbb{P}_{\boldsymbol{\lambda}}(s_j \in \{0, 1, 2\} | \mathbf{p}_1, \mathbf{p}_2).$$

Given $\boldsymbol{\lambda}$, the forward and backward probabilities are defined as $\alpha_j(s_j) = \mathbb{P}_{\boldsymbol{\lambda}}((p_{1t}, p_{2t})_{t=1}^j, s_j)$ and $\beta_j(s_j) = \mathbb{P}_{\boldsymbol{\lambda}}((p_{1t}, p_{2t})_{t=j+1}^J | s_j)$, respectively. The forward-backward procedure (Baum et al., 1970) can be used in the calculation. Specifically, we initialize $\alpha_1(s_1) = \pi_{s_1} f^{(s_1)}(p_{11}, p_{21})$ and $\beta_J(s_J) = 1$. We can obtain $\alpha_j(\cdot)$ and $\beta_j(\cdot)$ for $j = 1, \dots, J$ recursively by

$$\alpha_{j+1}(s_{j+1}) = \sum_{s_j=0}^3 \alpha_j(s_j) a_{s_j s_{j+1}} f^{(s_{j+1})}(p_{1,j+1}, p_{2,j+1})$$

and

$$\beta_j(s_j) = \sum_{s_{j+1}=0}^3 \beta_{j+1}(s_{j+1}) f^{(s_{j+1})}(p_{1,j+1}, p_{2,j+1}) a_{s_j s_{j+1}}.$$

Hence we have

$$\text{rLIS}_j = \frac{\sum_{s_j=0}^2 \alpha_j(s_j) \beta_j(s_j)}{\sum_{s_j=0}^3 \alpha_j(s_j) \beta_j(s_j)}.$$

The rejection rule can be written as

$$\delta_j = I(\text{rLIS}_j \leq t), \quad j = 1, \dots, J,$$

315 where $I(\cdot)$ is the indicator function.

We next derive the threshold \hat{t} for a pre-specified FDR level q . Total number of discoveries and the number of false discoveries are $R(t) = \sum_{j=1}^J I(\text{rLIS}_j \leq t)$ and $V(t) = \sum_{j=1}^J I(\text{rLIS}_j \leq t, s_j \in \{0, 1, 2\})$, respectively. We have

$$\begin{aligned} \mathbb{E}[V(t)] &= \mathbb{E} \left[\sum_{j=1}^J \{ \pi_0 I(\text{rLIS}_j \leq t | s_j = 0) + \pi_1 I(\text{rLIS}_j \leq t | s_j = 1) + \pi_2 I(\text{rLIS}_j \leq t | s_j = 2) \} \right] \\ &= \mathbb{E} \left[\sum_{j=1}^J I(\text{rLIS}_j \leq t) \text{rLIS}_j \right]. \end{aligned}$$

Let $\text{rLIS}_{(1)} \leq \text{rLIS}_{(2)} \leq \dots \leq \text{rLIS}_{(J)}$ be the order statistics and $H_{(1)}, \dots, H_{(J)}$ be the corresponding hypotheses. If k hypotheses are rejected, the number of false discoveries can be estimated by

$$\hat{V}(k) = \sum_{j=1}^k \text{rLIS}_{(j)},$$

and the FDR can be estimated by $\frac{1}{k} \sum_{j=1}^k \text{rLIS}_{(j)}$. We shall use the following step-up procedure to control the FDR at level q (Sun and Cai, 2009).

$$\text{let } \hat{k} = \max \left\{ k : \frac{1}{k} \sum_{j=1}^k \text{rLIS}_{(j)} \leq q \right\};$$

then reject all $H_{(j)}$ for $j = 1, \dots, \hat{k}$.

316 We provide an estimation of $\boldsymbol{\lambda}$ in the next section.

317 4.2.2 Data-driven testing procedure

To estimate the unknown parameters and functions in $\boldsymbol{\lambda}$, we first define two posterior probabilities $\gamma_j(s_j) = \mathbb{P}_{\boldsymbol{\lambda}}(s_j | \mathbf{p}_1, \mathbf{p}_2)$ and $\xi_j(s_j, s_{j+1}) = \mathbb{P}_{\boldsymbol{\lambda}}(s_j, s_{j+1} | \mathbf{p}_1, \mathbf{p}_2)$. By the definition, $\gamma_j(s_j) = \sum_{s_{j+1}=0}^3 \xi_j(s_j, s_{j+1})$. They can be obtained from the forward and backward probabil-

ities

$$\gamma_j(s_j) = \frac{\alpha_j(s_j)\beta_j(s_j)}{\sum_{s_j=0}^3 \alpha_j(s_j)\beta_j(s_j)}$$

and

$$\xi_j(s_j, s_{j+1}) = \frac{\alpha_j(s_j)\beta_{j+1}(s_{j+1})a_{s_j s_{j+1}} f^{(s_{j+1})}(p_{1,j+1}, p_{2,j+1})}{\sum_{s_j=0}^3 \sum_{s_{j+1}=0}^3 \alpha_j(s_j)\beta_{j+1}(s_{j+1})a_{s_j s_{j+1}} f^{(s_{j+1})}(p_{1,j+1}, p_{2,j+1})}.$$

The likelihood function of the complete data $(\mathbf{p}_1, \mathbf{p}_2, \mathbf{s})$ is given by

$$L(\boldsymbol{\lambda}; \mathbf{p}_1, \mathbf{p}_2, \mathbf{s}) = \pi_{s_1} \prod_{j=2}^J a_{s_{j-1} s_j} \cdot \prod_{j=1}^J f^{(s_j)}(p_{1j}, p_{2j}).$$

318 We develop a non-parametric EM algorithm (Dempster et al., 1977) to estimate the unknowns
 319 $\boldsymbol{\lambda} = (\boldsymbol{\pi}, \mathcal{A}, f_1, f_2)$ under the monotone likelihood ratio constraint (3). With an appropriate
 320 initialization of the unknowns, $\boldsymbol{\lambda}^{(0)} = (\boldsymbol{\pi}^{(0)}, \mathcal{A}^{(0)}, f_1^{(0)}, f_2^{(0)})$, the EM algorithm proceeds by
 321 iteratively implementing the following two steps.

E-step: Given current $\boldsymbol{\lambda}^{(t)} = (\boldsymbol{\pi}^{(t)}, \mathcal{A}^{(t)}, f_1^{(t)}, f_2^{(t)})$, the forward and backward probabilities $\alpha_j^{(t)}(s_j), \beta_j^{(t)}(s_j)$ and the posterior probabilities $\gamma_j^{(t)}(s_j), \xi_j^{(t)}(s_j, s_{j+1})$ are calculated. The conditional expectation of the log-likelihood function can be written as

$$\begin{aligned} D(\boldsymbol{\lambda}|\boldsymbol{\lambda}^{(t)}) &= \sum_{\mathbf{s}} \mathbb{P}_{\boldsymbol{\lambda}^{(t)}}(\mathbf{s}|\mathbf{p}_1, \mathbf{p}_2) \log L(\boldsymbol{\lambda}; \mathbf{p}_1, \mathbf{p}_2, \mathbf{s}) \\ &= \sum_{\mathbf{s}} \left\{ \mathbb{P}_{\boldsymbol{\lambda}^{(t)}}(\mathbf{s}|\mathbf{p}_1, \mathbf{p}_2) \left[\log \pi_{s_1} + \sum_{j=2}^J \log a_{s_{j-1} s_j} + \sum_{j=1}^J \log f^{(s_j)}(p_{1j}, p_{2j}) \right] \right\}. \end{aligned}$$

M-step: Update $\boldsymbol{\lambda}^{(t+1)}$ by

$$\boldsymbol{\lambda}^{(t+1)} = \arg \max_{\boldsymbol{\pi}, \mathcal{A}, f_1, f_2} D(\boldsymbol{\pi}, \mathcal{A}, f_1, f_2 | \boldsymbol{\lambda}^{(t)}).$$

We can update each component alternately. By using the Lagrange multiplier, we can calculate $\boldsymbol{\pi}^{(t+1)}$ and $\mathcal{A}^{(t+1)}$ as

$$\pi_s^{(t+1)} = \gamma_1^{(t)}(s), s \in \{0, 1, 2, 3\}$$

and

$$a_{kl}^{(t+1)} = \frac{\sum_{j=2}^J \xi_{j-1}^{(t)}(k, l)}{\sum_{j=2}^J \sum_{l=0}^3 \xi_{j-1}^{(t)}(k, l)}, k, l \in \{0, 1\}.$$

The two functions can be updated by

$$f_1^{(t+1)} = \arg \max_{f_1 \in \mathbb{H}} \left\{ \sum_{j=1}^J \left[\gamma_j^{(t)}(2) + \gamma_j^{(t)}(3) \right] \log f_1(p_{1j}) \right\} \quad (5)$$

and

$$f_2^{(t+1)} = \arg \max_{f_2 \in \mathbb{H}} \left\{ \sum_{j=1}^J \left[\gamma_j^{(t)}(1) + \gamma_j^{(t)}(3) \right] \log f_2(p_{2j}) \right\}, \quad (6)$$

322 where \mathbb{H} is a set of monotonic non-increasing density functions (Sun and Cai, 2007; Cao
 323 et al., 2013, 2022). We solve (5) and (6) independently using the non-parametric maximum
 324 likelihood estimation implemented with PAVA (Robertson et al., 1988).

The **E-step** and **M-step** are conducted iteratively until convergence. Detailed derivations of the algorithm are presented in the Supplemental Note A. With the estimate $\hat{\boldsymbol{\lambda}} = \{\hat{\boldsymbol{\pi}}, \hat{\mathcal{A}}, \hat{f}_1, \hat{f}_2\}$, we can calculate the test statistics $\widehat{\text{rLIS}}_j = \mathbb{P}_{\hat{\boldsymbol{\lambda}}}(s_j \in \{0, 1, 2\} | \mathbf{p}_1, \mathbf{p}_2)$. Let $\widehat{\text{rLIS}}_{(1)} \leq \dots \leq \widehat{\text{rLIS}}_{(J)}$ be the order statistics of $\widehat{\text{rLIS}}_j$, and denote $H_{(1)}, \dots, H_{(J)}$ as the

corresponding H_{0j} . The data-driven testing procedure works as follows.

$$\text{Let } \hat{k} = \max \left\{ i : \frac{1}{i} \sum_{j=1}^i \widehat{\text{rLIS}}_{(j)} \leq q \right\},$$

and reject $H_{(i)}$ for $i = 1, \dots, \hat{k}$.

325 4.3 Realistic simulation design

326 In simulation II, we perform realistic simulations to show the rankings of SNPs using rLIS in
327 two GWASs, where LD structures are derived from real data. Based on the CEU genotype data
328 and FIN genotype data from the Genetic European Variation in Disease project (Lappalainen
329 et al., 2013), we filter out SNPs with the same genotypes in all samples and obtain genotypes
330 of 16,764 SNPs in both studies. We specify 5 causal SNPs in each study, 4 of which are the
331 same in the two studies. Two of the 4 causal SNPs are close (separated by five SNPs), and
332 the other SNPs are selected randomly. Then in each study, for the i th subject ($i = 1, \dots, 78$
333 in the CEU study and $i = 1, \dots, 89$ in the FIN study), we generate continuous phenotypes
334 using the linear regression model

$$y_i = \beta_0 + \sum_{k=1}^5 G_{ik}^c \beta_k + \epsilon_i,$$

335 where β_0 is the intercept term, $G_{i1}^c, \dots, G_{i5}^c$ are the genotypes of the i th subject for the 5
336 causal SNPs, β_1, \dots, β_5 are regression coefficients, and ϵ_i is an error term generated from
337 $N(0, 1)$, a standard normal distribution. The intercept term and the regression coefficients
338 of causal SNPs $\beta_k, k = 0, 1, \dots, 5$, follow $N(0, \sigma^2)$, a normal distribution with mean 0 and
339 standard deviation σ . We set $\sigma = 0.6$ such that the SNP heritability h_g^2 , similar in spirit to
340 the R^2 in linear regression models, is centered between 0.2 and 0.3. p -values are obtained by
341 a marginal regression of each SNP on the phenotype. We repeat the above process 100 times

342 to get p -values of 1, 676, 400 SNPs.

343 4.4 Computation time

344 We compare the computation time of different methods. All methods are implemented in R,
345 in which STAREG and ReAD use Rcpp to speed up the computation. All computations are
346 carried out in an Intel(R) Core(TM) i7-9750H 2.6GHz CPU with 64 GB RAM laptop. Table
347 3 summarizes the results. We observe that all methods are quick to compute. The additional
348 time that ReAD takes in simulation studies is negligible in practice.

Table 3: Computation time (in seconds) of different methods.

| Dataset | # of SNPs | MaxP | STAREG | ReAD |
|---------------|-------------|--------|--------|--------|
| Simulation I | 10, 000 | 0.0033 | 0.0165 | 0.1871 |
| Simulation II | 1, 676, 400 | 0.1897 | 3.7237 | 35.159 |
| Asthma | 6, 222, 195 | 0.7317 | 169.82 | 105.47 |
| UC | 6, 232, 147 | 0.7428 | 9.5248 | 83.974 |

349 References

350 S. Alatab, S. G. Sepanlou, and et al. The global, regional, and national burden of inflammatory
351 bowel disease in 195 countries and territories, 1990–2017: a systematic analysis for the
352 global burden of disease study 2017. *The Lancet Gastroenterology and Hepatology*, 5(1):
353 17–30, 2020.

354 C. A. Anderson, G. Boucher, C. W. Lees, A. Franke, M. D’Amato, K. D. Taylor, J. C. Lee,
355 P. Goyette, M. Imielinski, A. Latiano, et al. Meta-analysis identifies 29 additional ulcerative
356 colitis risk loci, increasing the number of confirmed associations to 47. *Nature Genetics*, 43
357 (3):246–252, 2011.

- 358 J. J. Ashton, K. Latham, R. M. Beattie, and S. Ennis. the genetics of the human leucocyte
359 antigen region in inflammatory bowel disease. *Alimentary Pharmacology & Therapeutics*,
360 50(8):885–900, 2019.
- 361 L. E. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occurring in
362 the statistical analysis of probabilistic functions of markov chains. *Annals of Mathematical*
363 *Statistics*, 41(1):164–171, 1970.
- 364 C. G. Begley and L. M. Ellis. Raise standards for preclinical cancer research. *Nature*, 483
365 (7391):531–533, 2012.
- 366 Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful
367 approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Method-*
368 *ological)*, 57(1):289–300, 1995.
- 369 Y. Benjamini, R. Heller, and D. Yekutieli. Selective inference in complex research. *Philosoph-*
370 *ical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*,
371 367(1906):4255–4271, 2009.
- 372 C. Benner, C. C. Spencer, A. S. Havulinna, V. Salomaa, S. Ripatti, and M. Pirinen. Finemap:
373 efficient variable selection using summary data from genome-wide association studies. *Bioin-*
374 *formatics*, 32(10):1493–1501, 2016.
- 375 M. Bogomolov and R. Heller. Replicability across multiple studies. *arXiv preprint*
376 *arXiv:2210.00522*, 2022.
- 377 F. M. Busing. Monotone regression: A simple and fast $o(n)$ pava implementation. *Journal of*
378 *Statistical Software*, 102:1–25, 2022.
- 379 H. Cao, W. Sun, and M. R. Kosorok. The optimal power puzzle: scrutiny of the monotone
380 likelihood ratio assumption in multiple testing. *Biometrika*, 100(2):495–502, 2013.

- 381 H. Cao, J. Chen, and X. Zhang. Optimal false discovery rate control for large scale multiple
382 testing with auxiliary information. *Annals of Statistics*, 50(2):807–857, 2022.
- 383 S. J. Chanock, T. Manolio, L. D. Brooks, L. R. Cardon, M. Daly, and P. Donnelly. Replicating
384 genotype-phenotype associations. *Nature (London)*, 447(7145):655–660, 2007.
- 385 D. Chung, C. Yang, C. Li, J. Gelernter, and H. Zhao. Gpa: a statistical approach to prioritizing
386 gwas results by integrating pleiotropy and annotation. *PLoS Genetics*, 10(11):e1004787,
387 2014.
- 388 G. A. Churchill. Hidden markov chains and the analysis of genome structure. *Computers &*
389 *Chemistry*, 16(2):107–115, 1992.
- 390 K. M. De Lange, L. Moutsianas, J. C. Lee, C. A. Lamb, Y. Luo, N. A. Kennedy, L. Jostins,
391 D. L. Rice, J. Gutierrez-Achury, S.-G. Ji, et al. Genome-wide association study impli-
392 cates immune activation of multiple integrin genes in inflammatory bowel disease. *Nature*
393 *Genetics*, 49(2):256–261, 2017.
- 394 F. Demenais, P. Margaritte-Jeannin, K. C. Barnes, W. O. Cookson, J. Altmüller, W. Ang,
395 R. G. Barr, T. H. Beaty, A. B. Becker, J. Beilby, et al. Multiancestry association study
396 identifies new asthma risk loci that colocalize with immune-cell enhancer marks. *Nature*
397 *Genetics*, 50(1):42–53, 2018.
- 398 A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data
399 via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*,
400 39(1):1–22, 1977.
- 401 R. Durrett. *Probability: theory and examples*. Cambridge university press, 2019.
- 402 B. Efron. *Large-scale inference: empirical Bayes methods for estimation, testing, and predic-*
403 *tion*. Cambridge University Press, 2012.

404 M. A. Ferreira, M. C. Matheson, C. S. Tang, R. Granell, W. Ang, J. Hui, A. K. Kiefer,
405 D. L. Duffy, S. Baltic, P. Danoy, et al. Genome-wide association analysis identifies 11 risk
406 variants associated with the asthma with hay fever phenotype. *Journal of Allergy and*
407 *Clinical Immunology*, 133(6):1564–1571, 2014.

408 M. A. Ferreira, J. M. Vonk, H. Baurecht, I. Marenholz, C. Tian, J. D. Hoffman, Q. Helmer,
409 A. Tillander, V. Ullemar, J. Van Dongen, et al. Shared genetic origin of asthma, hay fever
410 and eczema elucidates allergic disease biology. *Nature Genetics*, 49(12):1752–1757, 2017.

411 L. P. Freedman, I. M. Cockburn, and T. S. Simcoe. The economics of reproducibility in
412 preclinical research. *PLoS Biology*, 13(6):e1002165, 2015.

413 R. Heller and D. Yekutieli. Replicability analysis analysis for genome-wide association studies.
414 *Annals of Applied Statistics*, 8(1):481–498, 2014.

415 R. Heller, S. Yaacoby, and D. Yekutieli. repfdr: a tool for replicability analysis for genome-wide
416 association studies. *Bioinformatics*, 30(20):2971–2972, 2014.

417 J. Huffman. Examining the current standards for genetic discovery and replication in the era
418 of mega-biobanks. *Nature Communications*, 9(1):5054, 2018.

419 J. Ioannidis, E. E. Ntzani, T. A. Trikalinos, and D. G. Contopoulos-Ioannidis. Replication
420 validity of genetic association studies. *Nature Genetics*, 29(3):306–309, 2001.

421 J. P. Ioannidis. Why most published research findings are false. *PLoS Medicine*, 2(8):e124,
422 2005.

423 T. Lappalainen, M. Sammeth, M. R. Friedländer, P. A. 't Hoen, J. Monlong, M. A. Rivas,
424 M. Gonzalez-Porta, N. Kurbatova, T. Griebel, P. G. Ferreira, et al. Transcriptome and
425 genome sequencing uncovers functional variation in humans. *Nature*, 501(7468):506–511,
426 2013.

- 427 N. Li and M. Stephens. Modeling linkage disequilibrium and identifying recombination
428 hotspots using single-nucleotide polymorphism data. *Genetics*, 165(4):2213–2233, 2003.
- 429 Q. Li, J. B. Brown, H. Huang, and P. J. Bickel. Measuring reproducibility of high-throughput
430 experiments. *Annals of Applied Statistics*, 5(3):1752–1779, 2011.
- 431 Y. Li, X. Zhou, R. Chen, X. Zhang, and H. Cao. Stareg: an empirical bayesian approach to
432 detect replicable spatially variable genes in spatial transcriptomic studies. *bioRxiv*, 2023.
- 433 J. Z. Liu, S. Van Sommeren, H. Huang, S. C. Ng, R. Alberts, A. Takahashi, S. Ripke, J. C.
434 Lee, L. Jostins, T. Shah, et al. Association analyses identify 38 susceptibility loci for
435 inflammatory bowel disease and highlight shared genetic risk across populations. *Nature*
436 *Genetics*, 47(9):979–986, 2015.
- 437 C. Lonjou, W. Zhang, A. Collins, W. J. Tapper, E. Elahi, N. Maniatis, and N. E. Morton.
438 Linkage disequilibrium in human populations. *Proceedings of the National Academy of*
439 *Sciences*, 100(10):6069–6074, 2003.
- 440 J. MacArthur, E. Bowler, M. Cerezo, L. Gil, P. Hall, E. Hastings, H. Junkins, A. McMahon,
441 A. Milano, J. Morales, et al. The new nhgri-ebi catalog of published genome-wide association
442 studies (gwas catalog). *Nucleic Acids Research*, 45(D1):D896–D901, 2017.
- 443 M. I. McCarthy, G. R. Abecasis, L. R. Cardon, D. B. Goldstein, J. Little, J. P. Ioannidis,
444 and J. N. Hirschhorn. Genome-wide association studies for complex traits: consensus,
445 uncertainty and challenges. *Nature Reviews Genetics*, 9(5):356–369, 2008.
- 446 D. McGuire, Y. Jiang, M. Liu, J. D. Weissenkampen, S. Eckert, L. Yang, F. Chen, A. Berg,
447 S. Vrieze, et al. Model-based assessment of replicability for genome-wide association meta-
448 analysis. *Nature Communications*, 12(1):1964, 2021.

449 N. A. Molodecky, S. Soon, D. M. Rabi, W. A. Ghali, M. Ferris, G. Chernoff, E. I. Benchimol,
450 R. Panaccione, S. Ghosh, H. W. Barkema, et al. Increasing incidence and prevalence of the
451 inflammatory bowel diseases with time, based on systematic review. *Gastroenterology*, 142
452 (1):46–54, 2012.

453 R. Moonesinghe, M. J. Khoury, T. Liu, and J. P. Ioannidis. Required sample size and non-
454 replicability thresholds for heterogeneous genetic associations. *Proceedings of the National
455 Academy of Sciences*, 105(2):617–622, 2008.

456 K. P. Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

457 J. K. Nowak, A. Glapa-Nowak, A. Banaszkiwicz, B. Iwańczak, J. Kwiecień, A. Szaflarska-
458 Popławska, U. Grzybowska-Chlebowczyk, M. Osiecki, J. Kierkuś, M. Hołubiec, et al. Hla-
459 dqɑ1* 05 associates with extensive ulcerative colitis at diagnosis: An observational study
460 in children. *Genes*, 12(12):1934, 2021.

461 T. A. Olafsdottir, F. Theodors, K. Bjarnadottir, U. S. Bjornsdottir, A. B. Agustsdottir, O. A.
462 Stefansson, E. V. Ivarsdottir, J. K. Sigurdsson, S. Benonisdottir, G. I. Eyjolfsson, et al.
463 Eighty-eight variants highlight the role of t cell regulation and airway remodeling in asthma
464 pathogenesis. *Nature Communications*, 11(1):393, 2020.

465 D. Philtrou, Y. Lyu, Q. Li, and D. Ghosh. Maximum rank reproducibility: a nonparametric
466 approach to assessing reproducibility in replicate experiments. *Journal of the American
467 Statistical Association*, 113(523):1028–1039, 2018.

468 J. Piñero, À. Bravo, N. Queralt-Rosinach, A. Gutiérrez-Sacristán, J. Deu-Pons, E. Centeno,
469 J. García-García, F. Sanz, and L. I. Furlong. Disgenet: a comprehensive platform integrating
470 information on human disease-associated genes and variants. *Nucleic Acids Research*, 2015:
471 gkw943, 2015.

472 F. Prinz, T. Schlange, and K. Asadullah. Believe it or not: how much can we rely on published
473 data on potential drug targets? *Nature Reviews Drug Discovery*, 10(9):712–712, 2011.

474 J. K. Pritchard and M. Przeworski. Linkage disequilibrium in humans: models and data.
475 *American Journal of Human Genetics*, 69(1):1–14, 2001.

476 L. Rabiner and B. Juang. An introduction to hidden markov models. *IEEE ASSP Magazine*,
477 3(1):4–16, 1986.

478 M. Reinshagen, C. Loeliger, P. Kuehnl, U. Weiss, B. Manfras, G. Adler, and B. Boehm. Hla
479 class ii gene frequencies in crohn’s disease: a population based analysis in germany. *Gut*,
480 38(4):538–542, 1996.

481 T. Robertson, R. L. Dykstra, and F. T. Wright. Order restricted statistical inference. In
482 *Wiley Series in Probability and Mathematical Statistics*. John Wiley and Sons, 1988.

483 A. D. Skol, L. J. Scott, G. R. Abecasis, and M. Boehnke. Joint analysis is more efficient than
484 replication-based analysis for two-stage genome-wide association studies. *Nature Genetics*,
485 38(2):209–213, 2006.

486 C. Sudlow, J. Gallacher, N. Allen, V. Beral, P. Burton, J. Danesh, P. Downey, P. Elliott,
487 J. Green, M. Landray, et al. Uk biobank: an open access resource for identifying the causes
488 of a wide range of complex diseases of middle and old age. *PLoS Medicine*, 12(3):e1001779,
489 2015.

490 W. Sun and T. Cai. Large-scale multiple testing under dependence. *Journal of the Royal*
491 *Statistical Society: Series B (Statistical Methodology)*, 71(2):393–424, 2009.

492 W. Sun and T. T. Cai. Oracle and adaptive compound decision rules for false discovery rate
493 control. *Journal of the American Statistical Association*, 102(479):901–912, 2007.

494 J. D. Wall and J. K. Pritchard. Haplotype blocks and linkage disequilibrium in the human
495 genome. *Nature Reviews Genetics*, 4(8):587–597, 2003.

496 G. Wang, A. Sarkar, P. Carbonetto, and M. Stephens. A simple new approach to variable
497 selection in regression, with application to genetic fine mapping. *Journal of the Royal
498 Statistical Society Series B: Statistical Methodology*, 82(5):1273–1300, 2020.

499 Z. Wei, W. Sun, K. Wang, and H. Hakonarson. Multiple testing in genome-wide association
500 studies via hidden markov models. *Bioinformatics*, 25(21):2802–2808, 2009.

501 D. Welter, J. MacArthur, J. Morales, T. Burdett, P. Hall, H. Junkins, A. Klemm, P. Flicek,
502 T. Manolio, L. Hindorff, et al. The nhgri gwas catalog, a curated resource of snp-trait
503 associations. *Nucleic Acids Research*, 42(D1):D1001–D1006, 2014.

504 X. Wen, Y. Lee, F. Luca, and R. Pique-Regi. Efficient integrative multi-snp association
505 analysis via deterministic approximation of posteriors. *The American Journal of Human
506 Genetics*, 98(6):1114–1129, 2016.

507 Y. Zhao, M. G. Sampson, and X. Wen. Quantify and control reproducibility in high-
508 throughput experiments. *Nature Methods*, 17(12):1207–1213, 2020.

509 Z. Zhu, P. H. Lee, M. D. Chaffin, W. Chung, P.-R. Loh, Q. Lu, D. C. Christiani, and L. Liang.
510 A genome-wide cross-trait analysis from uk biobank highlights the shared genetic architec-
511 ture of asthma and allergic diseases. *Nature Genetics*, 50(6):857–864, 2018.

512 Z. Zhu, Y. Guo, H. Shi, C.-L. Liu, R. A. Panganiban, W. Chung, L. J. O’Connor, B. E. Himes,
513 S. Gazal, K. Hasegawa, et al. Shared genetic and experimental links between obesity-related
514 traits and asthma subtypes in uk biobank. *Journal of Allergy and Clinical Immunology*,
515 145(2):537–549, 2020.