A powerful approach to identify replicable variants in genome-wide association studies

Authors

Yan Li, Haochen Lei, Xiaoquan Wen, Hongyuan Cao

Correspondence

hongyuancao@gmail.com

Replicability is one of the central issues in scientific research. We develop a robust and powerful method to assess replicability in GWAS via the repeated significance of SNPs. By accounting for the linkage disequilibrium structure, our method can identify replicable genetic loci that existing methods might otherwise miss.

Li et al., 2024, The American Journal of Human Genetics *111*, 966–978
May 2, 2024
https://doi.org/10.1016/j.ajhg.2024.04.004

CellPress

A powerful approach to identify replicable variants in genome-wide association studies

Yan Li,[1,2] Haochen Lei,[3] Xiaoquan Wen,[4] and Hongyuan Cao[3,*]

Summary

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 control. Through analysis of datasets from two asthma GWASs and two ulcerative colitis GWASs, we show that ReAD can identify replicable genetic loci that existing methods might otherwise miss.

Introduction

Genome-wide association studies (GWASs) allow for simultaneous study of millions of single-nucleotide polymorphisms (SNPs). Numerous genetic risk variants associated with various phenotypes and complex diseases have been reported over the past couple of decades.[1,2] These associations provide insights into the architecture of disease susceptibility. Despite these progresses, many reported genotype-phenotype associations fail to replicate in other studies.[3,4] An analysis of past studies indicates that the cumulative prevalence of irreplicable preclinical research (including GWASs) exceeds 50%[5–8]. Approximately 28 billion dollars annually is spent on preclinical research that is not replicable in the United States alone.[8] Irreplicable and/or inconsistent between-study associations might be spurious findings caused by confounding factors, such as population stratification, misclassification of phenotypes, genotyping errors, or technical biases, among others. Replicability is now considered a *sine qua non* for establishing credible genotype-phenotype associations in the era of GWASs.[9,10] We study conceptual replicability where consistent results are obtained using different processes and populations that target the same scientific question. For GWASs, replicability analysis aims to detect genetic risk loci that are significantly associated with the same phenotype across different studies.[11–13] By eliminating genetic associations that cannot be generalized across studies, replicability analysis provides stronger support for genuine scientific findings, avoids wasted resources, and improves efficiency of drug development. This helps the translation of bench discoveries to bedsided therapies.

In GWASs, millions of SNPs are tested simultaneously, requiring multiple testing adjustment. False discovery rate (FDR), defined as the expectation of the proportion of false discoveries over total discoveries, is a commonly used metric for type I error control.[14] A central characteristic of GWAS data is the linkage disequilibrium (LD) among SNPs with which alleles at nearby sites can co-occur on the same haplotype more often than by chance alone.[15,16] As a result, it is common to observe that phenotype-associated SNPs form clusters and exhibit high correlations within clusters.[17] An effective approach to account for the LD structure among SNPs is through the hidden Markov model (HMM).[18,19] Existing GWAS literature[17,19,20] using HMM for a single study is not applicable to replicability analysis of multiple studies. Furthermore, their approaches cannot be generalized to more than one study due to the heterogeneity of LD across different studies.[21,22] Replicability analysis of GWASs explicitly accounting for the LD structure remains understudied and critically important.

To claim replicability, an *ad hoc* approach is to implement an FDR control method, such as the Benjamini and Hochberg (BH) procedure,[14] for each study and intersect significant results from all studies as replicable findings. This approach does not control the FDR and moreover has low power as it does not borrow information from different studies. The maximum of p values across studies ($P_{\max}$) is a straightforward significance measure for replicability.[23] After summarizing data from multiple studies by $P_{\max}$, classic FDR control procedures such as BH are used for replicability analysis. This procedure is overly conservative as it guards against the worst scenario and does not

[1]School of Computer Science and Technology, Changchun University of Science and Technology, Changchun, Jilin 130022, China; [2]School of Mathematics, Jilin University, Changchun, Jilin 130012, China; [3]Department of Statistics, Florida State University, Tallahassee, FL 32306, USA; [4]Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA
*Correspondence: hongyuancao@gmail.com
https://doi.org/10.1016/j.ajhg.2024.04.004.

incorporate the composite null structure of replicability analysis. For independent features from high-throughput experiments, various methods were proposed for replicability analysis. These methods are not robust to heterogeneity of different studies,[24,25] require tuning parameters,[26] impose parametric assumptions on the p values,[12] or demand access to full datasets, which can be prohibitive due to privacy concerns or logistics.[27] Wang and Zhu[28] investigated replicability analysis for dependent data but imposed strong parametric assumptions.

We address the limitations of existing methods by developing an efficient method, ReAD (Replicability Analysis accounting for Dependence) to detect replicable genotype-phenotype associations across two GWASs by incorporating the LD structure. We use GWAS summary statistics such as p values, treating multiple studies symmetrically. Our approach models the clustered signals from two studies with a four-dimensional HMM, accounting for the heterogeneity of LD structures in different studies. Conditional on the HMM, we model the two p value sequences as a four-group mixture of SNPs.[29,30] The replicability null hypothesis consists of three components: zero effects in both studies, zero effect in one study, and non-zero effect in another study and vice versa. ReAD calculates the posterior probability of replicability null given data. Compared to other replicability analysis methods, ReAD is robust as it is non-parametric, jointly models the signal and non-signal from different studies, and accounts for the heterogeneity of different studies. ReAD provides more efficient rankings of importance for replicable SNPs by pooling information from two p value sequences via the forward and backward probabilities.[31,32] ReAD applies a step-up procedure to identify clusters of genotype-phenotype associated signals, improving the power of replicability analysis while effectively controlling the FDR. ReAD is computationally scalable to whole-genome studies with tens of millions of SNPs. Its implementation combines the non-parametric expectation maximization (EM) algorithm[33] and the pool-adjacent-violator algorithm (PAVA) in shape constraint inference,[34,35] without any tuning parameters. We conduct extensive simulation studies to evaluate the performance of our approach across a wide range of scenarios. By applying our procedure to summary statistics of two asthma GWASs, two ulcerative colitis (UC) GWASs, and two type 2 diabetes GWASs, we show that ReAD identifies more replicable genetic loci that otherwise might be missed when using existing methods that do not account for the LD structure. These identified association signals pinpoint potential loci related to metabolisms and immunity.

Material and methods

Method overview

ReAD takes p values from two independent GWASs with the same phenotype as input. Suppose we have J SNPs with corresponding p values $(p_{1j},p_{2j}), j = 1,\ldots,J$. We aim to identify replicable SNPs associated with the phenotype in both studies. Our method can handle SNPs in the whole genome where J is in the order of millions. We use θ_{ij} to represent the inferred association status of SNP j in study i. For each SNP, we consider its association analysis results are replicable if its corresponding θ values are consistently 1. The correlations between θs within a study are caused by LD among tested SNPs, and we model their dependence structure using a Markov chain. This is an effective way to model the correlations between observed p values. Given the observed p values are from both studies, the overall model structure can be represented by an HMM.[36]

We present our schematic in Figure 1. We use $s_j \in \{0,1,2,3\}$ to denote the joint inferred association status for SNP j, where $s_j = 0$ if $\theta_{1j} = \theta_{2j} = 0$, $s_j = 1$ if $\theta_{1j} = 0$ and $\theta_{2j} = 1$, $s_j = 2$ if $\theta_{1j} = 1$ and $\theta_{2j} = 0$, and $s_j = 3$ if $\theta_{1j} = \theta_{2j} = 1$. The composite null for replicability analysis corresponds to $s_j \in \{0,1,2\}$. To capture the local dependence of LD structure among SNPs, we impose a four-state HMM on $\boldsymbol{s} = (s_1,\ldots,s_J)$. The transition matrix is denoted 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 given by a_{kl}, and $\sum_{l=0}^{3} a_{kl} = 1$ for all k. An efficient EM algorithm in combination with the forward-backward procedure and PAVA is developed to estimate the unknown parameters and functions. We use the posterior probability of being replicability null, $\mathrm{rLIS}_j, j = 1,\ldots,J$, as the test statistic and obtain $\widehat{\mathrm{rLIS}}_j$ for all SNPs. By applying a step-up procedure on $\widehat{\mathrm{rLIS}}_j$, $j = 1,\ldots,J$, we get powerful testing results while controlling the FDR. More details of ReAD can be found below and in Note S1.

The HMM for replicability analysis

Suppose there are J SNPs in two independent GWASs. We are interested in testing whether the jth 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 jth SNP ($j = 1,\ldots,J$) is inferred associated with the phenotype in study i ($i = 1,2$) and $\theta_{ij} = 0$ otherwise. We use s_j ($j = 1,\ldots,J$) to denote the joint status.

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

The replicability null hypotheses is

$$H_{0j} : s_j \in \{0,1,2\}, j = 1,\ldots,J. \qquad \text{(Equation 1)}$$

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

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

where f_0 is the probability density function of p values when $\theta_{1j} = \theta_{2j} = 0$, and f_1 and f_2 are the p value density functions under non-null in study 1 and study 2, respectively. We assume f_0 follows the standard uniform distribution and impose the following monotone likelihood ratio condition.[37–39]

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

This condition naturally arises as small p values indicate evidence against the null. To capture the LD structure among SNPs,

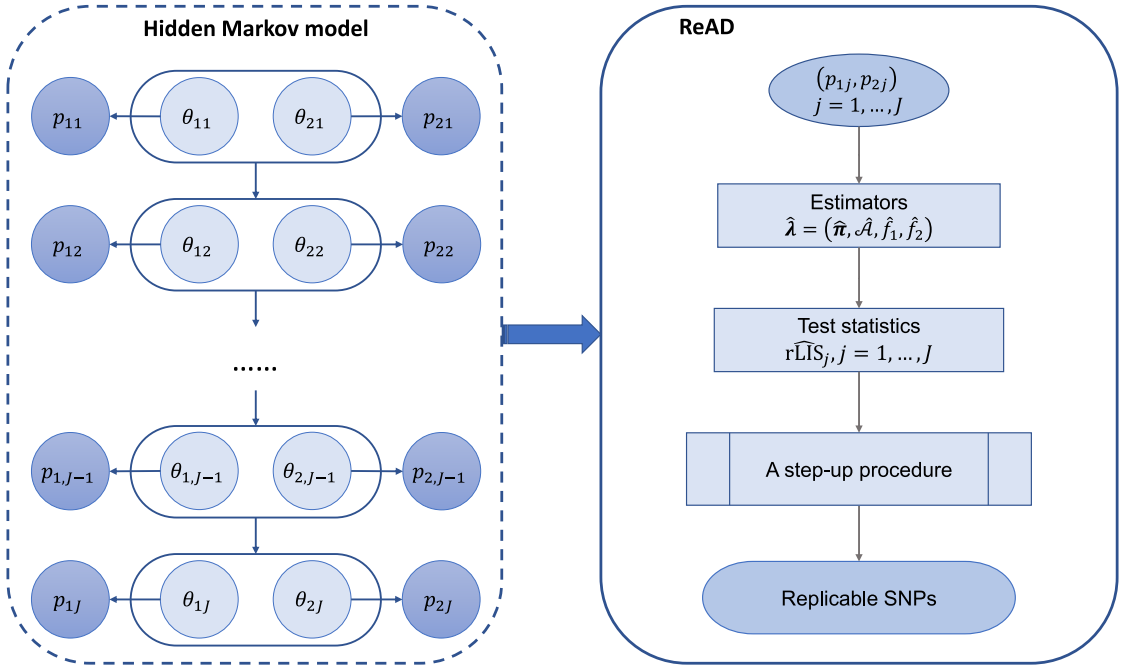

Figure 1. Schematic of ReAD

θ_{ij} represents the inferred association status of SNP j ($j = 1, \ldots, J$) in study i ($i = 1,2$). p_{ij} represents the p value of SNP j (j = 1,..., 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 an HMM.

we assume that $\boldsymbol{s} = (s_1, \ldots, s_J)$ follows a four-state stationary, irreducible, and aperiodic HMM. The transition probabilities

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

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 jth SNP are assumed to be conditionally independent satisfying

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

Based on the mixture model as in Equation 2, we have

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

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[40]) implies that

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

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

FDR control for replicability analysis accounting for LD

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,

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

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[41] 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 $j = 1, \ldots, J$ recursively by

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

and

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

Hence, we have

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

The rejection rule can be written as

$$\delta_j = I(\mathrm{rLIS}_j \le t), j = 1, \ldots, J,$$

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

$R(t) = \sum_{j=1}^{J} I(\mathrm{rLIS}_j \le t)$ and $V(t) = \sum_{j=1}^{J} I(\mathrm{rLIS}_j \le t, s_j \in \{0, 1, 2\})$, respectively. We have

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

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

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

and the FDR can be estimated by $\frac{1}{k}\sum_{j=1}^{k} \mathrm{rLIS}_{(j)}$. We shall use the following step-up procedure to control the FDR at level q.[20]

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

$$\text{then reject all } H_{(j)} \text{ for } j = 1, \ldots, \widehat{k}.$$

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

Data-driven testing procedure

To estimate the unknown parameters and functions in $\boldsymbol{\lambda}$, we first define two posterior probabilities, $\gamma_j(s_j) = \mathbb{P}_{\boldsymbol{\lambda}}(s_j | \boldsymbol{p}_1, \boldsymbol{p}_2)$ and $\xi_j(s_j, s_{j+1}) = \mathbb{P}_{\boldsymbol{\lambda}}(s_j, s_{j+1} | \boldsymbol{p}_1, \boldsymbol{p}_2)$. 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 probabilities

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

and

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

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

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

We develop a non-parametric EM algorithm[33] to estimate the unknowns $\boldsymbol{\lambda} = (\boldsymbol{\pi}, \mathcal{A}, f_1, f_2)$ under the monotone likelihood ratio constraint (Equation 3). With an appropriate initialization of the unknowns, $\boldsymbol{\lambda}^{(0)} = (\boldsymbol{\pi}^{(0)}, \mathcal{A}^{(0)}, f_1^{(0)}, f_2^{(0)})$, the EM algorithm proceeds by 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

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

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

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

We can update each component alternatingly. 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

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

and

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

where $\mathbb{H}$ is a set of monotonic non-increasing density functions.[37–39] We solve Equations 5 and 6 independently using the non-parametric maximum likelihood estimation implemented with PAVA.[34]

The E-step and M-step are conducted iteratively until convergence. Detailed derivations of the algorithm are presented in Note S1. With the estimate $\widehat{\boldsymbol{\lambda}} = \{\widehat{\boldsymbol{\pi}}, \widehat{\mathcal{A}}, \widehat{f}_1, \widehat{f}_2\}$, we can calculate the test statistics $\widehat{\mathrm{rLIS}}_j = \mathbb{P}_{\widehat{\boldsymbol{\lambda}}}(s_j \in \{0, 1, 2\} | \boldsymbol{p}_1, \boldsymbol{p}_2)$. Let $\widehat{\mathrm{rLIS}}_{(1)} \le \cdots \le \widehat{\mathrm{rLIS}}_{(J)}$ be the order statistics of $\widehat{\mathrm{rLIS}}_j$, and denote $H_{(1)}, \ldots, H_{(J)}$ as the corresponding H_{0j}. The data-driven testing procedure works as follows.

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

$$\text{and reject } H_{(i)} \text{ for } i = 1, \ldots, \widehat{k}.$$

Data analysis details

We apply ReAD to analyze two pairs of published GWAS datasets for identifying replicable associations, including two GWAS datasets for asthma and two GWAS datasets for UC. Informed consent was obtained from participants for all studies.[42–44] In the following, the locations of SNPs are mapped to Genome Assembly GRCh38/hg38.

Asthma GWASs

Asthma is a complex bronchial disease characterized by chronic inflammation and narrowing of the airways, which is caused by a combination of environmental and genetic factors. The prevalence of asthma varies across different populations and ethnicities. We use publicly available GWAS summary statistics for asthma from the Trans-National Asthma Genetic Consortium (TAGC) (https://www.ebi.ac.uk/gwas/downloads/summary-statistics) and UK Biobank (UKBB) (http://www.nealelab.is/uk-biobank) to conduct replicability analysis. The TAGC consortium deposits HapMap2-imputed, ancestry-specific, meta-analysis data from ethnically diverse populations.[43] The TAGC asthma GWAS data with high-density genotyped and imputed SNP based on the European-ancestry comprises 8,843,303 genetic variants for 19,954 asthma cases and 107,715 controls. UKBB is a large-scale

prospective cohort study with over half a million participants aged 40–69 years from the United Kingdom between 2006 and 2010.[45] The imputed asthma GWAS from UKBB contains summary statistics for 8,856,162 genetic variants measured on 39,049 self-reported asthma cases and 298,070 controls.

UC GWASs

Inflammatory bowel disease is a chronic, relapsing intestinal inflammatory disease. It has the highest age-standardized prevalence rate in the US followed by the UK[46] with increasing prevalence in Asia and developing countries.[47] UC is one of the two main forms of inflammatory bowel disease. We conduct replicability analysis of publicly available GWAS summary statistics for UC from the International Inflammatory Bowel Disease Genetics Consortium (IIBDGC) (https://www.ibdgenetics.org/) and the UKBB. The IIBDGC GWAS analyses 8,857,076 SNPs from 6,968 UC cases and 20,464 population controls of European descent.[48] The imputed UKBB GWAS data contain summary statistics of 8,856,162 SNPs genotyped on 1,795 self-reported UC cases and 335,324 controls from the United Kingdom.

Results

Simulation study

Simulation I

In simulation I, we evaluated the FDR and statistical power of ReAD based on the rLIS statistic across two studies. Here, power is defined as the average proportion of true discoveries among the total number of non-null hypotheses. We compare the FDR and power of ReAD with several replicability analysis methods developed under independence, including the *ad hoc* BH method, the MaxP method based on $P_{\max}$,[23] and the STAREG method based on the local false discovery rate (Lfdr).[49] Details of these methods can be found in Note S2. An extensive comparison with a replicability analysis method developed under dependence, repLIS,[28] can be found in Note S3.

In each simulation, the hidden states of 10,000 SNPs were generated from a four-state Markov chain. A detailed description of the data-generating process is provided in Note S3. In all simulations, we fix the initial distribution of four states as $\boldsymbol{\pi}^0 = (0.9, 0.025, 0.025, 0.05)$. The signals from two studies are generated from normal distributions with mean μ_i and variance $\sigma_i^2, i = 1, 2$. We vary the transition matrix $\mathcal{A} = \{a_{kl} : k, l = 0, 1, 2, 3\}$ and μ_2 while fix $\mu_1 = 2$, and $\sigma_1 = \sigma_2 = 1$, empirical FDR, and power are calculated from 100 replications for each setting. The results are summarized in Figure 2 (top: FDR; bottom: power). In Figure 2, each row corresponds to a different a_{00}, and each column corresponds to a different a_{33}. In each panel, we set μ_2 to 1.5,2, or 3. At FDR level 0.05, we observe that the *ad hoc* BH fails to control the FDR. MaxP is overly conservative across all settings. STAREG has a slight FDR inflation in some settings. By accounting for the local dependence structure via the rLIS statistic, ReAD properly controls the FDR and has substantial power gain compared to competing methods. The powers of all methods increase as μ_2 increases.

The forward-backward procedure of HMM implies that a small rLIS does not occur alone but in clusters. Therefore, ReAD tends to identify the entire cluster of genotype-phenotype associations. Such clusters are unlikely to occur by chance and are more plausible biological signals. To illustrate this, p values for two studies are generated following the above strategy by setting $a_{00} = 0.9, a_{33} = 0.7$, and $\mu_2 = 2$. We compare three methods for testing the composite replicability null hypotheses across two studies: the MaxP method,[23] the STAREG method,[49] and the ReAD method. Figure 3 presents results of different methods in one replication. It can be seen that MaxP is extremely conservative, which only identifies one single signal; STAREG rejects individual hypotheses with very small p values in both studies, whereas ReAD can identify clusters of replicable signals.

Simulation II

By incorporating the LD structure in GWASs through HMM, the rLIS statistic integrates information from adjacent locations. Therefore, the rankings of SNPs based on rLIS are different from the rankings from MaxP (based on $P_{\max}$) and STAREG (based on Lfdr). In simulation II, we perform simulation studies to evaluate FDR and power in GWASs with realistic LD patterns among SNPs. Data for two studies are generated based on two SNP matrices from the Genetic European Variation in Disease (GEUVADIS) project[50] at https://www.internationalgenome.org/data-portal/data-collection/geuvadis. The first genotype matrix is collected from 78 Utah residents (CEPH) with Northern and Western European ancestry (CEU), and the second genotype matrix is measured from 89 Finnish in Finland (FIN). CEU and FIN are both sub-populations of the European ancestry population, therefore they may have similar LD structures. Based on the CEU and FIN genotype matrices, we filter out SNPs with the same genotypes in all samples and obtain genotypes of 16,764 SNPs in both studies. We specify 4 causal SNPs that are approximately independent of each other in each study, 3 of which are replicable in both studies. In each study, for the ith subject ($i = 1, \ldots, 78$ in the CEU study and $i = 1, \ldots, 89$ in the FIN study), we generate continuous phenotypes using the linear regression model

$$y_i = \beta_0 + \sum_{k=1}^{4} G_{ik}^c \beta_k + \epsilon_i,$$

where β_0 is the intercept, $G_{i1}^c, \ldots, G_{i4}^c$ are the genotypes of the ith subject for the 4 causal SNPs, $\beta_1, \ldots, \beta_5$ are regression coefficients, and ϵ_i is an error term generated from $N(0, 1)$, a standard normal distribution. The intercept and the regression coefficients of causal SNPs $\beta_k, k = 0, 1, \ldots, 4$ are set to 1.5. The p values of 16,764 SNPs in two studies are obtained by a marginal regression of each SNP on the phenotype.

To evaluate FDR and power, following the clumping procedure in GWAS,[51] we define clustered signals as SNPs in LD ($r^2 \geq 0.5$) with pre-specified replicable causal SNPs. In addition, SNPs in LD ($r^2 \geq 0.5$) with clustered signals are added to the signal sets. We evaluate the FDR control and

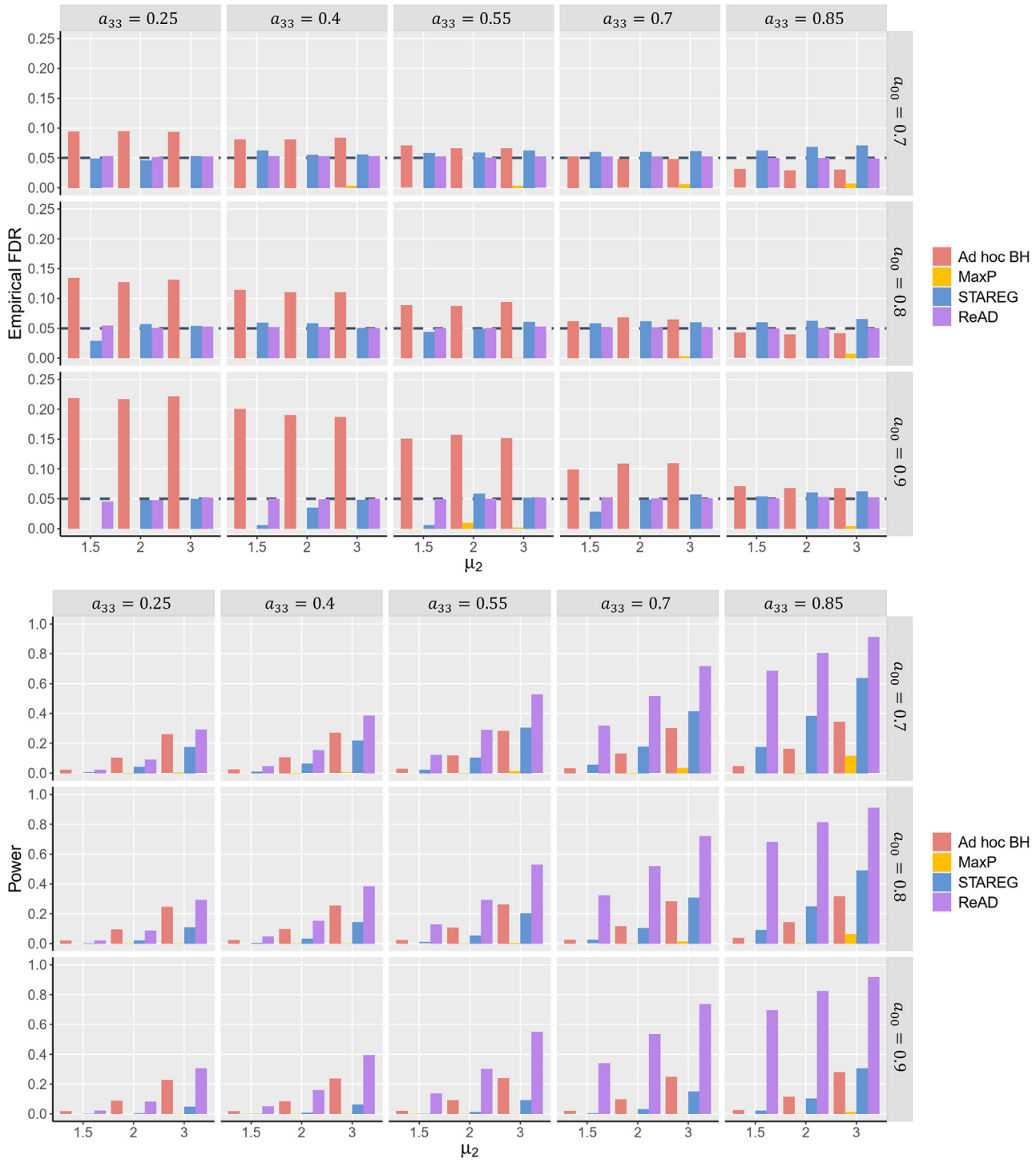

Figure 2. FDR control and power comparison of different methods

FDR control and power comparison are plotted in the top and bottom panels, respectively. Simulations were conducted under the setting of $J = 10,000$ and $\mu_1 = 2$. The initial distribution of four states is $\boldsymbol{\pi}^0 = (0.9, 0.025, 0.025, 0.05)$. The horizontal dashed line in the FDR plot represents the target FDR level of 0.05, and the results were calculated over 100 replications. Each row and column correspond to different transition probabilities, a_{00} and a_{33}, respectively. In each panel, we vary μ_2 from 1.5 to 3.

power of the top K hits based on the maximum of p values ($P_{\max}$), Lfdr, and rLIS from 100 runs. We define empirical FDR as the average proportion of false discoveries selected by the top K hits and power as the average proportion of true discoveries over the number of true signals. The estimated FDR of top K hits is obtained by averaging the ordered $P_{\max}$,Lfdr, and rLIS from the smallest to the Kth value. We set $K = 10,20,\ldots,1000$. Figure 4 summarizes the simulation results. We observe that MaxP is conservative, ReAD controls the FDR properly, and STAREG has FDR inflation for larger

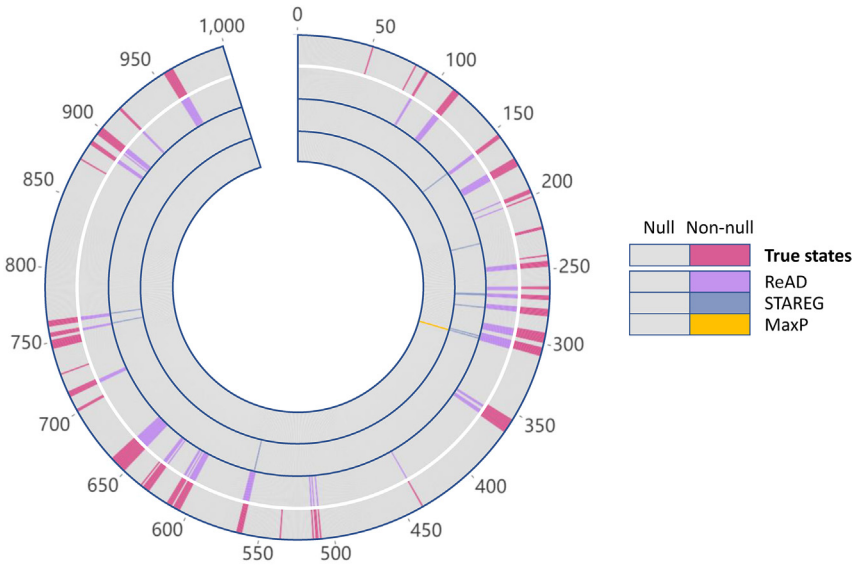

Figure 3. Methods comparison for cluster identification

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.

K. ReAD has the highest power based on selecting the top *K* hits.

We conduct additional simulation studies with real genotypes of samples from multiple population groups to assess the sensitivity and specificity of our method with respect to different LD structures. The detailed results are summarized in Note S3.4 and Figures S3 and S4. We observe that ReAD has robust performance with respect to variations of LD patterns in closely related populations. However, for distant populations, ReAD can suffer significant power loss.

Data analysis

Replicability analysis of asthma GWASs

We first perform replicability analysis on the asthma GWAS data obtained from TAGC and UKBB. We filter out SNPs with minor allele frequency (MAF) smaller than 0.05, resulting in 6,234,241 SNPs in the TAGC study and 6,242,120 SNPs in UKBB. After taking the intersection of SNPs in the two studies, we obtain paired p values of 6,222,195 SNPs to conduct replicability analysis.

As the *ad hoc* BH does not control FDR, we apply MaxP and STAREG on the paired p values for comparison. At FDR level 5×10^{-8}, MaxP identifies 2,853 SNPs, which are also identified by STAREG and ReAD. Compared to MaxP, STAREG identifies 909 additional significant SNPs. By capturing the local LD structure through HMM, ReAD identifies 10,084 significant SNPs, 6,328 of which are missed by MaxP or STAREG. This demonstrates the improved power of ReAD. The NHGRI-EBI GWAS Catalog[52] reported associations with asthma in published GWASs at SNP level and locus (cytogenetic region) level, which can be used as a validation criterion for the replicability analysis.

Among the 6,328 SNPs uniquely identified by ReAD, 158 are recorded in GWAS Catalog,[52] and 6,103 SNPs locate in known loci tagged by SNPs in GWAS Catalog. For the remaining 67 SNPs several of them can be mapped to asthma-associated genes. For instance, SNP rs864537 (rLIS: 3.2e-07; TAGC p value: 3.2e-05; UKBB p value: 2.0e-11) mapped to gene *CD247* indicates a significant locus associated with asthma.[53] Three SNPs, rs12700390 (rLIS: 2.7e-07; TAGC p value: 2.1e-05; UKBB p value: 2.6e-07), rs12700391 (rLIS: 3.3e-07; TAGC p value: 2.1e-05; UKBB p value: 2.8e-07), and rs7781534 (rLIS: 2.8e-07; TAGC p value: 2.8e-05; UKBB p value: 2.8e-07) are closest to gene *IL6*, which is a potential contributor to asthma and other inflammatory pulmonary diseases.[54,55]

To assess the replicability of GWAS loci across two studies, we state that a locus is replicated if at least one SNP within it is identified significantly replicable and is reported associated to asthma in the GWAS Catalog. If a locus contains multiple significant SNPs, the SNP with the strongest association is considered as the lead SNP. For instance, if we use STAREG with Lfdr as the test statistic, the SNP with the smallest Lfdr is the lead SNP. In this criterion, at FDR level 5×10^{-8}, MaxP identifies 12 loci, which are also identified by STAREG and ReAD. STAREG identifies 3 additional loci. ReAD identifies 28 genetic loci with replicable asthma associations, including 13 loci could not be detected by other methods. Figure 5 presents the Manhattan plots of MaxP, STAREG, and ReAD. In Figure 5, the vertical axis are $-\log_{10}$ transformations of test statistics for replicability analysis, i.e., $P_{\max}$ for MaxP, Lfdr for STAREG, and rLIS for ReAD.

Table 1 displays main characteristics of the 28 cytogenetic regions identified by ReAD. The mapped gene denotes genes overlapping or closest to the lead SNP in the identified locus. The 15 loci only identified by ReAD harbor signals closely related to asthma. For example, the lead SNP in locus 2p25.1, rs10174949, is in the intron of gene *LINC00299* and plays an important role in atopic dermatitis, including asthma, hay fever, and eczema in European and UK populations.[56–58] The 8q21.13 region is reported to be associated with asthma and hay fever in a European-ancestry study.[59] The lead SNP rs10957979 lies between gene *RPL13AP18* (chr8:80,265,528–80,266,155) and gene *RNU6-1213P* (chr8:80,405,516–80,405,609), and its association with asthma has been observed in several European-ancestry studies.[43,60]

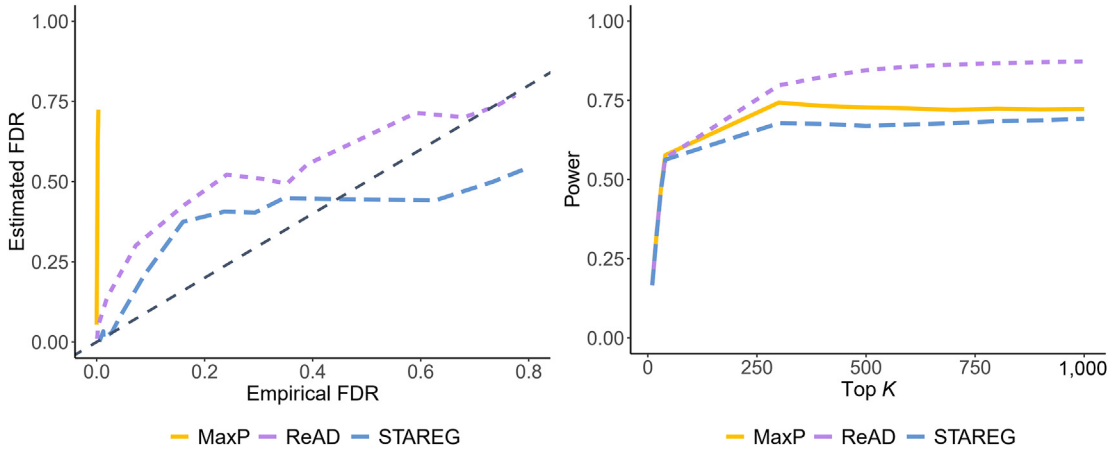

Figure 4. FDR control and power comparison of different methods

Replicability analysis of UC GWASs

Next, we perform replicability analysis of UC GWAS data from IIBDGC and UKBB. We filter out SNPs with MAF smaller than 0.05, resulting in 6,243,744 SNPs in the IIBDGC study and 6,242,120 SNPs in the UKBB. We use the paired p values of 6,232,147 SNPs common to both studies as input for replicability analysis.

We apply MaxP, STAREG, and ReAD on the paired p values. At FDR level 5×10^{-8}, MaxP identifies 1,239 significant SNPs in 1 locus. STAREG identifies 1,542 significant SNPs in 2 loci, one of which is also detected by MaxP. ReAD identifies 3,307 significant SNPs in 7 genetic loci, including 5 loci that are not detected by MaxP or STAREG. Figure 6 presents the Manhattan plots of MaxP, STAREG, and ReAD. ReAD uniquely identifies 1,766 additional UC-associated SNPs, among which 11 are recorded in the GWAS Catalog.[52] The remaining 1,755 SNPs locate in known loci tagged by SNPs in the GWAS Catalog.

We assess the replicability of genetic loci identified by different methods in the GWAS Catalog.[52] Table 2 presents the main characteristics of the 7 replicable genetic loci identified by ReAD. UC associations of these loci in cohorts of European descent have been reported in the literature. For instance, the lead SNP of loci 6p21.32, rs6927022, is in the intron of gene *HLA-DQA1*, and the HLA complex is associated with multiple risk alleles for inflammatory bowel disease, including UC.[61–63] The lead SNP harbored in loci 1q23.3, rs1801274, is only identified by ReAD and has confirmed associations with UC in several European-ancestry studies.[48,64,65] We have additional validations in DisGeNET, a versatile platform that contains a comprehensive catalog of genes and variants associated with human diseases.[66] Many mapped genes of the lead SNP only identified by ReAD have been reported to be associated with UC, such as *FCGR2A* in locus 1q23.3, *IL23R* in locus 1p31.3, *IL10* in locus 1q32.1, and *MST1* in locus 3p21.31.

Replicability analysis can be conducted on a wider range of traits beyond asthma and UC. To illustrate, we perform additional data analysis on GWAS data for type 2 diabetes. Details of the data and analysis can be found in Note S4 and Figure S5.

Discussion

In this paper, we present ReAD, an efficient method accounting for the LD structure to identify replicable associations from two GWAS datasets. We conducted extensive simulation studies and analyzed GWAS datasets for 3 traits.

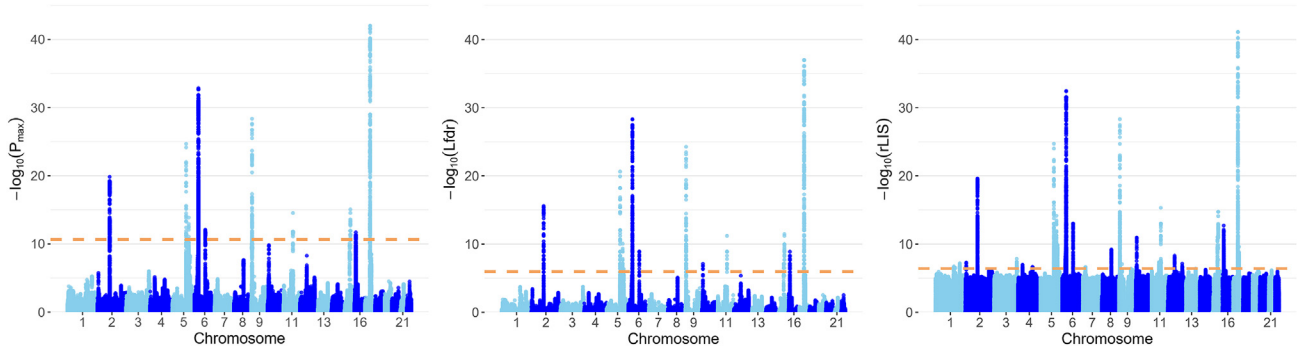

Figure 5. The Manhattan plots of asthma GWASs based on $P_{\max}$, Lfdr, and rLIS
The dashed horizontal lines denote the FDR cutoffs of 5×10^{-8} produced by MaxP, STAREG, and ReAD, respectively.

Table 1. Main characteristics of the 28 loci associated with asthma in the European-ancestry TAGC and UKBB GWASs identified by ReAD

Locus	Lead SNP	Lead SNP location	Mapped gene	P_{max}	Lfdr	rLIS
Replicable asthma loci identified by all methods						
2q12.1	rs3771180	chr2:102,337,157	*IL18R1, IL1RL1*	1.5e-20	2.5e-16	2.5e-20
5q22.1	rs10455025	chr5:111,069,301	*BCLAF1P1, TSLP*	2.0e-25	2.4e-21	1.9e-25
5q31.1	rs20541	chr5:132,660,272	*IL13, TH2LCRR*	1.4e-14	9.1e-11	7.0e-15
6p21.32	rs17843604	chr6:32,652,506	*HLA-DQA1, HLA-DQB1*	2.2e-33	5.0e-29	3.8e-33
6p21.33	rs2596464	chr6:31,445,184	*LINC01149*	1.8e-13	8.0e-10	4.2e-13
6q15	rs2325291	chr6:90,276,967	*BACH2*	8.6e-13	1.3e-09	1.0e-13
9p24.1	rs992969	chr9:6,209,697	*GTF3AP1, IL33*	4.3e-29	5.5e-25	4.8e-29
11q13.5	rs2155219	chr11:76,588,150	*LINC02757, EMSY*	2.9e-15	6.3e-12	4.8e-16
15q22.33	rs17228058	chr15:67,157,967	*SMAD3*	2.9e-15	6.3e-12	1.8e-15
16p13.13	rs12935657	chr16:11,125,184	*CLEC16A*	2.1e-12	1.3e-09	2.0e-13
17q12	rs2941522	chr17:39,754,115	*GRB7, IKZF3*	1.5e-38	3.6e-34	3.0e-38
17q21.1	rs2305479	chr17:39,905,964	*GSDMB*	1.0e-42	1.1e-37	8.0e-42
Replicable asthma loci identified by ReAD and STAREG but not by MaxP						
5q31.3	rs7705042	chr5:142,112,854	*NDFIP1*	6.8e-09	4.4e-07	5.1e-10
10p14	rs1775553	chr10:9,012,362	*LINC00709, LINC02676*	2.0e-10	1.5e-07	1.2e-11
15q22.2	rs11071558	chr15:60,777,222	*RORA*	8.3e-11	8.1e-08	9.9e-12
Replicable asthma loci only identified by ReAD						
1q32.1	rs7555556	chr1:203,121,848	*ADORA1*	5.5e-06	9.3e-04	6.9e-08
1q21.3	rs7521458	chr1:154,435,237	*IL6R*	1.1e-04	6.6e-04	2.9e-07
2p25.1	rs10174949	chr2:8,302,118	*LINC00299*	3.0e-06	5.7e-04	5.3e-08
3q28	rs2889896	chr3:188,384,928	*LPP*	1.0e-06	1.9e-04	1.5e-08
4p14	rs4833103	chr4:38,813,881	*TLR1*	1.5e-05	2.5e-03	2.1e-07
4q27	rs1904522	chr4:122,415,763	*ADAD1*	1.7e-05	2.5e-03	2.3e-07
6p22.2	rs766406	chr6:26,319,360	*H3C9P, H4C8*	2.8e-07	7.6e-05	1.3e-07
6p22.1	rs1117490	chr6:30,202,733	*TRIM26*	2.1e-08	8.6e-06	6.5e-10
8q21.13	rs10957979	chr8:80,377,552	*RPL13AP18, RNU6-1213P*	2.3e-08	8.6e-06	6.5e-10
11q12.2	rs174562	chr11:61,817,672	*FADS2, FADS1*	7.9e-06	1.4e-03	1.1e-7
12q13.3	rs3001425	chr12:57,115,786	*STAT6*	2.9e-07	1.1e-04	8.6e-09
12q24.31	rs625228	chr12:120,840,463	*SPPL3*	9.0e-06	5.8e-04	7.7e-08
17q21.33	rs17637472	chr17:49,384,071	*ZNF652-AS1, PHB1*	3.3e-09	3.3e-06	2.5e-10

The SNP with the strongest association within each locus is called lead SNP. The mapped gene denotes genes (or pseudogenes) overlapping or closest to the lead SNP in the identified locus. The locations of lead SNPs are mapped to Genome Assembly GRCh38/hg38.

Compared to conventional approaches that impose independence assumption among SNPs, ReAD provides effective FDR control. It has a substantial power gain in identifying genuine and replicable genetic loci. It is computationally scalable to hundreds of millions of SNPs and has no tuning parameters.

In this paper, we mainly consider irreplicable GWAS signals due to non-biological factors, e.g., batch effects. Our discussion focuses on assessing the replicability via commonly available summary statistics based on single-SNP association testing, which does not carry LD information. We acknowledge that, in the applications of genetic association analysis, varying LD patterns between studies can lead to inconsistent significant findings at the SNP level. Consequently, our assumption for simulations and real data analysis is that LD patterns in multiple GWASs are similar. This is illustrated in simulation II by using the real genotypes of different populations from the GEUVADIS project. In the presence of varying LD patterns, a more relevant question should be the consistency of underlying association signals within each interrogated locus across original and replication studies. To this end, we apply

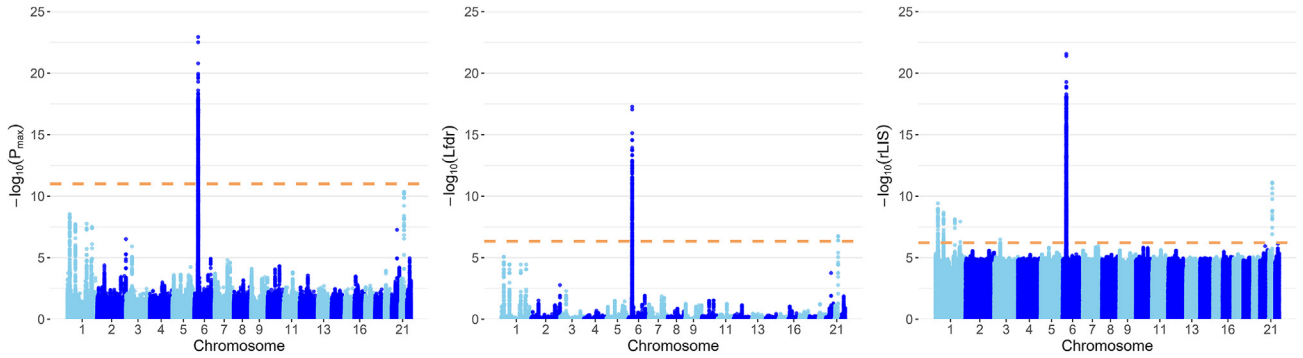

Figure 6. The Manhattan plots of UC GWASs based on P_{max}, Lfdr, and rLIS
The dashed horizontal lines denote the FDR cutoffs of 5×10^{-8} produced by MaxP, STAREG, and ReAD, respectively.

a simple and practical strategy requiring at least one SNP-level finding replicable. With the potential varying LD structures fully accounted for by the proposed HMM, we find this strategy intuitive and effective when applied to genomic loci with proper resolutions (as illustrated by our simulations and real data examples). Nevertheless, this locus-level criterion may be considered overly lenient. In addition, the allele frequencies of associated SNPs can differ drastically in different populations, making *p* values inappropriate metrics to assess the replicability of GWAS findings across populations. We will continue to explore alternative locus-level replicability assessment criteria for different populations in our future work.

In this work, we use repeated significance to assess replicability. We note that applying such a replicability criterion is debated in the scientific community.[67–69] While acknowledging its drawbacks, especially its conservativeness, we note the following context-specific factors. First, despite continued efforts to include more informative statistics summarizing GWAS findings, a large body of historical GWAS findings are *only* reported in *p* values (see GWAS Catalog[52]), which fundamentally limits applying alternative replicability criteria. Second, because complicated unknown confounders, e.g., population stratification and unobserved batch effects in genotyping experiments, often cause false positives in genetic association analysis, the genetics community has consistently advocated conservative replicability criteria to ensure the reliability of GWAS findings.[27,70] Third, we emphasize that our main statistical contribution is to account for the correlation structure between genetic variants, and our work can be naturally extended to applying other alternative replicability criteria.

On a related point, although we exclusively assume that GWAS results are reported in the form of single-SNP testing *p* values throughout this paper, the proposed statistical methodology can be extended to other forms of summary statistics. For example, probabilistic fine-mapping analysis of genetic association signals has become increasingly popular, thanks to the availability of efficient variable selection algorithms.[71–73] The fine-mapping result is typically given as a posterior inclusion probability (PIP) at the individual SNP level. With the ability to construct a Bayesian credible set for each underlying signal within a genomic locus, the PIPs have many advantages over single-SNP *p* values. Theoretically, our work can be straightforwardly extended to this setting by noting the connection that 1 − PIP is equivalent to the Lfdr in the Bayesian perspective. We leave this extension to our future work.

Table 2. Main characteristics of the 7 loci associated with UC in the European-ancestry IIBDGC and UKBB GWASs identified by ReAD

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

The SNP with the strongest association within each locus is called the lead SNP. The mapped gene denotes genes (or pseudogenes) overlapping or closest to the lead SNP in the identified locus. The locations of lead SNPs are mapped to Genome Assembly GRCh38/hg38.

Data and code availability

An R package ReAD implementing the proposed method is available on CRAN (https://CRAN.R-project.org/package=ReAD).

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2024.04.004.

Acknowledgments

We thank the AE and two anonymous reviewers for their helpful comments. The research of Hongyuan Cao is partially supported by NSF/DMS-2311249 and NIH 2UL1TR001427-5. NIH R35-GM138121 supports the research of X.W.

Declaration of interests

The authors declare no competing interests.

Received: August 19, 2023
Accepted: April 4, 2024
Published: May 2, 2024

References

1. McCarthy, M.I., Abecasis, G.R., Cardon, L.R., Goldstein, D.B., Little, J., Ioannidis, J.P., and Hirschhorn, J.N. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat. Rev. Genet. *9*, 356–369.
2. MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A., Milano, A., Morales, J., et al. (2017). The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res. *45*, D896–D901.
3. Ioannidis, J.P., Ntzani, E.E., Trikalinos, T.A., and Contopoulos-Ioannidis, D.G. (2001). Replication validity of genetic association studies. Nat. Genet. *29*, 306–309.
4. NCI-NHGRI Working Group on Replication in Association Studies, Chanock, S.J., Manolio, T., Boehnke, M., Boerwinkle, E., Hunter, D.J., Thomas, G., Hirschhorn, J.N., Abecasis, G., Altshuler, D., et al. (2007). Replicating genotype-phenotype associations. Nature (London) *447*, 655–660.
5. Ioannidis, J.P.A. (2005). Why most published research findings are false. PLoS Med. *2*, e124.
6. Prinz, F., Schlange, T., and Asadullah, K. (2011). Believe it or not: how much can we rely on published data on potential drug targets? Nat. Rev. Drug Discov. *10*, 712.
7. Begley, C.G., and Ellis, L.M. (2012). Raise standards for preclinical cancer research. Nature *483*, 531–533.
8. Freedman, L.P., Cockburn, I.M., and Simcoe, T.S. (2015). The economics of reproducibility in preclinical research. PLoS Biol. *13*, e1002165.
9. Moonesinghe, R., Khoury, M.J., Liu, T., and Ioannidis, J.P.A. (2008). Required sample size and nonreplicability thresholds for heterogeneous genetic associations. USA *105*, 617–622.
10. Huffman, J.E. (2018). Examining the current standards for genetic discovery and replication in the era of mega-biobanks. Nat. Commun. *9*, 5054.
11. Heller, R., and Yekutieli, D. (2014). Replicability analysis analysis for genome-wide association studies. Ann. Appl. Stat. *8*, 481–498.
12. Heller, R., Yaacoby, S., and Yekutieli, D. (2014). repfdr: a tool for replicability analysis for genome-wide association studies. Bioinformatics *30*, 2971–2972.
13. Bogomolov, M., and Heller, R. (2023). Replicability Across Multiple Studies. Stat. Sci. *38*, 602–620.
14. Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Stat. Soc. B *57*, 289–300.
15. Pritchard, J.K., and Przeworski, M. (2001). Linkage disequilibrium in humans: models and data. Am. J. Hum. Genet. *69*, 1–14.
16. Wall, J.D., and Pritchard, J.K. (2003). Haplotype blocks and linkage disequilibrium in the human genome. Nat. Rev. Genet. *4*, 587–597.
17. Wei, Z., Sun, W., Wang, K., and Hakonarson, H. (2009). Multiple testing in genome-wide association studies via hidden markov models. Bioinformatics *25*, 2802–2808.
18. Churchill, G.A. (1992). Hidden markov chains and the analysis of genome structure. Comput. Chem. *16*, 107–115.
19. Sesia, M., Bates, S., Candès, E., Marchini, J., and Sabatti, C. (2021). False discovery rate control in genome-wide association studies with population structure. USA *118*, e2105841118.
20. Sun, W., and Tony Cai, T. (2009). Large-scale multiple testing under dependence. J. Roy. Stat. Soc. B *71*, 393–424.
21. Lonjou, C., Zhang, W., Collins, A., Tapper, W.J., Elahi, E., Maniatis, N., and Morton, N.E. (2003). Linkage disequilibrium in human populations. USA *100*, 6069–6074.
22. Rahimmadar, S., Ghaffari, M., Mokhber, M., and Williams, J.L. (2021). Linkage disequilibrium and effective population size of buffalo populations of iran, turkey, pakistan, and egypt using a medium density snp array. Front. Genet. *12*, 608186.
23. Benjamini, Y., Heller, R., and Yekutieli, D. (2009). Selective inference in complex research. Philos. Trans. A Math. Phys. Eng. Sci. *367*, 4255–4271.
24. Li, Q., Brown, J.B., Huang, H., and Bickel, P.J. (2011). Measuring reproducibility of high-throughput experiments. Ann. Appl. Stat. *5*, 1752–1779.
25. Philtron, D., Lyu, Y., Li, Q., and Ghosh, D. (2018). Maximum rank reproducibility: a nonparametric approach to assessing reproducibility in replicate experiments. J. Am. Stat. Assoc. *113*, 1028–1039.
26. Zhao, Y., Sampson, M.G., and Wen, X. (2020). Quantify and control reproducibility in high-throughput experiments. Nat. Methods *17*, 1207–1213.
27. McGuire, D., Jiang, Y., Liu, M., Weissenkampen, J.D., Eckert, S., Yang, L., Chen, F., GWAS and Sequencing Consortium of Alcohol and Nicotine Use (GSCAN), Berg, A., Vrieze, S., et al. (2021). Model-based assessment of replicability for genome-wide association meta-analysis. Nat. Commun. *12*, 1964.
28. Wang, P., and Zhu, W. (2019). Replicability analysis in genome-wide association studies via cartesian hidden markov models. BMC Bioinf. *20*, 146.
29. Efron, B. (2012). Large-scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction (Cambridge University Press).
30. Chung, D., Yang, C., Li, C., Gelernter, J., and Zhao, H. (2014). GPA: a statistical approach to prioritizing GWAS results by integrating pleiotropy and annotation. PLoS Genet. *10*, e1004787.
31. Rabiner, L., and Juang, B. (1986). An introduction to hidden markov models. IEEE ASSP Mag. *3*, 4–16.

32. Murphy, K.P. (2012). Machine Learning: A Probabilistic Perspective (MIT press).
33. Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the em algorithm. J. Roy. Stat. Soc. B *39*, 1–22.
34. Robertson, T., Dykstra, R.L., and Wright, F.T. (1988). Order restricted statistical inference. In Wiley Series in Probability and Mathematical Statistics (John Wiley and Sons).
35. Busing, F.M.T.A. (2022). Monotone regression: A simple and fast O(n) PAVA implementation. J. Stat. Software *102*, 1–25.
36. Li, N., and Stephens, M. (2003). Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. Genetics *165*, 2213–2233.
37. Sun, W., and Cai, T.T. (2007). Oracle and adaptive compound decision rules for false discovery rate control. J. Am. Stat. Assoc. *102*, 901–912.
38. Cao, H., Sun, W., and Kosorok, M.R. (2013). The optimal power puzzle: scrutiny of the monotone likelihood ratio assumption in multiple testing. Biometrika *100*, 495–502.
39. Cao, H., Chen, J., and Zhang, X. (2022). Optimal false discovery rate control for large scale multiple testing with auxiliary information. Ann. Stat. *50*, 807–857.
40. Durrett, R. (2019). Probability: Theory and Examples (Cambridge university press).
41. Baum, L.E., Petrie, T., Soules, G., and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. Ann. Math. Stat. *41*, 164–171.
42. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The uk biobank resource with deep phenotyping and genomic data. Nature *562*, 203–209.
43. Demenais, F., Margaritte-Jeannin, P., Barnes, K.C., Cookson, W.O.C., Altmüller, J., Ang, W., Barr, R.G., Beaty, T.H., Becker, A.B., Beilby, J., et al. (2018). Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. Nat. Genet. *50*, 42–53.
44. Liu, Z., Liu, R., Gao, H., Jung, S., Gao, X., Sun, R., Liu, X., Kim, Y., Lee, H.-S., Kawai, Y., et al. (2023). Genetic architecture of the inflammatory bowel diseases across east asian and european ancestries. Nat. Genet. *55*, 796–806.
45. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al. (2015). Uk biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. *12*, e1001779.
46. GBD 2017 Inflammatory Bowel Disease Collaborators. (2020). The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990–2017: a systematic analysis for the global burden of disease study 2017. Lancet Gastroenterol. Hepatol. *5*, 17–30.
47. Molodecky, N.A., Soon, I.S., Rabi, D.M., Ghali, W.A., Ferris, M., Chernoff, G., Benchimol, E.I., Panaccione, R., Ghosh, S., Barkema, H.W., and Kaplan, G.G. (2012). Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. Gastroenterology *142*, 46–e30.
48. Liu, J.Z., Van Sommeren, S., Huang, H., Ng, S.C., Alberts, R., Takahashi, A., Ripke, S., Lee, J.C., Jostins, L., Shah, T., et al. (2015). Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat. Genet. *47*, 979–986.
49. Li, Y., Zhou, X., Chen, R., Zhang, X., and Cao, H. (2023). STAREG: an empirical bayesian approach to detect replicable spatially variable genes in spatial transcriptomic studies. Preprint at bioRxiv. https://doi.org/10.1101/2023.05.30.542607.
50. Lappalainen, T., Sammeth, M., Friedländer, M.R., 't Hoen, P.A., Monlong, J., Rivas, M.A., Gonzàlez-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P.G., et al. (2013). Transcriptome and genome sequencing uncovers functional variation in humans. Nature *501*, 506–511.
51. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.
52. Sollis, E., Mosaku, A., Abid, A., Buniello, A., Cerezo, M., Gil, L., Groza, T., Güneş, O., Hall, P., Hayhurst, J., et al. (2023). The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. Nucleic Acids Res. *51*, D977–D985.
53. Khatri, B., Tessneer, K.L., Rasmussen, A., Aghakhanian, F., Reksten, T.R., Adler, A., Alevizos, I., Anaya, J.-M., Aqrawi, L.A., Baecklund, E., et al. (2022). Genome-wide association study identifies sjögren's risk loci with functional implications in immune and glandular cells. Nat. Commun. *13*, 4287.
54. Rincon, M., and Irvin, C.G. (2012). Role of il-6 in asthma and other inflammatory pulmonary diseases. Int. J. Biol. Sci. *8*, 1281–1290.
55. Raita, Y., Zhu, Z., Camargo, C.A., Jr., Freishtat, R.J., Ngo, D., Liang, L., and Hasegawa, K. (2021). Relationship of soluble interleukin-6 receptors with asthma: a mendelian randomization study. Front. Med. *8*, 665057.
56. Zhu, Z., Guo, Y., Shi, H., Liu, C.-L., Panganiban, R.A., Chung, W., O'Connor, L.J., Himes, B.E., Gazal, S., Hasegawa, K., et al. (2020). Shared genetic and experimental links between obesity-related traits and asthma subtypes in uk biobank. J. Allergy Clin. Immunol. *145*, 537–549.
57. Zhu, Z., Lee, P.H., Chaffin, M.D., Chung, W., Loh, P.-R., Lu, Q., Christiani, D.C., and Liang, L. (2018). A genome-wide cross-trait analysis from uk biobank highlights the shared genetic architecture of asthma and allergic diseases. Nat. Genet. *50*, 857–864.
58. Ferreira, M.A., Vonk, J.M., Baurecht, H., Marenholz, I., Tian, C., Hoffman, J.D., Helmer, Q., Tillander, A., Ullemar, V., Van Dongen, J., et al. (2017). Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. Nat. Genet. *49*, 1752–1757.
59. Ferreira, M.A.R., Matheson, M.C., Tang, C.S., Granell, R., Ang, W., Hui, J., Kiefer, A.K., Duffy, D.L., Baltic, S., Danoy, P., et al. (2014). Genome-wide association analysis identifies 11 risk variants associated with the asthma with hay fever phenotype. J. Allergy Clin. Immunol. *133*, 1564–1571.
60. Olafsdottir, T.A., Theodors, F., Bjarnadottir, K., Bjornsdottir, U.S., Agustsdottir, A.B., Stefansson, O.A., Ivarsdottir, E.V., Sigurdsson, J.K., Benonisdottir, S., Eyjolfsson, G.I., et al. (2020). Eighty-eight variants highlight the role of t cell regulation and airway remodeling in asthma pathogenesis. Nat. Commun. *11*, 393.
61. Nowak, J.K., Glapa-Nowak, A., Banaszkiewicz, A., Iwańczak, B., Kwiecień, J., Szaflarska-Popławska, A., Grzybowska-Chlebowczyk, U., Osiecki, M., Kierkuś, J., Hołubiec, M., et al. (2021). Hladqa1* 05 associates with extensive ulcerative colitis at diagnosis: An observational study in children. Genes *12*, 1934.
62. Reinshagen, M., Loeliger, C., Kuehnl, P., Weiss, U., Manfras, B.J., Adler, G., and Boehm, B.O. (1996). Hla class ii gene frequencies in crohn's disease: a population based analysis in germany. Gut *38*, 538–542.

63. Ashton, J.J., Latham, K., Beattie, R.M., and Ennis, S. (2019). the genetics of the human leucocyte antigen region in inflammatory bowel disease. Aliment. Pharmacol. Ther. *50*, 885–900.
64. De Lange, K.M., Moutsianas, L., Lee, J.C., Lamb, C.A., Luo, Y., Kennedy, N.A., Jostins, L., Rice, D.L., Gutierrez-Achury, J., Ji, S.-G., et al. (2017). Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. Nat. Genet. *49*, 256–261.
65. Anderson, C.A., Boucher, G., Lees, C.W., Franke, A., D'Amato, M., Taylor, K.D., Lee, J.C., Goyette, P., Imielinski, M., Latiano, A., et al. (2011). Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. Nat. Genet. *43*, 246–252.
66. Piñero, J., Bravo, À., Queralt-Rosinach, N., Gutiérrez-Sacristán, A., Deu-Pons, J., Centeno, E., García-García, J., Sanz, F., and Furlong, L.I. (2015). Disgenet: a comprehensive platform integrating information on human disease-associated genes and variants. Nucleic Acids Res. *2015*, gkw943.
67. Goodman, S.N. (1992). A comment on replication, p-values and evidence. Stat. Med. *11*, 875–879.
68. Gibson, E.W. (2021). The role of p-values in judging the strength of evidence and realistic replication expectations. Stat. Biopharm. Res. *13*, 6–18.
69. Zhao, Y., and Wen, X. (2021). Statistical assessment of replicability via bayesian model criticism. Preprint at arXiv. https://doi.org/10.48550/arXiv.2105.03993.
70. Skol, A.D., Scott, L.J., Abecasis, G.R., and Boehnke, M. (2006). Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat. Genet. *38*, 209–213.
71. Benner, C., Spencer, C.C.A., Havulinna, A.S., Salomaa, V., Ripatti, S., and Pirinen, M. (2016). Finemap: efficient variable selection using summary data from genome-wide association studies. Bioinformatics *32*, 1493–1501.
72. Wang, G., Sarkar, A., Carbonetto, P., and Stephens, M. (2020). A simple new approach to variable selection in regression, with application to genetic fine mapping. J. R. Stat. Soc. Series B Stat. Methodol. *82*, 1273–1300.
73. Wen, X., Lee, Y., Luca, F., and Pique-Regi, R. (2016). Efficient integrative multi-snp association analysis via deterministic approximation of posteriors. Am. J. Hum. Genet. *98*, 1114–1129.