# An efficient Bayesian meta-analysis approach for studying 

## cross-phenotype genetic associations

Arunabha Majumdar ${ }^{1}$, Tanushree Haldar ${ }^{2}$, Sourabh Bhattacharya ${ }^{3}$, John S. Witte ${ }^{1 *}$<br>${ }^{1}$ Department of Epidemiology and Biostatistics, University of California, San Francisco<br>${ }^{2}$ Institute for Human Genetics, University of California, San Francisco<br>${ }^{3}$ Interdisciplinary Statistical Research Unit, Indian Statistical Institute, Kolkata

Address for correspondence:
John S. Witte
Department of Epidemiology and Biostatistics
University of California, San Francisco
$14503^{\text {rd }}$ St, Box 3110
San Francisco, CA 94158
*e-mail: JWitte@ucsf.edu

## 1 Abstract

Simultaneous analysis of genetic associations with multiple phenotypes may reveal shared genetic susceptibility across traits (pleiotropy). For a locus exhibiting overall pleiotropy, it is important to identify which specific traits underlie this association. We propose a Bayesian meta-analysis approach (termed CPBayes) that uses summary-level data across multiple phenotypes to simultaneously measure the evidence of aggregate-level pleiotropic association and estimate an optimal subset of traits associated with the risk locus. This method uses a unified Bayesian statistical framework based on a spike and slab prior. CPBayes performs a fully Bayesian analysis by employing the Markov chain Monte Carlo (MCMC) technique Gibbs sampling. It takes into account heterogeneity in the size and direction of the genetic effects across traits. It can be applied to both cohort data and separate studies of multiple traits having overlapping or non-overlapping subjects. Simulations show that CPBayes produces a substantially better accuracy in the selection of associated traits underlying a pleiotropic signal than the subset-based meta-analysis ASSET. We used CPBayes to undertake a genome-wide pleiotropic association study of 22 traits in the large Kaiser GERA cohort and detected nine independent pleiotropic loci associated with at least two phenotypes. This includes a locus at chromosomal region 1q24.2 which exhibits an association simultaneously with the risk of five different diseases: Dermatophytosis, Hemorrhoids, Iron Deficiency, Osteoporosis, and Peripheral Vascular Disease. The GERA cohort analysis suggests that CPBayes is more powerful than ASSET with respect to detecting independent pleiotropic variants. We provide an R-package 'CPBayes' implementing the proposed method.

Keywords: Pleiotropy, selection, non-null traits, summary statistics, Bayes factor, MCMC.

## 2 Introduction

Genome-wide association studies (GWASs) have detected loci associated with multiple different traits and diseases (i.e., pleiotropy). For example, pleiotropy has been observed for different types of cancers [Sakoda et al., 2013], immune-mediated diseases [Parkes et al., 2013], and psychiatric disorders [Parkes et al., 2013]. As a specific example, Ellinghaus et al. [2016] demonstrated shared genetic susceptibility to five chronic inflammatory diseases: ankylosing spondylitis, Crohn's disease, psoriasis, primary sclerosing cholangitis, and ulcerative colitis. Pickrell et al. [2016] systematically compared genetic architecture of 42 phenotypes and reported substantial pleiotropy. Analyzing pleiotropy provides a better understanding of shared pathways and biological mechanisms common to multiple different diseases/phenotypes. From the perspective of clinical genetics, the discovery of a locus simultaneously associated with multiple diseases can support the use of a common therapeutic intervention.

When evaluating a group of phenotypes, one may only expect a subset of them to exhibit pleiotropy. For example, the Global Lipids Genetics Consortium [2013] discovered novel pleiotropic loci associated with different subsets of blood lipid traits. In particular, variants in the genes $R S P O 3, F T O, V E G F A, P E P D$ were associated with HDL and triglycerides, but not with total cholesterol or LDL. Hence, in addition to evaluating the evidence of overall pleiotropic association, it is crucial to determine which traits are associated with the risk locus to better interpret the pleiotropic signal. Another important consideration is the availability of individual level data from multiple GWASs of different phenotypes. When accessing individual level data is difficult, one can use more readily available genome-wide summary statistics for various phenotypes.

With such data, pleiotropy can be assessed using fixed-effects meta-analysis approach which assumes that the true effects across studies are homogeneous. However, the effects of a genetic variant on various traits may be more heterogeneous than the effects of a genetic variant on a specific phenotype across different studies. The random-effects meta-analysis allows for heterogeneity but can be under-powered in association studies. Moreover, neither the fixed-effects nor the random-effects meta-analysis facilitates a simultaneous selection of associated/non-null traits. To address this, a subset-based fixed-effects meta-analysis ASSET [Bhattacharjee et al., 2012] simultaneously provides a p-value evaluating the evidence of aggregate-level pleiotropic association and an optimal subset of associated/non-null traits. It can adjust for correlation between summary statistics due to sharing of subjects across studies. Recent studies [Ellinghaus et al., 2016; Carty et al., 2014] have used ASSET as a primary tool for pleiotropy analysis.

In this article, we propose a Bayesian meta-analysis approach (termed CPBayes) that simultaneously provides a measure of the evidence of aggergate-level pleiotropic association and an optimal subset of associated traits underlying a pleiotropic signal. The evidence of aggregate-level pleiotropic association is measured by a Bayes factor (BF) and a posterior probability of null association (PPNA). CPBayes explicitly takes into account correlation between summary statistics. For multiple case-control studies of different diseases, the summary statistics across traits become correlated mainly due to sample overlap between studies (e.g., con-
trols). Similarly, for a cohort study, summary statistics across traits are correlated due to assessing multiple phenotypes for the same group of individuals. CPBayes considers heterogeneity both in the size and direction of genetic effects across phenotypes. It also estimates the posterior probability of each phenotype being associated with the risk locus that quantifies the relative contribution of the traits underlying a pleiotropic signal.

The Bayesian framework of CPBayes is based on the spike and slab prior, which is commonly used due to its appropriateness and simplicity in solving two-class classification problems. The application of the spike and slab prior in genetic association studies is gradually increasing [Zhou et al., 2013; Wen and Stephens, 2014; Vilhjálmsson et al., 2015]. With a spike and slab prior, the spike element represents a null effect, and the slab component represents a non-null effect. The spike part can either be a positive mass at zero (Dirac spike) or be a normal distribution with mean zero and a small variance (continuous spike). We design the Gibbs samplers for these two type of prior spikes for both uncorrelated and correlated summary statistics across traits.

We demonstrate by simulations that the continuous spike offers better accuracy in the selection of associated traits than the Dirac spike. The former is also computationally much faster than the latter due to simpler analytic expressions of the full conditional posterior distributions of the model parameters. Hence, we adopted the continuous spike for constructing CPBayes. We explore the performance of CPBayes in various simulation scenarios and compare its efficiency in selecting the non-null traits compared to ASSET. CPBayes resembles ASSET in that both methods simultaneously draw inference on the evidence of aggregate-level pleiotropic association and on the optimal subset of non-null traits underlying a pleiotropic signal. But, we show here that the key advantage of CPBayes is that it selects the non-null traits with substantially higher specificity (proportion of null traits discarded from the optimal subset) than ASSET while maintaining a good level of sensitivity (proportion of associated traits included in the subset). We also compare CPBayes and ASSET in the analysis of 22 phenotypes in the large Kaiser "Resource for Genetic Epidemiology Research on Adult Health and Aging" (GERA) cohort [dbGaP Study Accession: phs000674.v1.p1]. CPBayes identified nine independent pleiotropic loci associated with at least two phenotypes including a locus at chromosomal region $1 q 24.2$ that exhibits an association with five different diseases: Dermatophytosis, Hemorrhoids, Iron Deficiency, Osteoporosis, and Peripheral Vascular Disease. ASSET identified larger number of independent pleiotropic loci associated with more than one trait, but selected many phenotypes with very weak genetic effects. We provide an R-package 'CPBayes' implementing the proposed method for a general use by other investigators.

## 3 Material and methods

Let $Y_{1}, \ldots, Y_{K}$ denote $K$ phenotypes, $G$ denote genotype at a single nucleotide polymorphism (SNP), and $W$ denote a set of covariates. Suppose, a generalized linear model (GLM) is separately fit for each phenotype as: $g\left(E\left(Y_{j}\right)\right)=\alpha_{j}+\beta_{j} G+\gamma_{j}^{\prime} W, j=1, \ldots, K$. Let $\hat{\beta}_{1}, \ldots, \hat{\beta}_{K}$ denote the estimates (e.g., maximum likelihood estimates) of $\beta_{1}, \ldots, \beta_{K}$ with the corresponding standard errors $s_{1}, \ldots, s_{K}$. Let $\hat{\boldsymbol{\beta}}=\left(\hat{\beta}_{1}, \ldots, \hat{\beta}_{K}\right), \boldsymbol{\beta}=$ $\left(\beta_{1}, \ldots, \beta_{K}\right)$, and $\boldsymbol{s}=\left(s_{1}, \ldots, s_{K}\right)$. Now suppose that we only have the summary statistics (e.g., $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{s}$ ). For a large sample size, we can assume that $\hat{\beta}_{j} \mid \beta_{j}, \sigma_{j} \sim \mathrm{~N}\left(\beta_{j}, \sigma_{j}^{2}\right)$. Since $s_{j}$ is a consistent estimator of $\sigma_{j}$, it is commonly used in place of $\sigma_{j}$. Hence, we assume that $\hat{\beta}_{j} \mid \beta_{j} \sim \mathrm{~N}\left(\beta_{j}, s_{j}^{2}\right)$. If $\hat{\beta}_{1}, \ldots, \hat{\beta}_{K}$ are uncorrelated, $\hat{\beta}_{j} \mid \beta_{j} \stackrel{\text { ind }}{\sim} \mathrm{N}\left(\beta_{j}, s_{j}^{2}\right) ; j=1, \ldots, K$. If $\hat{\beta}_{1}, \ldots, \hat{\beta}_{K}$ are correlated with a covariance matrix $S$, we assume that $\hat{\boldsymbol{\beta}} \mid \boldsymbol{\beta} \sim \operatorname{MVN}(\boldsymbol{\beta}, S)$.

### 3.1 Continuos spike

The continuous spike and slab prior in our context [George and McCulloch, 1993; Malsiner-Walli and Wagner, 2011] is described as: for $j=1, \ldots, K$,

$$
\begin{gather*}
\beta_{j} \mid z_{j}, \tau, d \stackrel{\text { ind }}{\sim}\left(1-z_{j}\right) \times N\left(0, \tau^{2}\right)+z_{j} \times N\left(0,\left(\frac{\tau}{d}\right)^{2}\right) ; \tau>0,0<d<1,\left(\frac{\tau}{d}\right)^{2}>\tau^{2} \\
P\left(z_{j}=1 \mid q\right)=q ; P\left(z_{j}=0 \mid q\right)=(1-q) ; 0<q<1  \tag{1}\\
q\left|c_{1}, c_{2} \sim \operatorname{Beta}\left(c_{1}, c_{2}\right) ; d\right| e_{1}, e_{2} \sim \operatorname{Beta}\left(e_{1}, e_{2}\right)
\end{gather*}
$$

The latent variable $z_{j}$ denotes the association status of $Y_{j}$. When $z_{j}=0, \beta_{j} \sim N\left(0, \tau^{2}\right)$, and when $z_{j}=1$, $\beta_{j} \sim N\left(0,\left(\frac{\tau}{d}\right)^{2}\right)$, where $\left(\frac{\tau}{d}\right)^{2}>\tau^{2}$. The usefulness of such a formulation is that $\tau$ can be set small enough so that, if $z_{j}=0,\left|\beta_{j}\right|$ would probably be very small to safely be considered as zero ( $Y_{j}$ is not associated with the SNP), and $d$ can be chosen sufficiently small (so $\frac{1}{d} \gg 1$ ) such that, if $z_{j}=1, \beta_{j}$ can be considered as non-zero ( $Y_{j}$ is associated with the SNP). Of note, without the latent variables $Z=\left(z_{1}, \ldots, z_{K}\right)$, it is not possible to distinguish between a null and a non-null effect, because $\beta_{j}$ is always non-zero under both the spike and slab distributions. The proportion of traits having a non-null genetic effect is denoted by $q$. For simplicity and reduction in computational cost, we consider $\tau$ as fixed. We also choose $c_{1}=c_{2}=e_{1}=e_{2}=1$ which correspond to the uniform $(0,1)$ distribution. The parameter $d$ is updated in a given range subject to the slab variance $\left(\frac{\tau}{d}\right)^{2}$ varying in a pre-fixed range. We describe the continuous spike and slab prior in the context of modeling pleiotropy with an example diagram in Figure 1.

### 3.2 Dirac spike

The Dirac spike in the current context [Mitchell and Beauchamp, 1988; Malsiner-Walli and Wagner, 2011] is given by: for $j=1, \ldots, K$,

$$
\begin{gather*}
\beta_{j} \mid q, b \stackrel{i . i . d .}{ }(1-q) \times \delta_{\{0\}}\left(\beta_{j}\right)+q \times \mathrm{N}\left(0, b^{2}\right)  \tag{2}\\
q \mid c_{1}, c_{2} \sim \operatorname{Beta}\left(c_{1}, c_{2}\right), 0<q<1
\end{gather*}
$$

Here, $\delta_{\{0\}}\left(\beta_{j}\right)=1$ if $\beta_{j}=0$, and $\delta_{\{0\}}\left(\beta_{j}\right)=0$ if $\beta_{j} \neq 0$. Clearly, under no association, $\beta_{j}=0$. The proportion of associated traits is given by $q$. Note that, the Dirac spike can be obtained from the continuous spike by first setting $\tau=0$ and $\frac{\tau}{d}=b$ in Equation 1, and then integrating out the latent variables $Z$ from the model.

### 3.3 Statistical inference on pleiotropy by employing MCMC

To perform a fully Bayesian analysis, we implement MCMC by the Gibbs sampling algorithm to generate posterior samples of the model parameters based on which we draw the statistical inference for pleiotropy. We derive the Gibbs samplers for both uncorrelated and correlated summary statistics. Here we describe the inference procedure for the continuous spike. The Gibbs sampling algorithm for the continuous spike (Algorithm 1) is outlined in the Appendix, and the algorithm for the Dirac spike (Algorithm 2) is stated in the supplementary materials. The mathematical derivation of the full conditional posterior distributions underlying the Gibbs samplers are also given in the supplementary materials.

Let $\left\{\boldsymbol{\beta}^{(i)}, Z^{(i)}, q^{(i)}, d^{(i)} ; i=1, \ldots, N\right\}$ denote $N$ posterior samples of $(\boldsymbol{\beta}, Z, q, d)$ obtained by MCMC after a certain burn-in period. First, we want to test the global null hypothesis of no association $\left(H_{0}\right)$ against the global alternative hypothesis of association with at least one trait $\left(H_{1}\right)$. Since, for the continuous spike, the latent association status distinguishes between an association being null or non-null, we set $H_{0}: z_{1}=\ldots=$ $z_{K}=0(Z=0)$ versus $H_{1}$ : at least one of $z_{1}, \ldots, z_{K}=1(Z \neq 0)$.

### 3.3.1 Bayes factor (BF)

Let $D$ denote the summary statistic data at a SNP across traits. The Bayes Factor for testing $H_{1}$ against $H_{0}$ is given by:

$$
\begin{equation*}
\mathrm{BF}=\frac{P\left(D \mid H_{1}\right)}{P\left(D \mid H_{0}\right)}=\frac{P\left(H_{1} \mid D\right)}{P\left(H_{0} \mid D\right)} \frac{P\left(H_{0}\right)}{P\left(H_{1}\right)}=\frac{P(Z \neq 0 \mid D)}{P(Z=0 \mid D)} \frac{P(Z=0)}{P(Z \neq 0)}=\frac{\text { Posterior odds }}{\text { Prior odds }} \tag{3}
\end{equation*}
$$

The posterior odds of $H_{1}$ vs. $H_{0}=\frac{P(Z \neq 0 \mid D)}{P(Z=0 \mid D)}=\frac{1-P(Z=0 \mid D)}{P(Z=0 \mid D)}$, and the prior odds of $H_{1}$ vs. $H_{0}=\frac{P(Z \neq 0)}{P(Z=0)}=$ $\frac{1-P(Z=0)}{P(Z=0)}$. Of note, $P\left(z_{j}=1\right)=E(q)=\frac{c_{1}}{c_{1}+c_{2}}=p_{1}$. Let $p_{0}=1-p_{1}$. Since, $z_{j}$ s are independently distributed in the prior, $P(Z=0)=p_{0}^{K}$, and $P(Z \neq 0)=1-p_{0}^{K}$. So the prior odds $=\frac{1-p_{0}^{K}}{p_{0}^{K}}$. The analytic calculation of $P(Z=0 \mid D)$ is intractable. Hence, we estimate this conditional probability based on the posterior sample of
the model parameters. Note that $P(Z=0 \mid D)=\iiint P(Z=0 \mid \boldsymbol{\beta}, q, d, D) f(\boldsymbol{\beta}, q, d \mid D) d(\boldsymbol{\beta}) d(q) d(d)$. Thus,

$$
\begin{equation*}
P(Z=0 \mid D)=E_{\boldsymbol{\beta}, q, d \mid D} P(Z=0 \mid \boldsymbol{\beta}, q, d, D) \approx \frac{1}{N} \sum_{i=1}^{N} P\left(Z=0 \mid \boldsymbol{\beta}^{(i)}, q^{(i)}, d^{(i)}, D\right) \tag{4}
\end{equation*}
$$

where $\left(\boldsymbol{\beta}^{(i)}, q^{(i)}, d^{(i)}\right)$ denotes the $i^{t h}$ posterior sample of $(\boldsymbol{\beta}, q, d)$ obtained by the MCMC. We note that the full conditional posterior distributions of $z_{1}, \ldots, z_{K}$ are independent (see step 5 in Algorithm 1 and the derivation of full conditional distributions of $z_{1}, \ldots, z_{K}$ in the supplementary materials). Hence, $P\left(Z=0 \mid \boldsymbol{\beta}^{(i)}, q^{(i)}, d^{(i)}, D\right)=\prod_{j=1}^{K} P\left(z_{j}=0 \mid \boldsymbol{\beta}^{(i)}, q^{(i)}, d^{(i)}, D\right)$. This independence property of the full conditional posterior distributions of $z_{1}, \ldots, z_{K}$ is crucial for the explicit estimation of the Bayes factor.

### 3.3.2 Posterior probability of null association (PPNA)

We consider another measure for evaluating the aggregate-level pleiotropic association, termed the posterior probability of null association (PPNA). This is based on the posterior probability of association (PPA) introduced in Stephens and Balding [2009]. The posterior odds (PO) of $H_{1}$ versus $H_{0}$ is $\frac{P\left(H_{1} \mid D\right)}{P\left(H_{0} \mid D\right)}$. The PPA is given by:

$$
\begin{equation*}
\mathrm{PPA}=\frac{\mathrm{PO}}{1+\mathrm{PO}} \tag{5}
\end{equation*}
$$

We define PPNA $=1$ - PPA, which can be viewed as a Bayesian analog of the p-value [Stephens and Balding, 2009]. If the data supports $H_{1}$, the PPNA should be close to zero, and if the data supports $H_{0}$, it should be close to one (similar to a p-value). The posterior odds is computed in the same way as described above for the Bayes factor.

### 3.3.3 Selection of optimal subset of associated traits

For $i=1, \ldots, N$, let $S_{i}=\left\{Y_{j}: z_{j}^{(i)}=1 ; j=1, \ldots, K\right\}$ denote the subset of associated traits detected in the $i^{\text {th }}$ posterior sample. That subset of traits which is observed with the maximum frequency in the posterior sample is estimated as the optimal subset of associated traits. Note that it is the maximum a posteriori (MAP) estimate of the optimal subset. Let $\mathrm{PPA}_{j}$ denote the trait-specific posterior probability of association which is estimated as $\frac{1}{N} \sum_{i=1}^{N} z_{j}^{(i)}$ for the phenotype $Y_{j} . \mathrm{PPA}_{j}$ provides a better insight into a pleiotropic signal. It quantifies the relative contribution of the traits underlying a pleiotropic signal. We note that, even if a trait is not selected as non-null, the estimated $\mathrm{PPA}_{j}$ for the trait may not be negligible, e.g., $25 \%$. An interpretation of such a phenomenon is that even though the estimated genetic effect on a phenotype was not substantial enough to make into the optimal subset, the possibility of the genetic variant having a pleiotropic effect on the trait along with those in the optimal subset seems promising.

The direction of association between each non-null trait and a genetic variant can be estimated based on the posterior sample of $\boldsymbol{\beta}$. The posterior probability that $Y_{j}$ is positively associated is estimated as the
proportion of positive $\beta_{j}$ among the posterior sample of $\beta_{j} . Y_{j}$ is classified as being positively associated if this estimated proportion is greater than half.

The posterior mean, median, and the $95 \%$ credible interval (Bayesian analog of the frequentist confidence interval) of the true genetic effect on each phenotype can be computed based on the posterior sample of the association parameters.

### 3.4 Specifying the variance of the spike and slab distributions

We set $\tau^{2}$ (the variance parameter of the spike distribution) to a fixed value $10^{-4}$ after extensive experimentation with simulated data. If $\beta\left(\log (\right.$ odds ratio $)$ ) follows $N\left(0,10^{-4}\right)$, then $P\left(0.98<e^{\beta}<1.02\right)=0.954$. It implies that under the spike distribution (no association), the odds ratio for association between a variant and single trait will vary between 0.98 and 1.02 with a prior probability of $95.4 \%$. In the MCMC, we updated the slab variance $\left(\frac{\tau}{d}\right)^{2}$ in the range $(0.8-1.2)$ with the median value equal to 1 . If $\beta \sim N(0,1)$, then $P\left(e^{\beta}<0.95\right.$ or $\left.e^{\beta}>1.05\right)=0.96$, which implies that under the slab distribution (association) with variance one, the odds ratio is smaller than 0.95 (a negative association) or larger than 1.05 (a positive association) with a prior probability of $96 \%$. We also explored other choices for these parameters by simulations, such as, $\tau^{2}=10^{-3}, 10^{-2}$, and $\left(\frac{\tau}{d}\right)^{2}$ in a range ( $\left.0.5-1.0\right),(0.7-1.1)$, etc. The values used here $\left(\tau^{2}=10^{-4},\left(\frac{\tau}{d}\right)^{2}\right.$ in the range $(0.8-1.2)$ ) gave a high level of specificity while maintaining an overall good level of sensitivity. We note that the choice of the spike variance $\left(\tau^{2}\right)$ and the ratio of the slab and spike variances $\left(\frac{1}{d^{2}}\right)$ directly impact the selection accuracy [George and McCulloch, 1993].

### 3.5 Estimating the correlation between summary statistics

The summary statistics across traits can be correlated due to overlap or close genetic relatedness among subjects across different studies. For case-control studies, Zaykin and Kozbur [2010] and Lin and Sullivan [2009] derived a simple formula of correlation among $\hat{\beta}_{1}, \ldots, \hat{\beta}_{K}$. For $k, l \in\{1, \ldots, K\}$ and $k \neq l$,

$$
\begin{equation*}
\operatorname{corr}\left(\hat{\beta}_{k}, \hat{\beta}_{l}\right)=\left(n_{k l}^{(11)} \sqrt{\frac{n_{k}^{(0)} n_{l}^{(0)}}{n_{k}^{(1)} n_{l}^{(1)}}}+n_{k l}^{(00)} \sqrt{\frac{n_{k}^{(1)} n_{l}^{(1)}}{n_{k}^{(0)} n_{l}^{(0)}}}\right) / \sqrt{n_{k} n_{l}} \tag{6}
\end{equation*}
$$

Here $n_{k}^{(1)}, n_{k}^{(0)}$, and $n_{k}$ (or $n_{l}^{(1)}, n_{l}^{(0)}$, and $n_{l}$ ) denote the number of cases, controls, and total sample size for the study of $Y_{k}\left(\right.$ or $\left.Y_{l}\right) ; n_{k l}^{(11)}$ and $n_{k l}^{(00)}$ denote the number of cases and controls shared between the studies of $Y_{k}$ and $Y_{l}$. Let $n_{k l}^{(10)}$ be the number of overlapping subjects that are cases for $Y_{k}$ but controls for $Y_{l}$; similarly, let $n_{k l}^{(01)}$ be the number of shared subjects that are controls for $Y_{k}$ but cases for $Y_{l}$. Here, the above formula can be generalized to:

$$
\begin{equation*}
\operatorname{corr}\left(\hat{\beta}_{k}, \hat{\beta}_{l}\right)=\left(n_{k l}^{(11)} \sqrt{\frac{n_{k}^{(0)} n_{l}^{(0)}}{n_{k}^{(1)} n_{l}^{(1)}}}-n_{k l}^{(10)} \sqrt{\frac{n_{k}^{(0)} n_{l}^{(1)}}{n_{k}^{(1)} n_{l}^{(0)}}}-n_{k l}^{(01)} \sqrt{\frac{n_{k}^{(1)} n_{l}^{(0)}}{n_{k}^{(0)} n_{l}^{(1)}}}+n_{k l}^{(00)} \sqrt{\frac{n_{k}^{(1)} n_{l}^{(1)}}{n_{k}^{(0)} n_{l}^{(0)}}}\right) / \sqrt{n_{k} n_{l}} \tag{7}
\end{equation*}
$$

This formula is accurate when none of the phenotypes $Y_{1}, \ldots, Y_{K}$ is associated with the SNP or environmental covariates. An alternative strategy [Zhu et al., 2015; Pickrell et al., 2016] is based on using genome-wide summary statistics data to estimate the correlation structure, which is useful when the environmental covariates are associated with the phenotypes or the number of cases or controls shared across studies are not available.

### 3.6 A combined strategy for correlated summary statistics

For strongly correlated summary statistics, when a majority of the traits are associated with the risk locus (non-sparse scenario), the Gibbs sampler can sometimes be trapped in a local mode rather than the global mode of the posterior distribution due to possible multi-modality of the posterior distribution of model parameters. We observed this pattern in our simulation study. It may result in an incorrect selection of associated traits, reducing the robustness of CPBayes. We noticed that, in such a scenario, if the summary statistics are assumed to be uncorrelated, the MCMC does not get trapped in a local mode and moves around the global mode. But ignoring the correlation can give a lower Bayes factor and sensitivity of the selected traits. Hence, for correlated summary statistics we combine the correlated and the uncorrelated versions of CPBayes as follows. First, we execute CPBayes considering the correlation among $\hat{\beta}_{1}, \ldots, \hat{\beta}_{K}$. Let $A$ denote the selected subset of non-null traits that contains $K_{1}$ traits. Let $B$ denote the subset of $K_{1}$ traits that have the smallest univariate association p-values. If $A$ and $B$ match, we accept the results; otherwise, we implement CPBayes assuming that $\hat{\beta}_{1}, \ldots, \hat{\beta}_{K}$ are uncorrelated and take the results obtained. Note that, if $A$ is empty, we accept the results obtained by the correlated version of CPBayes.

## 4 Simulation study

We consider multiple case-control studies with or without shared controls. We also consider a cohort study where the data on multiple disease states are available for a group of individuals. First, we specify the simulation model and generate the phenotype and genotype data. After computing the summary statistics based on the simulated data, we assume that only the summary-level data are available. For case-control studies with no overlapping subjects, we implement the uncorrelated version of CPBayes and ASSET. For case-control studies with overlapping subjects or a cohort study, we estimate the correlation structure of summary statistics based on the formula given in the Equation 7, and use the combined strategy of CPBayes and the correlated version of ASSET.

Let $K_{1}$ denote the number of associated phenotypes among $K$ phenotypes. Suppose, $K_{1}^{+}$traits are positively associated and $K_{1}^{-}$traits are negatively associated ( $K_{1}=K_{1}^{+}+K_{1}^{-}$). We consider two different choices of the minor allele frequency (MAF) at the risk SNP (denoted by $m$ ): 0.3 and 0.1 .

For non-overlapping case-control studies, we consider a separate group of 7000 cases and 10000 controls
in each study. For overlapping case-control studies, we consider a distinct set of 7000 cases in each study, and a common set of 10000 controls shared across all the studies. For multiple case-control studies, we consider $K=5\left(K_{1}=0,1,2,3,4\right), K=10\left(K_{1}=0,2,4,6,8\right)$, and $K=15\left(K_{1}=0,3,6,9\right)$. We simulate the odds ratio for a non-null trait at random from $(1.05-1.25)$ in each replication under a simulation scenario. We consider the genetic effects to be all positive, and also both positive and negative. For each disease, we assume an overall disease prevalence of $10 \%$ in the whole population. We simulate the genotype data in cases and controls based on the standard logistic model of disease probability conditioning on the genotype: $\operatorname{logit}(P($ case $\mid G))=\alpha+\beta G$, where $G$ is the genotype at the risk SNP coded as the minor allele count $(0,1$, or 2$)$. In the cohort study, we consider 15000 individuals. First, we generate the continuous traits from a multivariate normal distribution based on the simulation model described in Majumdar et al. [2016]. We choose the trait-specific heritability due to the quantitative trait locus (QTL) at random from $(0.2 \%-0.5 \%)$. Finally, we dichotomize each continuous phenotype into a binary trait subject to an overall disease prevalence of $10 \%$. For the cohort studies, we considered $K=5\left(K_{1}=0,1,2,3\right), K=10\left(K_{1}=\right.$ $0,2,4,6)$, and $K=15\left(K_{1}=0,3,6,9\right)$ to save computing time and space.

Note that the Bayes factor and PPNA are not comparable to the p-value. While evaluating the aggregatelevel pleiotropic association, we provide various summary measures of $\log _{10}$ (Bayes factor) (abbreviated as $\left.\log _{10} \mathrm{BF}\right),-\log _{10}(\mathrm{PPNA})\left(\right.$ denoted as $\left.-\log _{10} \mathrm{PPNA}\right)$, and $-\log _{10}\left(\mathrm{ASSET}\right.$ p-value) (denoted by $\left.-\log _{10} \mathrm{ASTpv}\right)$ obtained across 500 replications. Under the global null hypothesis of no association $\left(K_{1}=0\right)$, we describe the summary measures in Table 1 (for non-overlapping case-control studies), 2 (for overlapping case-control studies), and S2 (for a cohort study). The summary measures in these three tables show that $\log _{10} \mathrm{BF}$ and $-\log _{10}$ PPNA are very well-controlled under the global null hypothesis. The maximum of $\log _{10} \mathrm{BF}$ in the three tables is observed to be negative $(\mathrm{BF}<1)$. Hence, all the quantiles of $\log _{10} \mathrm{BF}$ also appear to be negative. For example, in Table 1, for $K=5$ and $m=0.3,75 \%, 95 \%, 99 \%$ quantiles and the maximum of $\log _{10} B F$ are $-2.68,-2.35,-1.94$, and -1.46 , respectively. We also observe that $95 \%, 99 \%$ quantiles and the maximum of $-\log _{10} \mathrm{PPNA}$ are smaller than those for $-\log _{10} \mathrm{ASTpv}$ in all the three tables. For example, in Table 2 , for $K=10$ and $m=0.1,95 \%$ and $99 \%$ quantiles, and the maximum of $-\log _{10} \mathrm{PPNA}$ are $0.07,0.15$, and 1.01, respectively; whereas the same for $-\log _{10} \mathrm{ASTpv}$ are $0.98,1.91$, and 3.42 , respectively.

When at least one of the phenotypes is associated $\left(K_{1} \geq 1\right)$, we present the summary measures in Table $\mathrm{S} 3, \mathrm{~S} 4, \mathrm{~S} 5, \mathrm{~S} 6, \mathrm{~S} 7$ (for non-overlapping case-control studies); S8, S9, S10, S11, S12 (for overlapping casecontrol studies); and S13, S14, S15 (for cohort study). To save space, we provide all of these tables in the supplementary material. As expected, given a choice of $K, \log _{10} \mathrm{BF}$ and $-\log _{10} \mathrm{PPNA}$ increase as $K_{1}$ increases. For example, in Table S 3 (for 5 non-overlapping case-control studies), for $K_{1}=1\left(K_{1}^{+}=1, K_{1}^{-}=0\right)$ and $m=$ 0.3, the mean and median of $\log _{10} \mathrm{BF}$ are 11.05 and $5.12 ;$ whereas for $K_{1}=2\left(K_{1}^{+}=2, K_{1}^{-}=0\right)$ and $m=0.3$, the mean and median of $\log _{10} \mathrm{BF}$ are 31.70 and 27.99 . Similarly, in Table S 10 (for 10 overlapping case-control studies $)$, for $K_{1}=2\left(K_{1}^{+}=2, K_{1}^{-}=0\right)$ and $m=0.1$, the mean and median of $\log _{10}$ BF are 9.70 and 3.30 ;
whereas for $K_{1}=4\left(K_{1}^{+}=4, K_{1}^{-}=0\right)$ and $m=0.1$, the mean and median of $\log _{10} \mathrm{BF}$ are 25.97 and 14.82. For overlapping case-control studies and a cohort study, for the same choice of $K_{1}$, when the non-null effects are both positive and negative, $\log _{10} \mathrm{BF}$ and $-\log _{10} \mathrm{PPNA}$ tend to increase in comparison with when all the non-null effects are positive. For example, in Table S10 (for 10 overlapping case-control studies), for $K_{1}^{+}=4, K_{1}^{-}=0\left(K_{1}=4\right)$ and $m=0.3$, the mean and median of $\log _{10} \mathrm{BF}$ are 57.66 and 53.68 ; whereas for $K_{1}^{+}=2, K_{1}^{-}=2\left(K_{1}=4\right)$ and $m=0.3$, the mean and median of $\log _{10} \mathrm{BF}$ are 77.81 and 72.52 . Similarly, in Table S14 (for binary cohort with 10 phenotypes), for $K_{1}^{+}=2, K_{1}^{-}=0\left(K_{1}=2\right)$ and $m=0.3$, the mean and median of $\log _{10} \mathrm{BF}$ are 15.43 and 6.94 ; whereas for $K_{1}^{+}=1, K_{1}^{-}=1\left(K_{1}=2\right)$ and $m=0.3$, the mean and median of $\log _{10} \mathrm{BF}$ are 25.82 and 19.64. However, for non-overlapping case-control studies, we do not observe such a trend with respect to the direction of the non-null effects. We also observe that $-\log _{10} \mathrm{ASTpv}$ behaves similarly to $\log _{10} \mathrm{BF}$ and $-\log _{10}$ PPNA.

In Figure 2, we have plotted $\log _{10} \mathrm{BF}$ and $-\log _{10} \mathrm{PPNA}$ for 100 replications in the same set-up of five non-overlapping and overlapping case-control studies considered above. We chose MAF $=0.3$. As expected, when none of the phenotypes is associated, $\log _{10} \mathrm{BF}$ cluster below zero and $-\log _{10}$ PPNA cluster close to zero (Figure 2). And, the measures tend to increase as the number of associated traits increases (Figure 2).

For overlapping case-control studies and cohort study, we implemented the combined strategy of CPBayes. For each simulation scenario in Table $\mathrm{S} 8, \mathrm{~S} 9, \mathrm{~S} 10, \mathrm{~S} 11, \mathrm{~S} 12$ (for overlapping case-control studies), and S13, S14, S15 (for a cohort study), we provide the percentage of replications in which the combined strategy chose the uncorrelated version (denoted by uncor\%). From the tables, we observe that uncor\% increases as $K_{1}$ increases for a given choice of $K$. For example, in Table S11 (for 10 overlapping case control studies), for $m=0.3$, uncor \% increases from $3 \%$ to $6.4 \%$ as $K_{1}$ increases from $6\left(K_{1}^{+}=6, K_{1}^{-}=0\right)$ to $8\left(K_{1}^{+}=8, K_{1}^{-}=\right.$ 0 ). Similarly, in Table S15 (for a cohort study with 15 phenotypes), for $m=0.1$, uncor $\%$ increases from $5.8 \%$ to $8.4 \%$ as $K_{1}$ increases from $6\left(K_{1}^{+}=6, K_{1}^{-}=0\right)$ to $9\left(K_{1}^{+}=9, K_{1}^{-}=0\right)$. For a sparse scenario when less than half of the traits are associated (e.g., $K_{1}=3$ and $K=15$ ), uncor\% is substantially smaller than that for a non-sparse scenario (e.g., $K_{1}=9$ and $K=15$ ). For example, in Table S12 (for 15 overlapping case-control studies), for $m=0.1$, uncor $\%$ is $11 \%$ when $K_{1}=3\left(K_{1}^{+}=2, K_{1}^{-}=1\right)$, and it increases to $26.2 \%$ when $K_{1}=9\left(K_{1}^{+}=5, K_{1}^{-}=4\right)$. Similarly, in Table S15 (for a cohort study with 15 phenotypes), for $m=0.3$, uncor\% increases from $4.8 \%$ to $7.6 \%$ when $K_{1}$ increases from $3\left(K_{1}^{+}=2, K_{1}^{-}=1\right)$ to $9\left(K_{1}^{+}=5, K_{1}^{-}=4\right)$.

To evaluate the accuracy of selection of associated traits in a simulation scenario, we computed the average specificity and sensitivity obtained across 500 replications. We present the comparison of the selection accuracy between CPBayes and ASSET in Figure 3 (for non-overlapping case-control studies), Figure 4 (for overlapping case-control studies), and Figure S2 (for cohort study). CPBayes yielded a very good level of specificity (consistently more than $95 \%$ ) which is substantially higher than that of ASSET. For example, in Figure 3 (for non-overlapping case-control studies), while CPBayes' specificity is $>97 \%$, the specificity of ASSET varies in the range $46 \%-98 \%$ when $K=5$, and in a range $53 \%-99 \%$ when $K=10$. Similary, in

Figure 4 (for overlapping case-control studies), the specificity of ASSET varies in a range $49 \%-99 \%$ when $K=5$, and in $62 \%-100 \%$ when $K=10$.

CPBayes also offers an overall good level of sensitivity. For five non-overlapping case-control studies (Figure 3), CPBayes produced a sensitivity of $79 \%-89 \%$ when $\mathrm{MAF}=0.3$ and $64 \%-76 \%$ when $\mathrm{MAF}=$ 0.1; ASSET yielded slightly higher sensitivity of $86 \%-100 \%$ (with $50 \%-98 \%$ specificity) when MAF $=$ 0.3 , and a higher sensitivity of $85 \%-98 \%$ (with $46 \%-90 \%$ specificity) when MAF $=0.1$. In Figure 3 , we observe a similar pattern for $K=10$ and $K=15$. ASSET's higher sensitivity than CPBayes appears to come at the expense of a substantially lower specificity.

For overlapping case-control studies (Figure 4), CPBayes gave a substantially better sensitivity than ASSET (along with substantially better specificity) for a majority of the simulation scenarios. For example, when $K=10$ and $m=0.3$, the sensitivity was $82 \%-89 \%$ for CPBayes, and $39 \%-73 \%$ for ASSET, except for the case $K_{1}^{+}=1, K_{1}^{-}=1$ when ASSET had $100 \%$ sensitivity. Similarly, for $K=10$ and $m=0.1$, the sensitivity was $57 \%-73 \%$ for CPBayes and $39 \%-71 \%$ for ASSET except for the choice $K_{1}^{+}=1, K_{1}^{-}=1$ when ASSET had $96 \%$ sensitivity. For $K=15$ and $m=0.3$, the sensitivity was $82 \%-87 \%$ for CPBayes and $38 \%-82 \%$ for ASSET. For $K=15$ and $m=0.1$, the sensitivity was $60 \%-70 \%$ for CPBayes and $35 \%-79 \%$ for ASSET.

For a cohort study, CPBayes and ASSET both consistently exhibited good levels of sensitivity. For some cases, ASSET had a higher sensitivity than CPBayes, though generally with lower specificity. We also observe that when the non-null effects are all positive, the specificity of ASSET is smaller compared to when the non-null effects are both positive and negative. However, CPBayes performed more robustly with respect to the direction of non-null effects.

We also carried out a simulation study for 50 traits. We considered the same set-up of overlapping casecontrol studies considered above and $K_{1}=0,5,10$. Since ASSET is computationally very slow for 50 traits due to an extremely large number of possible subsets of traits, we only implemented CPBayes. Different summary measures of $\log _{10} \mathrm{BF}$ and $-\log _{10} \mathrm{PPNA}$ obtained across 200 replications are described in Table S16. The mean specificity and sensitivity are provided in Table S17. We observe that CPBayes performs similarly as for 5,10 , or 15 overlapping studies (discussed above).

## 5 GERA cohort analysis

To investigate the performance of CPBayes using real data, we analyzed the large genome-wide association study from the Kaiser "Resource for Genetic Epidemiology Research on Adult Health and Aging" (GERA) cohort data obtained from dbGaP [dbGaP Study Accession: phs000674.v1.p1]. We also analyzed the data by ASSET. For simplicity's sake, we restricted our analysis to 62,318 European-American individuals, who constitute more than $75 \%$ of the dbGaP data. We tested 657,184 SNPs genotyped across 22 autosomal
chromosomes for their potential pleiotropic effects on 22 phenotypes in the GERA cohort (Table S18). Note that in the dbGaP data, the cancers are collapsed into a single variable (any cancer). Therefore, we could only use an overall cancer categorization even though the genetic architecture is likely heterogeneous across different cancers. We provide the trait-trait correlation matrix in Table S24. The phenotypes are correlated modestly with a maximum correlation of 0.36 observed between Hypertension and Dyslipidemia.

Before our analysis, we undertook the following QC steps. First, we removed individuals with: over $3 \%$ of genotypes missing; any missing information on covariates (described below); genotype heterozygosity outside six standard deviations; first degree relatives; or discordant sex information. This left us with 53,809 individuals. Next, we removed SNPs with: MAF $<0.01 ; 10 \%$ or more missingness; or deviation from HWE at a level of significance $10^{-5}$. This leaves 601,175 SNPs that were tested for pleiotropic association by CPBayes and ASSET. We adjusted the analysis for the following covariates: age, gender, smoking status, BMI category, and 10 principal components of ancestry (PCs). We tested the single-trait association for each of the 22 phenotypes by a logistic regression of the case-control status on the genotype incorporating the same set of adjusting covariates. We used SNP-trait effect estimates (log odds ratios) and their standard errors in CPBayes and ASSET. Since the summary statistics are correlated, we used the combined strategy of CPBayes and the correlated version of ASSET.

As we have environmental covariates in the GERA study, we estimated the correlation matrix of the effect estimates using the genome-wide summary statistics data [Zhu et al., 2015]. First, we extracted all of the SNPs for each of which the trait-specific univariate p -values are $>0.1$ across 22 traits. This ensures that each SNP is either weakly or not associated with any of the 22 phenotypes. Then we selected a set of 24,510 independent SNPs from the initial set of null SNPs by using a threshold of $r^{2}<0.1$ ( $r$ : the correlation between the genotypes at a pair of SNPs). Finally, we computed the correlation matrix of the effect estimates as the sample correlation matrix of $\hat{\beta_{1}}, \ldots, \hat{\beta_{22}}$ across the selected 24,510 independent null SNPs.

We apply the conventional genome-wide (GW) level of statistical significance $5 \times 10^{-8}$ for ASSET (equivalent to $-\log _{10}($ ASSET p-value $\left.)>7.30\right)$. It is tough to decide on an appropriate GW cut-off for $\log _{10} \mathrm{BF}$ of CPBayes. Of note, Bhattacharjee et al. [2012] and Majumdar et al. [2016] demonstrated by simulations that ASSET appropriately controls for the false positive rate. We observed in our simulation study that, under the global null hypothesis of no association, CPBayes always produced a negative $\log _{10} \mathrm{BF}$ (Table 1, 2, S2). Hence $\log _{10} \mathrm{BF}$ was always smaller than $-\log _{10} \mathrm{ASTpv}$ under the null hypothesis. So, for contrasting the results obtained by the two methods, we set the cut-off of $\log _{10} \mathrm{BF}$ as 7.30 , the same as for $-\log _{10} \mathrm{ASTpv}$. This cut-off may be somewhat conservative for CPBayes, because in the simulation study, CPBayes produced substantially smaller $\log _{10} \mathrm{BF}$ than $-\log _{10} \mathrm{ASTpv}$ under the global null hypothesis (Table 1, 2, S2). Based on the GW cut-off, CPBayes identified 314 SNPs and ASSET detected 394 SNPs. By definition, each of these SNPs is associated with at least one of the 22 phenotypes. We note that CPBayes and ASSET identified a common set of 253 SNPs.

Many of the associated SNPs are expected to be in linkage disequilibrium (LD). On each chromosome, we identified the LD blocks by using a threshold of $r^{2}=0.25$. For CPBayes, we identified 49 associated LD blocks, and for ASSET, we detected 30 associated LD blocks. For each of the 394 SNPs detected by ASSET, the optimal subset of non-null traits always included more than one phenotype. So for ASSET, within each LD block, we chose the SNP having the minimum p-value of aggregate-level pleiotropic association. We present the results for these lead SNPs in Table S20, S21, S22, and S23. CPBayes selected more than one trait for 63 among 314 SNPs. Within each LD block identified by CPBayes, we chose the SNP associated with the maximum number of phenotypes. If multiple SNPs satisfy this criterion, we chose the one having the maximum $\log _{10} \mathrm{BF}$. If every SNP in a block is associated with one trait, we chose the SNP with the maximum $\log _{10} \mathrm{BF}$. In Table 3, we present the results for the independent pleiotropic SNPs at which CPBayes selected at least two phenotypes. In Table S19, we report the selected independent SNPs at which CPBayes detected one trait. In all the tables describing the results of CPBayes and ASSET, we provide the traitspecific univariate association p-values for contrast's sake. In all the tables for CPBayes, we also present the estimated trait-specific posterior probability of association $\left(\mathrm{PPA}_{j}\right)$ and the direction of association for the selected phenotypes (genotype was coded as the number of wild allele).

Even though CPBayes detected a smaller number of GW-significant SNPs (314) than ASSET (394), the former identified a substantially larger number of LD blocks than the later (49 versus 30). For example, CPBayes identified one SNP on chromosome 18 associated with Peripheral Vascular Disease, but ASSET did not detect any SNP on this chromosome. Specifically, CPBayes detected rs8092654 $\left(\log _{10} B F=12.24\right)$ at which ASSET yielded a p-value of 0.13 . In the NHGRI-EBI GWAS catalog, rs8092654 is reported as an eQTL hit for the ZNF611 gene in the peripheral blood monocytes tissue. As another example, on chromosome 11, CPBayes detected three LD blocks. From Table S19 for CPBayes, we observe that the lead SNPs of the blocks are rs1799963 (11p11.2), rs964184 (11q23.3), and rs55975204 (11q13.2) which are associated with Peripheral Vascular Disease (univariate p-value: $1.19 \times 10^{-8}$ ), Dyslipidemia (univariate pvalue: $5.49 \times 10^{-28}$ ), and Osteoporosis (univariate p-value: $2.0 \times 10^{-9}$ ), respectively. But, ASSET detected only one LD block on chromosome 11, which contains SNPs that are mainly associated with Dyslipidemia. In this block, the lead SNP for ASSET (rs964184) also turned out to be the lead SNP for CPBayes in the corresponding LD block containing the SNPs associated with Dyslipidemia. So, ASSET missed the signals for Peripheral Vascular Disease and Osteoporosis. We note that, in the NHGRI-EBI GWAS catalog, rs1799963 is reported to be associated with venous thromboembolism, and rs964184 is reported as associated with LDL cholesterol, triglycerides, and total cholesterol. And, rs55975204 is in LD ( $r^{2}=0.9$ ) with rs12286536 which is an eQTL hit for the CPT1A and MTL5 genes in the whole blood tissue. These findings suggest that in certain situations CPBayes may detect associations not detected by ASSET.

For each SNP detected by ASSET, the selected subset of non-null traits always contained more than one phenotype (Table S20, S21, S22, S23). For a majority of the SNPs, the subset of non-null traits
included some phenotypes that have large univariate association p-values (so, weak effects). For example, rs2300430 (1q31.3) was detected by both the methods (first SNP in Table S20 and S19); ASSET selected Allergic Rhinitis, Depressive Disorder, Dermatophytosis, Hemorrhoids, Insomnia, Macular Degeneration, Osteoporosis, and Peptic Ulcer, which have trait-specific univariate p-values equal to: $0.34,0.18,0.09,0.36$, $0.47,2.57 \times 10^{-77}, 0.09$, and 0.74 , respectively. In contrast, CPBayes only selected Macular Degeneration (Table S20). This suggests that CPBayes selects only those phenotypes having substantially strong genetic effects, while ASSET may select many more traits with lower specificity as seen in our simulation study.

CPBayes detected nine independent GW significant pleiotropic SNPs, for which it selected at least two phenotypes as non-null (Table 3). For example, at rs6025 (1q24.2), it selected a maximum of 5 phenotypes: Dermatophytosis, Hemorrhoids, Iron Deficiency, Osteoporosis, and Peripheral Vascular Disease, which have univariate $p$-values equal to $0.0018,0.0014,0.0004,0.0002$, and $6.81 \times 10^{-14}$, respectively. We present a forest plot for this pleiotropic signal in Figure 5. ASSET also identified this SNP and selected the same five traits as CPBayes. Interestingly, the SNP was positively associated with Dermatophytosis, Hemorrhoids, and Iron Deficiency, but negatively associated with Osteoporosis and Peripheral Vascular Disease (Figure 5). At rs10455872 (6q25.3), CPBayes selected Cardiovascular Disease, Dyslipidemia, and Peripheral Vascular Disease (Figure 6). ASSET selected these three traits and five more phenotypes with large univariate p-values. At rs651007 (9q34.2), CPBayes detected pleiotropy with Dyslipidemia and Peripheral Vascular Disease, whereas ASSET detected both of these and other phenotypes with weak genetic effects (Figure 7). CPBayes detected two independent pleiotropic SNPs in the chromosomal region 6 p 21.32 : rs 3830123 and rs9275476 (Table 3). The $r^{2}$ value between rs3830123 and rs9275476 was 0.016 . We also found that a disjoint set of phenotypes are associated with the two SNPs: Asthma, Type 2 Diabetes, and Macular Degeneration are associated with rs9275476; Cancers and Dyslipidemia are associated with rs3830123. CPBayes also detected four other pleiotropic loci: rs7601401 (2p16.1) (Abdominal Hernia and Osteoarthritis), rs17661572 (2p21) (Allergic Rhinitis and Insomnia), rs387608 (6p21.33) (Cancers and Macular Degeneration), and rs4506565 (10q25.2) (Type 2 Diabetes and Dyslipidemia). CPBayes detected 22 pair-wise trait-trait pleiotropic signals which we present in Table 4.

For CPBayes, the trait-specific posterior probability of association $\left(\mathrm{PPA}_{j}\right)$ provides a better insight into the relative strength of association between a pleiotropic variant and the selected non-null traits. For example, at rs6025, $\mathrm{PPA}_{j}$ for Dermatophytosis, Hemorrhoids, Iron Deficiency, Osteoporosis, and Peripheral Vascular Disease are $70 \%, 73.6 \%, 95.3 \%, 87.8 \%$, and $100 \%$, respectively (Table 3). This implies that the association with Peripheral Vascular Disease is the strongest among the five selected phenotypes. At some of the GW significant SNPs detected by CPBayes (Table 3 and S19), a few phenotypes produced a nonnegligible value of $\mathrm{PPA}_{j}$ but were left out from the optimal subset of non-null traits. In Table 5, we list these SNPs and the corresponding phenotypes having a $\mathrm{PPA}_{j}$ larger than $25 \%$. For example, at rs849135 (7p15.1), CPBayes only selected Type 2 Diabetes (Table 3), but Asthma also produced a $\mathrm{PPA}_{j}$ of $40.2 \%$.

Thus, even though the effect of rs849135 on Asthma was not strong enough to make into the optimal subset, a further consideration of the pleiotropic effect of rs849135 on Type 2 Diabetes and Asthma looks promising. At rs76075198 (19q13.31), CPBayes only selected Dyslipidemia (Table 3), but Peripheral Vascular Disease also produced a $\mathrm{PPA}_{j}$ of $44.9 \%$. We observe a similar phenomenon for the other two SNPs in Table 5 .

We note that a majority of the SNPs detected by the two methods (Table 3 , S19, S20, and S21) are already reported in the NHGRI-EBI GWAS catalog. For example, rs6025 (1q24.2) has been associated with inflammatory bowel disease and venous thromboembolism. rs10455872 (6q25.3) has been associated with myocardial infarction, response to statins (LDL cholesterol change), coronary artery disease, and aortic valve calcification. rs651007 (9q34.2) has been associated with iron status biomarkers (ferritin levels), blood metabolite levels, serum alkaline phosphatase levels, and end-stage coagulation. We also note that the combined strategy of CPBayes used the uncorrelated version only for 19 SNPs among all the 601,175 SNPs analyzed, and for none of the 314 SNPs identified in the primary genome-wide screening.

## 6 Discussion

We have proposed a Bayesian meta-analysis approach for pleiotropic association analysis based on summarylevel data. It simultaneously evaluates the evidence of aggregate-level pleiotropic association and estimates an optimal subset of traits associated with the risk locus under a unified Bayesian statistical framework. The method is implemented by Gibbs sampling designed for both uncorrelated and correlated summary statistics. We have conducted an extensive simulation study and analyzed the large GERA cohort for evaluating the performance of CPBayes.

An appealing feature of CPBayes is that, in addition to $\log _{10} \mathrm{BF}, \mathrm{PPNA}$, and an optimal subset of non-null traits, it simultaneously provides other interesting insights into an observed pleiotropic signal. For example, it estimates a trait-specific posterior probability of association $\left(\mathrm{PPA}_{j}\right)$, the direction of association, posterior mean/median, and the credible interval of the unknown true effect. As demonstrated in the real data application, even if CPBayes does not select a phenotype in the optimal subset of non-null traits which is defined as the MAP estimate (see the methods section), $\mathrm{PPA}_{j}$ for the phenotype may not be negligible. It may help an investigator to better explain a pleiotropic signal. One can also define the optimal subset of associated traits as $\left\{Y_{j}: \mathrm{PPA}_{j}>p\right\}$, where $p$ can be chosen as 0.5 (known as the median model), or other values. Based on our simulations, the MAP estimate and the median model lead to almost the same level of specificity and sensitivity. Such flexibility in making inference on pleiotropy are mainly due to the MCMC construction underlying CPBayes.

In contrast to ASSET, the major advantage of CPBayes is that it selects the non-null traits underlying a pleiotropic signal with a substantially higher accuracy. A possible reason behind this is that CPBayes performs the selection probabilistically through updating the latent association status by MCMC. ASSET
selects that subset of traits as non-null which maximizes the observed value of a weighted linear combination of the normalized univariate association statistics corresponding to the phenotypes belonging to a subset. So, given the summary statistics, ASSET does not select the non-null traits probabilistically based on the distribution of the summary statistics. We also note that ASSET is based on the framework of a fixed effects meta analysis and assumes that the effects in a given direction (positive/negative) are the same in size. But we observed in our real data application that, in a given direction, the effects of a variant across phenotypes may often be heterogeneous. CPBayes allows heterogeneity simultaneously in the direction and size of the effects. Table S1 summarizes key features of CPBayes and ASSET.

While assessing the selection accuracy, we have placed more emphasis on specificity than sensitivity. This was because a higher sensitivity at the expense of a lower specificity can lead to a false selection of too many traits as non-null. CPBayes consistently maintained a very good level of specificity while offering a good level of sensitivity across a wide range of simulation scenarios. While CPBayes produced a limited number of pleiotropic SNPs associated with more than one phenotype in the analysis of GERA cohort, these pleiotropic signals seem highly promising. Hence, the non-null traits for a pleiotropic variant selected by CPBayes are more reliable than those detected by ASSET.

Note that the continuous spike inherits the infinitesimal-model assumption that every SNP contributes to the variation of a trait, and the distinction is made between a negligible and a significant contribution, whereas the Dirac spike assigns the null effects explicitly to zero. From the perspective of heritability estimation, the infinitesimal-model assumption is more relevant since many SNPs with small effects underlie the variation of a phenotype. We conducted a simulation study (provided in the supplementary material) to compare the continuous spike and the Dirac spike. We found that the continuous spike offers better accuracy in the selection of non-null traits than the Dirac spike. The continuous spike is also computationally much faster (2-3 times) than the Dirac spike. Hence, we adopted the continuous spike for constructing CPBayes. We also note that, the latent association status $(Z)$ could only be used in the model for the continuous spike. For the Dirac spike, the inclusion of $Z$ in the model makes the corresponding MCMC reducible, and hence non-convergent to its stationary distribution (details not provided). Also, for the continuous spike, the full conditional posterior distributions of $z_{1}, \ldots, z_{K}$ are independent which leads to an explicit estimation of the Bayes factor based on the MCMC sample. But, for the Dirac spike, the explicit estimation of the Bayes factor appears to be extremely difficult in the correlated case, because the full conditional posterior distributions of $\beta_{1}, \ldots, \beta_{K}$ are not independent for correlated summary statistics.

In related work, Han and Eskin [2012] proposed a modified random effects meta analysis for combining heterogeneous studies coupled with a Bayesian approach to provide a better interpretation of an observed signal of aggregate-level association. They investigated how to combine heterogeneous genetic studies across different populations/ethnicities. However, they did not address how to account for a possible correlation between the summary statistics while selecting the most important studies underlying an observed signal of
aggregate-level association. Hence we compared CPBayes only with ASSET.
We note that we did not explicitly compare CPBayes and ASSET with respect to power, because the Bayes factor and the p-value are not directly comparable. Computing a p-value based on the Bayes factor is computationally very expensive, and moreover, the approximation may not be accurate at a genome-wide scale. On the other hand, determining the cut-off of the Bayes factor corresponding to a given choice of the false positive rate is also time consuming and may become computationally infeasible for a false positive rate at the genome-wide scale, e.g., $5 \times 10^{-8}$. Hence, we preserved the fully Bayesian essence of CPBayes. In the GERA cohort analysis, CPBayes primarily detected smaller number of SNPs than ASSET (314 versus 394) by using a seemingly conservative cut-off of $\log _{10} \mathrm{BF}$ (discussed in the real data application section). But, among the primarily identified genome-wide significant SNPs, CPBayes detected substantially larger number of associated LD blocks than ASSET (49 versus 30). ASSET completely missed some loci which were detected by CPBayes. These findings indicate that CPBayes is powerful to identify pleiotropic variants.

We have evaluated CPBayes primarily for SNP by SNP analysis. However, one possible approach to implement the method for a gene-based association analysis is as follows. First, implement PrediXcan [Gamazon et al., 2015] to impute the expression level of a gene. Next, regress each phenotype individually on the imputed gene expression level and compute the estimate of the association parameter $(\beta)$ along with the corresponding standard error. Finally, we can implement CPBayes on these summary statistics to conduct a gene-level pleiotropy analysis. We note that the method can also be applied to observational studies of non-genetic exposures.

For a larger number of phenotypes, CPBayes is computationally faster than ASSET. For example, in the analysis of 22 traits in the GERA cohort, CPBayes took an average run time of 4.5 hours for 1,000 SNPs, and ASSET took an average run time of 9 hours for 1,000 SNPs. However, as the number of traits decreases, ASSET gradually becomes faster due to the reduction in the number of all possible subsets of traits. That said, CPBayes is computationally feasible and can be implemented at a genome-wide scale. As expected, the uncorrelated version of CPBayes is at least twice as fast as the correlated version of CPBayes. In future work, we aim to investigate whether the computing speed of CPBayes can be increased by using a variational Bayes approach or by using an optimization technique (e.g., EM algorithm or its variants) instead of using MCMC, while preserving the efficiency of the method. We also plan to explore whether fitting multiple (3 or 4) slab distributions instead of a single slab can better model the non-null effects in the presence of extreme effect size heterogeneity.

In summary, CPBayes is a sensitive and specific approach for detecting associated traits underlying a pleiotropic signal. CPBayes has a strong theoretical foundation and allows for heterogeneity in both the direction and size of effects. In addition to parameters of primary interest (e.g., the measures of overall pleiotropic association, the optimal subset of associated traits), it provides other interesting insights into a pleiotropic signal (e.g., the trait-specific posterior probability of association, the direction of association). It
bioRxiv preprint doi: https://doi.org/10.1101/101543; this version posted January 19, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Bayesian meta-analysis for studying cross-phenotype genetic associations
is computationally feasible and faster than ASSET for a larger number of traits. A user-friendly R-package 'CPBayes' is provided for general use by other investigators.

## Appendix A

Here we state the Gibbs sampling algorithm for the continuous spike described in Equation 1. It is a desirable practice to provide a MCMC with a good initial value of the model parameters for faster convergence to its stationary distribution. Hence we use the false discovery rate controlling procedure proposed by Benjamini and Yekutieli [2001] (BY procedure) which is robust to arbitrary correlation structure of multiple test statistics. We apply the BY procedure on the univariate association p-values of $K$ traits at an FDR level of 0.05 and assign $z_{j}=1$ if $Y_{j}$ is found to be significantly associated, otherwise set $z_{j}=0 ; j=1, \ldots, K$. We also choose an initial value of $q$ as the proportion of non-null traits detected by the BY procedure (the boundary situations of no/all non-null traits are taken care of appropriately).

Define $\Sigma_{2}=\operatorname{diag}\left(\tau_{1}^{2}, \ldots, \tau_{K}^{2}\right)$ (a diagonal matrix with diagonal elements $\tau_{1}^{2}, \ldots, \tau_{K}^{2}$ ), where $\tau_{j}=\tau$ if $z_{j}=0 ;$ and $\tau_{j}=\frac{\tau}{d}$ if $z_{j}=1 ; j=1, \ldots, K$. So $\boldsymbol{\beta} \mid Z \sim \operatorname{MVN}\left(\underset{\sim}{0}, \Sigma_{2}\right)$. Let $\Sigma_{1}=S$. Also, let $\boldsymbol{\beta}_{-j}=$ $\left(\beta_{1}, \ldots, \beta_{j-1}, \beta_{j+1}, \ldots, \beta_{K}\right)$, and $Z_{-j}=\left(z_{1}, \ldots, z_{j-1}, z_{j+1}, \ldots, z_{K}\right)$.

```
Algorithm 1 Gibbs sampling for continuous spike in correlated case
    Start:
    Assign the initial values of \(Z\) and \(q\) as described above.
    loop:
    Simulate \(\boldsymbol{\beta}\) from its full conditional posterior distribution: \(\boldsymbol{\beta} \mid Z, q, d, \hat{\boldsymbol{\beta}} \sim \operatorname{MVN}\left[\left(\Sigma_{1}^{-1}+\right.\right.\)
    \(\left.\left.\Sigma_{2}^{-1}\right)^{-1} \Sigma_{1}^{-1} \hat{\boldsymbol{\beta}},\left(\Sigma_{1}^{-1}+\Sigma_{2}^{-1}\right)^{-1}\right]\).
    For \(j=1, \ldots, K\), update \(z_{j}\) using the full conditional posterior probability: \(P\left(z_{j}=0 \mid Z_{-j}, \boldsymbol{\beta}, q, d, \hat{\boldsymbol{\beta}}\right)=\)
    \(\frac{1}{1+\text { ratio }_{j}}\), where ratio \({ }_{j}=\frac{q}{1-q} d \exp \left[-\frac{\beta_{j}^{2}}{2 \tau^{2}}\left(d^{2}-1\right)\right]\).
    Let \(k_{1}=\sum_{j=1}^{K} z_{j}, k_{0}=K-k_{1}\). Update \(q\) using \(q \mid \boldsymbol{\beta}, Z, d, \hat{\boldsymbol{\beta}} \sim \operatorname{Beta}\left(c_{1}+k_{1}, c_{2}+k_{0}\right)\).
    We assume that \(e_{1}=e_{2}=1\). Update \(d\) from its full conditional posterior distribution in a fixed range
    so that the slab variance \(\left(\frac{\tau}{d}\right)^{2}\) varies in a given range \(\left(v_{0}, v_{1}\right)\), and let the corresponding range of \(d\) be
    given by: \(d_{0}<d<d_{1}\). If \(k_{1}=\sum_{j=1}^{K} z_{j}>0\), then \(d=\sqrt{\frac{y}{2 C}}\), where \(C=\frac{1}{2 \tau^{2}} \sum_{j: z_{j}=1} \beta_{j}^{2}\), and \(y\) follows
    a truncated \(\left(2 C d_{0}^{2}<y<2 C d_{1}^{2}\right)\) central \(\chi_{k_{1}+1}^{2}\) distribution. If \(k_{1}=0, d\) is updated from the truncated
    \(\left(d_{0}<d<d_{1}\right) \operatorname{Beta}(1,1)\) distribution.
    goto loop until all the MCMC iterations are finished.
```

We note that, $d$ can be updated using the truncated central $\chi^{2}$ distribution as long as the second shape parameter of its Beta prior $\left(e_{2}\right)$ is 1 .

If the summary statistics are uncorrelated, step 4 of Algorithm 1 is modified as: for $j=1, \ldots, K$, update $\beta_{j}$ by sampling from its full conditional posterior distribution: $\beta_{j} \mid \boldsymbol{\beta}_{-j}, Z, q, d, \hat{\boldsymbol{\beta}} \sim \mathrm{~N}\left(\frac{\sigma_{j}^{2}}{s_{j}^{2}} \hat{\beta}_{j}, \sigma_{j}^{2}\right)$, where $\frac{1}{\sigma_{j}^{2}}=\frac{1}{s_{j}^{2}}+\frac{1}{\tau_{j}^{2}}$. All the other steps remain the same as in the Algorithm 1.

## Acknowledgements

This work was supported by the National Institutes of Health grants R01CA088164, U01CA127298, R25CA112355, and the UCSF Goldberg-Benioff Program in Cancer Translational Biology. We thank Thomas Hoffmann, Prasenjit Ghosh, and Moumita Das for important discussions relating to this work. The GERA cohort data came from a grant, the Resource for Genetic Epidemiology Research in Adult Health and Aging (RC2 AG033067; Schaefer and Risch, PIs) awarded to the GERA Permanente Research Program on Genes, Environment, and Health (RPGEH) and the UCSF Institute for Human Genetics. The RPGEH was supported by grants from the Robert Wood Johnson Foundation, the Wayne and Gladys Valley Foundation, the Ellison Medical Foundation, GERA Permanente Northern California, and the GERA Permanente National and Northern California Community Benefit Programs. The RPGEH and the Resource for Genetic Epidemiology Research in Adult Health and Aging are described in the following publication, Schaefer C. et al., The GERA Permanente Research Program on Genes, Environment and Health: Development of a Research Resource in a Multi-Ethnic Health Plan with Electronic Medical Records, In preparation. The authors do not have any conflict of interest

## References

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. Annals of statistics, pages 1165-1188.

Bhattacharjee, S., Rajaraman, P., Jacobs, K. B., Wheeler, W. A., Melin, B. S., Hartge, P., Yeager, M., Chung, C. C., Chanock, S. J., Chatterjee, N., et al. (2012). A subset-based approach improves power and interpretation for the combined analysis of genetic association studies of heterogeneous traits. The American Journal of Human Genetics, 90(5):821-835.

Carty, C. L., Bhattacharjee, S., Haessler, J., Cheng, I., Hindorff, L. A., Aroda, V., Carlson, C. S., Hsu, C.-N., Wilkens, L., Liu, S., et al. (2014). Comparative analysis of metabolic syndrome components in over 15,000 african americans identifies pleiotropic variants: Results from the page study. Circulation: Cardiovascular Genetics, pages 505-513.

Ellinghaus, D., Jostins, L., Spain, S. L., Cortes, A., Bethune, J., Han, B., Park, Y. R., Raychaudhuri, S., Pouget, J. G., Hübenthal, M., et al. (2016). Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. Nature genetics.

Gamazon, E. R., Wheeler, H. E., Shah, K. P., Mozaffari, S. V., Aquino-Michaels, K., Carroll, R. J., Eyler, A. E., Denny, J. C., Nicolae, D. L., Cox, N. J., et al. (2015). A gene-based association method for mapping traits using reference transcriptome data. Nature genetics, 47(9):1091-1098.

George, E. I. and McCulloch, R. E. (1993). Variable selection via gibbs sampling. Journal of the American Statistical Association, 88(423):881-889.

Han, B. and Eskin, E. (2012). Interpreting meta-analyses of genome-wide association studies. PLoS Genet, 8(3):e1002555.

Lin, D.-Y. and Sullivan, P. F. (2009). Meta-analysis of genome-wide association studies with overlapping subjects. The American Journal of Human Genetics, 85(6):862-872.

Majumdar, A., Haldar, T., and Witte, J. S. (2016). Determining which phenotypes underlie a pleiotropic signal. Genetic epidemiology.

Malsiner-Walli, G. and Wagner, H. (2011). Comparing spike and slab priors for bayesian variable selection. Austrian Journal of Statistics, 40(4):241-264.

Mitchell, T. J. and Beauchamp, J. J. (1988). Bayesian variable selection in linear regression. Journal of the American Statistical Association, 83(404):1023-1032.

Parkes, M., Cortes, A., van Heel, D. A., and Brown, M. A. (2013). Genetic insights into common pathways and complex relationships among immune-mediated diseases. Nature Reviews Genetics, 14(9):661-673.

Pickrell, J. K., Berisa, T., Liu, J. Z., Ségurel, L., Tung, J. Y., and Hinds, D. A. (2016). Detection and interpretation of shared genetic influences on 42 human traits. Nature genetics.

Sakoda, L. C., Jorgenson, E., and Witte, J. S. (2013). Turning of cogs moves forward findings for hormonally mediated cancers. Nat Genet, 45(4):345-8.

Stephens, M. and Balding, D. J. (2009). Bayesian statistical methods for genetic association studies. Nature Reviews Genetics, 10(10):681-690.

Vilhjálmsson, B. J., Yang, J., Finucane, H. K., Gusev, A., Lindström, S., Ripke, S., Genovese, G., Loh, P.-R., Bhatia, G., Do, R., et al. (2015). Modeling linkage disequilibrium increases accuracy of polygenic risk scores. The American Journal of Human Genetics, 97(4):576-592.

Wen, X. and Stephens, M. (2014). Bayesian methods for genetic association analysis with heterogeneous subgroups: from meta-analyses to gene-environment interactions. The annals of applied statistics, 8(1):176.

Zaykin, D. V. and Kozbur, D. O. (2010). P-value based analysis for shared controls design in genome-wide association studies. Genetic epidemiology, 34(7):725-738.

Zhou, X., Carbonetto, P., and Stephens, M. (2013). Polygenic modeling with bayesian sparse linear mixed models. PLoS Genet, 9(2):e1003264.
bioRxiv preprint doi: https://doi.org/10.1101/101543; this version posted January 19, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Bayesian meta-analysis for studying cross-phenotype genetic associations

Zhu, X., Feng, T., Tayo, B. O., Liang, J., Young, J. H., Franceschini, N., Smith, J. A., Yanek, L. R., Sun, Y. V., Edwards, T. L., et al. (2015). Meta-analysis of correlated traits via summary statistics from gwass with an application in hypertension. The American Journal of Human Genetics, 96(1):21-36.

Global Lipids Genetics Consortium (2013). Discovery and refinement of loci associated with lipid levels. Nature genetics, 45(11):1274-1283.

Figure 1: An example diagram of the continuous spike and slab prior used by CPBayes to model pleiotropy. In this diagram, the spike variance is chosen as 0.1 . However, we set this value to $10^{-4}$ in our simulation study and real data analysis (a diagram corresponding to this choice is presented in Figure S1 in the supplementary material.)

Continuous spike and slab prior modeling pleiotropy


Table 1: Simulation study results. Summary of measures of the evidence of the overall pleiotropic association under the global null hypothesis of no association when multiple case-control studies are considered without any overlapping subjects. We considered a separate group of 7000 cases and 10000 controls in each study.

| K | m |  | mean |  |  | quantiles |  |  |  |  |  | max |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  | sd | min | 5\% | 25\% | 50\% | 75\% | 95\% | 99\% |  |
| 5 | 0.3 | $\log _{10} \mathrm{BF}$ | -2.76 | 0.22 | -3.02 | -2.99 | -2.92 | -2.82 | -2.68 | -2.35 | -1.94 | -1.46 |
|  | 0.1 | - $\log _{10}$ PPNA | 0.03 | 0.02 | 0.01 | 0.01 | 0.02 | 0.02 | 0.03 | 0.06 | 0.13 | 0.32 |
|  |  | - $\log _{10}$ ASTpv | 0.48 | 0.47 | 0.00 | 0.01 | 0.13 | 0.33 | 0.68 | 1.45 | 1.97 | 2.88 |
|  |  | $\log _{10} \mathrm{BF}$ | -2.57 | 0.28 | -2.86 | -2.81 | -2.74 | -2.65 | -2.49 | -2.12 | -1.39 | -0.26 |
|  |  | - $\log _{10}$ PPNA | 0.05 | 0.08 | 0.02 | 0.02 | 0.02 | 0.03 | 0.04 | 0.09 | 0.35 | 1.25 |
|  |  | - $\log _{10}$ ASTpv | 0.43 | 0.44 | 0.00 | 0.005 | 0.11 | 0.30 | 0.60 | 1.24 | 2.14 | 2.63 |
| 10 | 0.3 | $\log _{10} \mathrm{BF}$ | -4.29 | 0.22 | -4.59 | -4.51 | -4.42 | -4.35 | -4.22 | -3.92 | -3.54 | -2.41 |
|  |  | - $\log _{10}$ PPNA | 0.03 | 0.04 | 0.01 | 0.01 | 0.02 | 0.02 | 0.03 | 0.05 | 0.11 | 0.70 |
|  |  | $-\log _{10}$ ASTpv | 0.32 | 0.42 | 0.00 | 0.00 | 0.04 | 0.17 | 0.44 | 1.14 | 1.82 | 2.99 |
|  | 0.1 | $\log _{10} \mathrm{BF}$ | -4.09 | 0.23 | -4.42 | -4.34 | -4.24 | -4.15 | -4.01 | -3.65 | -3.18 | -2.55 |
|  |  | $-\log _{10}$ PPNA | 0.04 | 0.04 | 0.02 | 0.02 | 0.02 | 0.03 | 0.04 | 0.09 | 0.22 | 0.59 |
|  |  | $-\log _{10}$ ASTpv | 0.30 | 0.38 | 0.00 | 0.00 | 0.04 | 0.17 | 0.44 | 1.05 | 1.66 | 2.53 |
| 15 | 0.3 | $\log _{10} \mathrm{BF}$ | -5.76 | 0.20 | -6.04 | -5.96 | -5.88 | -5.81 | -5.70 | -5.36 | -4.99 | -4.20 |
|  |  | $-\log _{10}$ PPNA | 0.03 | 0.03 | 0.01 | 0.02 | 0.02 | 0.02 | 0.03 | 0.06 | 0.12 | 0.48 |
|  |  | - $\log _{10}$ ASTpv | 0.24 | 0.38 | 0.00 | 0.00 | 0.001 | 0.08 | 0.34 | 0.89 | 1.42 | 4.37 |
|  | 0.1 | $\log _{10} \mathrm{BF}$ | -5.55 | 0.23 | -5.85 | -5.78 | -5.69 | -5.61 | -5.49 | -5.13 | -4.61 | -4.16 |
|  |  | - $\log _{10}$ PPNA | 0.05 | 0.04 | 0.02 | 0.02 | 0.03 | 0.03 | 0.04 | 0.09 | 0.25 | 0.51 |
|  |  | $-\log _{10}$ ASTpv | 0.24 | 0.41 | 0.00 | 0.00 | 0.00 | 0.08 | 0.32 | 1.05 | 1.78 | 4.40 |

Legend: $K$ - total number of phenotypes, $m$ - allele frequency at the marker $\mathrm{SNP} ; K=5,10$, 15 , and $m=0.3$, 0.1 . The following abbreviations denote $-\log _{10} \mathrm{BF}: \log _{10}$ (Bayes factor), $-\log _{10} \mathrm{PPNA}:-\log _{10}$ (posterior probability of null association), $-\log _{10}$ ASTpv: $-\log _{10}$ (ASSET p-value). For multiple studies with no overlapping subjects, the uncorrelated version of CPBayes is implemented. Different summary measures obtained across 500 replications are provided: mean, standard deviation (sd), minimum (min), $5 \%$ quantile, $25 \%$ quantile, $50 \%$ quantile, $75 \%$ quantile, $95 \%$ quantile, $99 \%$ quantile, and maximum (max).

Table 2: Summary of measures for the evidence of the overall pleiotropic association under the global null hypothesis of no association when multiple case-control studies with overlapping subjects are considered. We considered a distinct set of 7000 cases in each study and a common set of 10000 controls shared across all the studies.

| K | m |  | quantile |  |  |  |  |  |  |  |  | max |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | mean | sd | min | 5\% | 25\% | 50\% | 75\% | 95\% | 99\% |  |
| 5 | 0.3 | $\log _{10} \mathrm{BF}$ | -2.88 | 0.18 | -3.10 | -3.06 | -3.00 | -2.93 | -2.81 | -2.49 | -2.12 | -2.04 |
|  |  | $-\log _{10}$ PPNA | 0.02 | 0.01 | 0.01 | 0.01 | 0.01 | 0.02 | 0.02 | 0.04 | 0.09 | 0.11 |
|  |  | - $\log _{10}$ ASTpv | 0.26 | 0.38 | 0.00 | $6.18 \mathrm{E}-05$ | 0.02 | 0.10 | 0.34 | 1.03 | 1.75 | 2.34 |
|  | 0.1 | $\log _{10} \mathrm{BF}$ | -2.67 | 0.24 | -2.94 | -2.89 | -2.83 | -2.75 | -2.58 | -2.18 | -1.79 | -1.22 |
|  |  | $-\log _{10}$ PPNA | 0.03 | 0.04 | 0.02 | 0.02 | 0.02 | 0.02 | 0.03 | 0.08 | 0.18 | 0.46 |
|  |  | $-\log _{10}$ ASTpv | 0.27 | 0.41 | 0.00 | $2.35 \mathrm{E}-05$ | 0.02 | 0.10 | 0.32 | 1.17 | 1.72 | 3.43 |
| 10 | 0.3 | $\log _{10} \mathrm{BF}$ | -4.40 | 0.18 | -4.64 | -4.59 | -4.52 | -4.45 | -4.34 | -4.04 | -3.76 | -3.41 |
|  |  | $-\log _{10}$ PPNA | 0.02 | 0.01 | 0.01 | 0.01 | 0.01 | 0.02 | 0.02 | 0.04 | 0.07 | 0.15 |
|  |  | $-\log _{10}$ ASTpv | 0.15 | 0.36 | 0.00 | 0.00 | 0.00 | 0.0003 | 0.04 | 0.98 | 1.62 | 2.19 |
|  | 0.1 | $\log _{10} \mathrm{BF}$ | -4.20 | 0.23 | -4.47 | -4.43 | -4.35 | -4.26 | -4.12 | -3.78 | -3.39 | -2.05 |
|  |  | $-\log _{10}$ PPNA | 0.03 | 0.05 | 0.01 | 0.02 | 0.02 | 0.02 | 0.03 | 0.07 | 0.15 | 1.01 |
|  |  | $-\log _{10}$ ASTpv | 0.15 | 0.39 | 0.00 | 0.00 | 0.00 | $6.71 \mathrm{E}-06$ | 0.04 | 0.98 | 1.91 | 3.42 |
| 15 | 0.3 | $\log _{10} \mathrm{BF}$ | -5.90 | 0.15 | -6.12 | -6.07 | -6.01 | -5.93 | -5.84 | -5.60 | -5.38 | -5.19 |
|  |  | $-\log _{10}$ PPNA | 0.02 | 0.01 | 0.01 | 0.01 | 0.01 | 0.02 | 0.02 | 0.03 | 0.06 | 0.08 |
|  |  | - $\log _{10}$ ASTpv | 0.10 | 0.33 | 0.00 | 0.00 | 0.00 | 0.00 | 0.002 | 0.81 | 1.62 | 2.27 |
|  | 0.1 | $\log _{10} \mathrm{BF}$ | -5.68 | 0.21 | -5.96 | -5.89 | -5.81 | -5.73 | -5.61 | -5.32 | -4.82 | -4.15 |
|  |  | $-\log _{10}$ PPNA | 0.03 | 0.03 | 0.02 | 0.02 | 0.02 | 0.03 | 0.03 | 0.06 | 0.17 | 0.52 |
|  |  | - $\log _{10}$ ASTpv | 0.13 | 0.42 | 0.00 | 0.00 | 0.00 | 0.00 | 0.003 | 1.03 | 2.22 | 3.37 |

Legend: $K$ - total number of phenotypes, $m$ - allele frequency at the marker SNP; $K=5,10$, 15 , and $m=0.3$, 0.1 . The following abbreviations denote $-\log _{10} \mathrm{BF}: \log _{10}$ (Bayes factor), $-\log _{10} \mathrm{PPNA}:-\log _{10}$ (posterior probability of null association), $\log _{10}$ ASTpv: $-\log _{10}$ (ASSET p-value). For multiple studies with overlapping subjects, since the summary statistics are correlated, the combined strategy of CPBayes is implemented. Different summary measures obtained across 500 replications are provided: mean, standard deviation (sd), minimum (min), $5 \%$ quantile, $25 \%$ quantile, $50 \%$ quantile, $75 \%$ quantile, $95 \%$ quantile, $99 \%$ quantile, and maximum (max).

Bayesian meta-analysis for studying cross-phenotype genetic associations

Figure 2: Simulation study results. Bayes factor and PPNA for five non-overlapping and overlapping casecontrol studies across 100 replications. The red colored band presents (mean - standard deviation, mean + standard deviation) of the Bayes factor or PPNA across the replications.


Figure 3: Simulation study results. Comparison of the accuracy of selection of associated traits by CPBayes and ASSET for multiple non-overlapping case-control studies. The total number of phenotypes/studies is denoted by $K$ and $m$ denotes the minor allele frequency at the risk SNP.


Figure 4: Comparison of the accuracy of selection of associated traits by CPBayes and ASSET for multiple overlapping case-control studies. The total number of phenotypes/studies is denoted by $K$ and $m$ denotes the minor allele frequency at the risk SNP.


Table 3: Independent pleiotropic SNPs detected by CPBayes which are associated with at least two phenotypes.

| rsID | chrom band | CPBayes $\log _{10} \mathrm{BF}$ | CPBayes PPNA | Subset of associated traits | $\mathrm{PPA}_{j}$ | Direction | Univariate p-values |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| rs6025 | 1q24.2 | 270.68 | $4.97 \mathrm{E}-278$ | Dermatophytosis | 70\% | positive | 0.0018 |
|  |  |  |  | Hemorrhoids | 73.6\% | positive | 0.0014 |
|  |  |  |  | Iron Deficiency | 95.3\% | positive | 0.0004 |
|  |  |  |  | Osteoporosis | 87.8\% | negative | 0.0002 |
|  |  |  |  | Peripheral Vascular Disease | 100\% | negative | $6.81 \mathrm{E}-14$ |
| rs7601401 | 2p16.1 | 9.69 | $4.91 \mathrm{E}-17$ | Abdominal Hernia | 100\% | positive | $3.88 \mathrm{E}-12$ |
|  |  |  |  | Osteoarthritis | 86.7\% | positive | $3.46 \mathrm{E}-06$ |
| rs17661572 | 2p21 | 8.47 | $8.10 \mathrm{E}-16$ | Allergic Rhinitis | 94\% | negative | $4.32 \mathrm{E}-05$ |
|  |  |  |  | Insomnia | 98.4\% | positive | 0.0001 |
| rs387608 | 6p21.33 | 29.06 | $2.07 \mathrm{E}-36$ | Cancers | 54.7\% | positive | 0.0001 |
|  |  |  |  | Macular Degeneration | 100\% | positive | $4.29 \mathrm{E}-12$ |
| rs10455872 | 6 q 25.3 | 26.20 | $1.49 \mathrm{E}-33$ | Cardiovascular Disease | 72.9\% | negative | $6.14 \mathrm{E}-05$ |
|  |  |  |  | Dyslipidemia | 100\% | negative | $6.97 \mathrm{E}-15$ |
|  |  |  |  | Peripheral Vascular Disease | 66.8\% | negative | 0.0002 |
| rs3830123 | 6p21.32 | 22.31 | $1.16 \mathrm{E}-29$ | Cancers | 99.9\% | negative | $1.01 \mathrm{E}-07$ |
|  |  |  |  | Dyslipidemia | 100\% | negative | $5.55 \mathrm{E}-13$ |
| rs9275476 | 6p21.32 | 22.10 | $1.88 \mathrm{E}-29$ | Asthma | 100\% | negative | $5.61 \mathrm{E}-08$ |
|  |  |  |  | Type 2 Diabetes | 100\% | negative | $4.27 \mathrm{E}-06$ |
|  |  |  |  | Macular Degeneration | 56\% | positive | 0.0011 |
| rs651007 | 9 q 34.2 | 24.72 | $4.57 \mathrm{E}-32$ | Dyslipidemia | 100\% | negative | $2.89 \mathrm{E}-15$ |
|  |  |  |  | Peripheral Vascular Disease | 100\% | negative | $9.55 \mathrm{E}-08$ |
| rs4506565 | $10 \mathrm{q} 25.2$ | $117.08$ | $1.96 \mathrm{E}-124$ | Type 2 Diabetes | 100\% | negative | $2.02 \mathrm{E}-55$ |
|  |  |  |  | Dyslipidemia | 60.6\% | negative | $1.33 \mathrm{E}-06$ |

Legend: The chromosome band of a SNP is denoted by 'chrom band'. Direction means whether the SNP is positively or negatively associated with the phenotype.
Table 4: Pairwise trait-trait pleiotropic signals detected by CPBaye

| Abdominal Hernia |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  | $\begin{aligned} & \text { rs7601401 } \\ & (2 \mathrm{p} 16.1) \\ & \hline \end{aligned}$ |  |  |  |
| Allergic Rhinitis |  |  | $\begin{aligned} & \text { rs17661572 } \\ & (2 \mathrm{p} 21) \end{aligned}$ |  |  |  |  |  |  |
| Asthma |  |  |  |  | $\begin{aligned} & \text { rs9275476 } \\ & (6 \mathrm{p} 21.32) \end{aligned}$ |  |  |  | $\begin{aligned} & \hline \text { rs9275476 } \\ & (6 \mathrm{p} 21.32) \end{aligned}$ |
| Cancers | $\begin{aligned} & \hline \text { rs3830123 } \\ & (6 \mathrm{p} 21.32) \\ & \hline \end{aligned}$ |  |  |  | $\begin{aligned} & \mathrm{rs} 387608 \\ & (6 \mathrm{p} 21.33) \\ & \hline \end{aligned}$ |  |  |  |  |
| Cardiovascular Disease | $\begin{aligned} & \text { rs10455872 } \\ & (6 \mathrm{q} 25.3) \\ & \hline \end{aligned}$ |  |  |  |  |  |  | $\begin{aligned} & \hline \text { rs10455872 } \\ & (6 \mathrm{q} 25.3) \\ & \hline \end{aligned}$ |  |
| Dermatophytosis |  | $\begin{aligned} & \hline \text { rs6025 } \\ & (1 \mathrm{q} 24.2) \\ & \hline \end{aligned}$ |  | $\begin{aligned} & \hline \text { rs6025 } \\ & (1 \mathrm{q} 24.2) \\ & \hline \end{aligned}$ |  |  | $\begin{aligned} & \hline \text { rs6025 } \\ & (1 \mathrm{q} 24.2) \end{aligned}$ | $\begin{aligned} & \text { rs6025 } \\ & (1 \mathrm{q} 24.2) \\ & \hline \end{aligned}$ |  |
| Dyslipidemia |  |  |  |  |  |  |  | $\begin{aligned} & \text { rs10455872 } \\ & (6 \mathrm{q} 25.3) \\ & \text { rs651007 } \\ & (9 \mathrm{q} 34.2) \\ & \hline \end{aligned}$ | $\begin{aligned} & \text { rs4506565 } \\ & (10 q 25.2) \end{aligned}$ |
| Hemorrhoids |  |  |  | $\begin{aligned} & \hline \text { rs6025 } \\ & (1 \mathrm{q} 24.2) \\ & \hline \end{aligned}$ |  |  | $\begin{aligned} & \hline \text { rs6025 } \\ & (1 \mathrm{q} 24.2) \\ & \hline \end{aligned}$ | $\begin{aligned} & \hline \text { rs6025 } \\ & (1 \mathrm{q} 24.2) \\ & \hline \end{aligned}$ |  |
| Iron Deficiency |  |  |  |  |  |  | $\begin{aligned} & \hline \text { rs6025 } \\ & (1 \mathrm{q} 24.2) \end{aligned}$ | $\begin{aligned} & \hline \text { rs6025 } \\ & (1 \mathrm{q} 24.2) \\ & \hline \end{aligned}$ |  |
| Macular Degeneration |  |  |  |  |  |  |  |  | $\begin{aligned} & \text { rs9275476 } \\ & \text { (6p21.32) } \\ & \hline \end{aligned}$ |
| Osteoporosis |  |  |  |  |  |  |  | $\begin{aligned} & \hline \text { rs6025 } \\ & (1 \mathrm{q} 24.2) \\ & \hline \end{aligned}$ |  |

Bayesian meta-analysis for studying cross-phenotype genetic associations

Table 5: Pleiotropy results by CPBayes for those SNPs in Table 3 and S19, at which some phenotypes (colored blue) were not selected in the optimal subset of non-null traits but produced a non-negligible value of trait-specific posterior probability of association $\left(\mathrm{PPA}_{j}\right)$.

| rsID | chrom <br> band | CPBayes <br> log10BF | CPBayes <br> PPNA | Important <br> phenotypes | $\mathrm{PPA}_{j}$ | Univariate <br> p-values |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| rs115946033 | 3 q 25.32 | 8.59 | $6.07 \mathrm{E}-16$ | Depressive Disorder | $37.2 \%$ | 0.0008 |
|  |  |  |  | Type 2 Diabetes | $100 \%$ | $3.83 \mathrm{E}-07$ |
| rs849135 | 7 p 15.1 | 7.66 | $5.22 \mathrm{E}-15$ | Asthma | $40.2 \%$ | $2.19 \mathrm{E}-05$ |
|  |  |  |  | Type 2 Diabetes | $100 \%$ | $1.79 \mathrm{E}-14$ |
| rs76517520 | 10 p 13 | 13.60 | $5.95 \mathrm{E}-21$ | Macular Degeneration | $25.7 \%$ | 0.02 |
|  |  |  |  | Peptic Ulcer | $100 \%$ | $1.39 \mathrm{E}-05$ |
| rs76075198 | 19 q 13.31 | 41.56 | $6.53 \mathrm{E}-49$ | Dyslipidemia | $100 \%$ | $5.29 \mathrm{E}-11$ |
|  |  |  |  | Peripheral Vascular Disease | $44.9 \%$ | 0.0067 |

Bayesian meta-analysis for studying cross-phenotype genetic associations

Figure 5: Forest plot for pleiotropic signal at rs6025 on chromosome 1. Phenotypes selected by either of the two methods are plotted. Blue colored bands present the primary trait-specific univariate odds ratio estimate with the corresponding $95 \%$ confidence interval. Red colored bands present the posterior mean of the trait-specific odds ratio along with the corresponding $95 \%$ credible interval obtained by CPBayes. The $y$-axis represents the value of odds ratio as 1 (null association). The $\log _{10} \mathrm{BF}$ obtained by CPBayes and $-\log _{10}\left(\right.$ ASSET p-value) $\left(-\log _{10} \mathrm{ASTpv}\right)$ are provided. Whether a phenotype was selected by a method is indicated by 'Yes' or 'No'. Trait-specific univariate association p-values are also provided.

| Pleiotropy signal at rs6025: $\log 10 B F=270.68,-\log 10(A S T p v)=10.86$ |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: |
| Disease name | p-values | CPBayes | $\begin{aligned} & \text { OR estimate \& 95\% CI } \\ & \text { CPBayes summary } \end{aligned}$ | ASSET |
| Dermatophytosis | 0.0018 | Yes |  | Yes |
| Hemorrhoids | 0.0014 | Yes |  | Yes |
| Iron Deficiency | 4e-04 | Yes |  | Yes |
| Osteoporosis | 2e-04 | Yes |  | Yes |
| Peripheral Vascular Disease | 6.81e-14 | Yes |  | Yes |
|  |  |  | dds ratio ${ }^{1.5}$ |  |

Bayesian meta-analysis for studying cross-phenotype genetic associations

Figure 6: Forest plot for pleiotropic signal at rs104558872 on chromosome 6. Phenotypes selected by either of the two methods are plotted. Blue colored bands present the primary trait-specific univariate odds ratio estimate with the corresponding $95 \%$ confidence interval. Red colored bands present the posterior mean of the trait-specific odds ratio along with the corresponding $95 \%$ credible interval obtained by CPBayes. The $y$-axis represents the value of odds ratio as 1 (null association). The $\log _{10} \mathrm{BF}$ obtained by CPBayes and $-\log _{10}$ (ASSET p-value) $\left(-\log _{10} \mathrm{ASTpv}\right)$ are provided. Whether a phenotype was selected by a method is indicated by 'Yes' or 'No'. Trait-specific univariate association p-values are also provided.

| Pleiotropy signal at rs10455872: $\log 10 B F=26.20,-\log 10($ ASTpv $)=10.26$ |  |  |  | ASSET |
| :---: | :---: | :---: | :---: | :---: |
| Disease name | p -values | CPBayes | OR estimate \& $95 \% \mathrm{Cl}$ <br> - CPBayes summary |  |
| Allergic Rhinitis | 0.071 | No |  | Yes |
| Cardiovascular Disease | $6.14 \mathrm{e}-05$ | Yes |  | Yes |
| Dyslipidemia | 6.97e-15 | Yes |  | Yes |
| Insomnia | 0.37 | No |  | Yes |
| Iron Deficiency | 0.57 | No |  | Yes |
| Irritable Bowel Syndrome | 0.34 | No |  | Yes |
| Osteoarthritis | 0.12 | No |  | Yes |
| Peripheral Vascular Disease | 2e-04 | Yes |  | Yes |
|  |  |  | dds ratio ${ }^{1.16}$ |  |

Bayesian meta-analysis for studying cross-phenotype genetic associations

Figure 7: Forest plot for pleiotropic signal at rs651007 on chromosome 9. Phenotypes selected by either of the two methods are plotted. Blue colored bands present the primary trait-specific univariate odds ratio estimate with the corresponding $95 \%$ confidence interval. Red colored bands present the posterior mean of the trait-specific odds ratio along with the corresponding $95 \%$ credible interval obtained by CPBayes. The $y$-axis represents the value of odds ratio as 1 (null association). The $\log _{10} \mathrm{BF}$ obtained by CPBayes and $-\log _{10}$ (ASSET p-value) $\left(-\log _{10} \mathrm{ASTpv}\right)$ are provided. Whether a phenotype was selected by a method is indicated by 'Yes' or 'No'. Trait-specific univariate association p-values are also provided.

| Pleiotropy signal at rs651007: $\log 10 B F=24.72,-\log 10(A S T p v)=15.51$ |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: |
| Disease name | p-values | CPBayes | $\begin{aligned} & \text { OR estimate \& } 95 \% \mathrm{Cl} \\ & \text { CPBayes summary } \end{aligned}$ | ASSET |
| Depressive Disorder | 0.094 | No |  | Yes |
| Dyslipidemia | 2.89e-15 | Yes |  | Yes |
| Hypertension | 0.03 | No |  | Yes |
| Hemorrhoids | 0.17 | No |  | Yes |
| Abdominal Hernia | 0.34 | No |  | Yes |
| Peripheral Vascular Disease | $9.55 \mathrm{e}-08$ | Yes |  | Yes |
| Peptic Ulcer | 0.025 | No |  | Yes |
| Varicose Veins | 0.0027 | No |  | Yes |
|  |  |  | Ids ratio ${ }_{\text {1.16 }}$ |  |

# Supplementary materials for "An efficient Bayesian meta-analysis approach for studying cross-phenotype genetic associations" by 

 Arunabha Majumdar, Tanushree Haldar, Sourabh Bhattacharya,John S. Witte.

## 7 Outline of mathematical derivation of the full conditional posterior distributions for the continuous spike

### 7.1 Correlated case

Here we derive the full conditional posterior distributions of the model parameters to perform Gibbs sampling for correlated summary statistics. If $S$ is the covariance matrix of $\hat{\boldsymbol{\beta}}$,

$$
\hat{\boldsymbol{\beta}} \mid \boldsymbol{\beta}, S \sim \operatorname{MVN}(\boldsymbol{\beta}, S)
$$

For $j=1, \ldots, K$,

$$
\begin{gather*}
\beta_{j} \mid z_{j}, \tau, d \stackrel{i n d}{\sim}\left(1-z_{j}\right) N\left(0, \tau^{2}\right)+z_{j} N\left(0,\left(\frac{\tau}{d}\right)^{2}\right) ; \quad \tau>0,0<d<1,\left(\frac{\tau}{d}\right)^{2}>\tau^{2} \\
P\left(z_{j}=1 \mid q\right)=q ; \quad P\left(z_{j}=0 \mid q\right)=(1-q) ; 0<q<1 \\
q \mid c_{1}, c_{2} \sim \operatorname{Beta}\left(c_{1}, c_{2}\right) \Rightarrow P\left(z_{j}=1\right)=E(q)=\frac{c_{1}}{c_{1}+c_{2}}  \tag{8}\\
d \mid e_{1}, e_{2} \sim \operatorname{Beta}\left(e_{1}, e_{2}\right)
\end{gather*}
$$

So, $\beta_{j} \mid z_{j}=0, \tau \sim \mathrm{~N}\left(0, \tau^{2}\right)$ and $\beta_{j} \mid z_{j}=1, \tau, d \sim \mathrm{~N}\left(0,\left(\frac{\tau}{d}\right)^{2}\right)$. Let $\Sigma_{1}=S$. Define, $\Sigma_{2}=\operatorname{diag}\left(\tau_{1}^{2}, \ldots, \tau_{K}^{2}\right)$ (a diagonal matrix with the diagonal elements $\tau_{1}^{2}, \ldots, \tau_{K}^{2}$ ), where $\tau_{j}=\tau$ if $z_{j}=0$, and $\tau_{j}=\frac{\tau}{d}$ if $z_{j}=1$, $j=1, \ldots, K$. So,

$$
\begin{equation*}
\boldsymbol{\beta} \mid Z, \tau, d \sim \operatorname{MVN}\left(\underset{\sim}{0}, \Sigma_{2}\right) \tag{9}
\end{equation*}
$$

7.1.1 Full conditional posterior distribution of $\boldsymbol{\beta}$

Let $[U]$ denote a generic notation of the probability distribution of a random variable $U$, and $\left[U_{1} \mid U_{2}\right]$ denote a generic notation of the conditional probability distribution of $U_{1}$ given $U_{2}$. We have considered fixed choice of $\tau$. Hence we drop it from the set of conditional parameters while writing the expressions of the full conditional distributions. For example, we write $[\boldsymbol{\beta} \mid Z, q, d, \hat{\boldsymbol{\beta}}]$ instead of $[\boldsymbol{\beta} \mid Z, q, \tau, d, \hat{\boldsymbol{\beta}}]$. Note that,

$$
\begin{equation*}
[\boldsymbol{\beta} \mid Z, q, d, \hat{\boldsymbol{\beta}}] \propto[\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}, Z, q, d] \propto[\hat{\boldsymbol{\beta}} \mid \boldsymbol{\beta}][\boldsymbol{\beta} \mid Z, d] \tag{10}
\end{equation*}
$$

Applying standard techniques from linear algebra and distribution theory of multivariate normal, one can obtain that:

$$
\begin{equation*}
\boldsymbol{\beta} \mid Z, q, d, \hat{\boldsymbol{\beta}} \sim \operatorname{MVN}\left[\left(\Sigma_{1}^{-1}+\Sigma_{2}^{-1}\right)^{-1} \Sigma_{1}^{-1} \hat{\boldsymbol{\beta}},\left(\Sigma_{1}^{-1}+\Sigma_{2}^{-1}\right)^{-1}\right] \tag{11}
\end{equation*}
$$

Note that $\Sigma_{2}$ is dependent on $Z$, and $Z$ influences the full conditional posterior distribution of $\boldsymbol{\beta}$ through the specification of $\Sigma_{2}$.

### 7.1.2 Full conditional posterior distribution of $Z$

Let $Z_{-j}=\left(z_{1}, \ldots, z_{j-1}, z_{j+1}, \ldots, z_{K}\right)$. Note that, $P\left(z_{j}=0 \mid Z_{-j}, \boldsymbol{\beta}, q, d, \hat{\boldsymbol{\beta}}\right) \propto\left[\beta_{j} \mid z_{j}=0\right]\left[z_{j}=0 \mid q\right]$. Similarly, $P\left(z_{j}=1 \mid Z_{-j}, \boldsymbol{\beta}, q, d, \hat{\boldsymbol{\beta}}\right) \propto\left[\beta_{j} \mid z_{j}=1, d\right]\left[z_{j}=1 \mid q\right]$. Combining these two equations, we obtain that:

$$
\begin{equation*}
P\left(z_{j}=0 \mid Z_{-j}, \boldsymbol{\beta}, q, d, \hat{\boldsymbol{\beta}}\right)=\frac{1}{1+\mathrm{ratio}_{j}}, \text { where } \text { ratio }_{j}=\frac{q}{1-q} d \exp \left[-\frac{\beta_{j}^{2}}{2 \tau^{2}}\left(d^{2}-1\right)\right] \tag{12}
\end{equation*}
$$

We note that the full conditional posterior distribution of $z_{j}$ does not depend on $Z_{-j}$. Hence, the full conditional distributions of $z_{1}, \ldots, z_{K}$ are independent.

### 7.1.3 Full conditional posterior distribution of $q$

Note that, $[q \mid \boldsymbol{\beta}, Z, d, \hat{\boldsymbol{\beta}}] \propto[Z \mid q][q] \propto \prod_{j=1}^{K}\left[z_{j} \mid q\right][q]$. Let $k_{1}=\sum_{j=1}^{K} z_{j}$, and $k_{0}=K-k_{1}$. Then it can easily be derived that:

$$
\begin{equation*}
q \mid \boldsymbol{\beta}, Z, d, \hat{\boldsymbol{\beta}} \sim \operatorname{Beta}\left(c_{1}+k_{1}, c_{2}+k_{0}\right) \tag{13}
\end{equation*}
$$

### 7.1.4 Full conditional posterior distribution of $d$

We assume that $e_{2}=1$ and derive a closed-form full conditional posterior distribution of $d$ under this restriction. Let $\sum_{j: z_{j}=1} z_{j}=k_{1}$.

Case 1: $k_{1}>0$
Of note, $[d \mid \boldsymbol{\beta}, Z, q, \hat{\boldsymbol{\beta}}] \propto \prod_{j: z_{j}=1}\left[\beta_{j} \mid z_{j}=1, d\right][d] \propto \exp \left[-C \times d^{2}\right] \times d^{\left(k_{1}+e_{1}\right)-1}$, where $C=\frac{1}{2 \tau^{2}} \sum_{j: z_{j}=1} \beta_{j}^{2}$. We consider the following transformation of variable: $y=2 C d^{2}$. It can be derived that: $y \sim \chi_{k_{1}+e_{1}}^{2}$ under the assumption that $y>0$.

Suppose that, we want to update the slab variance $\left(\frac{\tau}{d}\right)^{2}$ in a range (say, $v_{0}-v_{1}$ ) such that the corresponding range of $d$ is given by: $d_{0}<d<d_{1}$. Using the above transformation of variable, the corresponding range of $y$ is given by: $2 C d_{0}^{2}<y<2 C d_{1}^{2}$. Hence, $y \sim$ truncated central $\chi_{k_{1}+e_{1}}^{2}$, where $2 C d_{0}^{2}<y<2 C d_{1}^{2}$. Finally, the updated $d$ can be obtained by using the transformation $d=\sqrt{\frac{y}{2 C}}$.

## Case 2: $k_{1}=0$

It can easily be shown that: $d \mid \boldsymbol{\beta}, Z, q, \hat{\boldsymbol{\beta}} \sim \operatorname{truncated} \operatorname{Beta}\left(e_{1}, 1\right)$, where $d_{0}<d<d_{1}$.

In the Algorithm 1 in main text, we considered $e_{1}=1$, which is a natural choice under the absence of any prior information.

### 7.2 Uncorrelated case

When the summary statistics are uncorrelated, the full conditional posterior distributions of all the parameters except $\boldsymbol{\beta}$ remain the same as in the correlated case (described above). Now the full conditional distributions of $\beta_{1}, \ldots, \beta_{K}$ become independent. For $j=1, \ldots, K$,

$$
\begin{equation*}
\hat{\beta}_{j} \mid \beta_{j} \stackrel{\text { ind }}{\sim} \mathrm{N}\left(\beta_{j}, s_{j}^{2}\right), \text { and }\left[\beta_{j} \mid \boldsymbol{\beta}_{-j}, Z, q, d, \hat{\boldsymbol{\beta}}\right] \propto\left[\hat{\beta}_{j} \mid \beta_{j}\right] \times\left[\beta_{j} \mid z_{j}, d\right] \tag{14}
\end{equation*}
$$

Using the above equation, it's easy to derive the full conditional distribution as: for $j=1, \ldots, K$,
$\beta_{j} \mid \boldsymbol{\beta}_{-j}, Z, q, d, \hat{\boldsymbol{\beta}} \stackrel{\text { ind }}{\sim} \mathrm{N}\left(\frac{\sigma_{j}^{2}}{s_{j}^{2}} \hat{\beta}_{j}, \sigma_{j}^{2}\right)$, where $\frac{1}{\sigma_{j}^{2}}=\frac{1}{s_{j}^{2}}+\frac{1}{\tau_{j}^{2}}$.

## 8 Gibbs sampling algorithm for the Dirac spike

Here we outline the Gibbs sampler for the Dirac spike. We apply the BY procedure [Benjamini and Yekutieli, 2001] on the univariate association p-values of $K$ traits at an FDR level of 0.05 and assign $\beta_{j}=\hat{\beta}_{j}$ (since $\hat{\beta}_{j}$ is a consistent estimator of $\beta_{j}$ ) if $Y_{j}$ is found to be associated, otherwise we set $\beta_{j}=0 ; j=1, \ldots, K$. We also choose an initial value of $q$ as the proportion of non-null/associated traits detected by the BY procedure (the boundary situations of no/all non-null traits are taken care of appropriately).

Let $\boldsymbol{\beta}_{-j}=\left\{\beta_{1}, \ldots, \beta_{j-1}, \beta_{j+1}, \ldots, \beta_{K}\right\}$. Consider the following partition of $S$ :

$$
S=\left[\begin{array}{cc}
s_{j}^{2} & S_{j,-j} \\
S_{-j, j} & S_{-j,-j}
\end{array}\right]
$$

Let $\bar{\sigma}_{j}^{2}=s_{j}^{2}-\left(S_{j,-j} \times S_{-j,-j}^{-1} \times S_{-j, j}\right)$, and $m_{j,-j}=\hat{\beta}_{j}-\left(S_{j,-j} \times S_{-j,-j}^{-1} \times\left(\hat{\boldsymbol{\beta}}_{-j}-\boldsymbol{\beta}_{-j}\right)\right)$.

```
Algorithm 2 Gibbs sampling for the Dirac spike for correlated summary statistics
    Start:
    Assign the initial values of \(\boldsymbol{\beta}\) and \(q\) as discussed above.
    loop:
    For \(j=1, \ldots, K\), update \(\beta_{j}\) as follows: set \(\beta_{j}=0\) with probability \(P\left(\beta_{j}=0 \mid \boldsymbol{\beta}_{-j}, q, b, \hat{\boldsymbol{\beta}}\right)=\frac{1}{1+\text { ratio }_{j}}\),
    where \(\operatorname{ratio}_{j}=\frac{q}{(1-q)} \frac{\sigma_{j}}{b} e^{\frac{m_{j,-j}^{2} \times \sigma_{j}^{2}}{2 \sigma_{j}^{4}}}\). If \(\beta_{j}\) is selected to be non-zero, simulate it from \(\beta_{j} \mid \boldsymbol{\beta}_{-j}, q, b, \hat{\boldsymbol{\beta}} \sim\)
    \(\mathrm{N}\left(\frac{\sigma_{j}^{2} m_{j,-j}}{\bar{\sigma}_{j}^{2}}, \sigma_{j}^{2}\right)\), where \(\frac{1}{\sigma_{j}^{2}}=\frac{1}{\bar{\sigma}_{j}^{2}}+\frac{1}{b^{2}}\).
    5: Let \(k_{0}=\#\left\{\beta_{j}: \beta_{j}=0, j=1, \ldots, K\right\}\). Update \(q\) using it's full conditional posterior distribution which is
    a mixture of \(\left(k_{0}+1\right)\) Beta distributions as follows: for \(j=0,1, \ldots, k_{0}, q \mid \boldsymbol{\beta}, b, \hat{\boldsymbol{\beta}} \sim \operatorname{Beta}\left(c_{1}+K-j, c_{2}+j\right)\)
    with probability \(\propto\binom{k_{0}}{j}\left(\frac{1}{\sqrt{2 \pi} \times b}\right)^{K-j} \times \operatorname{Beta}\left(c_{1}+K-j, c_{2}+j\right)\). Here \(\operatorname{Beta}\left(r_{1}, r_{2}\right)\) denotes the normalizing
    constant of a \(\operatorname{Beta}\left(r_{1}, r_{2}\right)\) distribution.
6: Let \(b=\frac{1}{v}\), where \(v>0\). Suppose, we want to update \(b\) in a given range \(\left(b_{0}, b_{1}\right)\), and the corresponding range of \(v\) is given by: \(\left(v_{0}, v_{1}\right)\). We consider a uniform prior on \(v\). Let \(k_{1}=K-k_{0}\) and \(t_{j}=K-j+1\). If \(k_{1}>0\), let \(C=\sum_{j=1}^{K} \beta_{j}^{2}\). We update \(v\) using the transformation: \(v=\frac{\sqrt{y}}{\sqrt{C}}\), where \(y\) follows a mixture of \(\left(k_{0}+1\right)\) distributions - the \(j^{t h}\) distribution is a truncated (between \(v_{0}^{2} C-v_{1}^{2} C\) ) central chi-square distribution with degree of freedom \(t_{j}, j=0,1, \ldots, k_{0}\). The \(j^{t h}\) distribution is selected with probability \(w_{j} \propto w_{j}^{(1)} w_{j}^{(2)} w_{j}^{(3)} w_{j}^{(4)} ; w_{j}^{(1)}=\frac{1}{(2 \pi)^{\frac{K-j}{2}}}\binom{k_{0}}{j}(1-q)^{j} q^{K-j}, w_{j}^{(2)}=\frac{1}{2\left(v_{1}-v_{0}\right)} \frac{1}{C^{\frac{t_{j}}{2}}}, w_{j}^{(3)}=\) \(2^{\frac{t_{j}}{2}} \times \Gamma\left(\frac{t_{j}}{2}\right), w_{j}^{(4)}=P\left(v_{0}^{2} C<\chi_{t_{j}}^{2}<v_{1}^{2} C\right)\), where \(\chi_{t_{j}}^{2}\) is a central chi-square distribution with d.f. \(t_{j}\).
If \(k_{1}=0\), the full conditional posterior distribution of \(v\) is a mixture of \((K+1)\) distributions. For \(j=0,1, \ldots, K\), the \(j^{t h}\) distribution of the mixture is given by: \(t_{j} \times \frac{v^{K-j}}{v_{1}^{t_{j}}-v_{0}^{t_{j}}}, v_{0}<v<v_{1}\); and it is selected with probability \(w_{j} \propto w_{j}^{(1)} \times w_{j}^{(2)}\), where \(w_{j}^{(1)}=\frac{1}{v_{1}-v_{0}} \frac{1}{(2 \pi)^{\frac{K-j}{2}}}\binom{K}{j}(1-q)^{j} q^{K-j}\) and \(w_{j}^{(2)}=\frac{1}{t_{j}}\left(v_{1}^{t_{j}}-v_{0}^{t_{j}}\right)\). goto loop until all MCMC iterations are finished.
```

If the summary statistics are uncorrelated, step 4 of Algorithm 2 is modified as: for $j=1, \ldots, K$, set $\beta_{j}=0$ with probability $P\left(\beta_{j}=0 \mid \boldsymbol{\beta}_{-j}, q, b, \hat{\boldsymbol{\beta}}\right)=\frac{1}{1+\text { ratio }_{j}}$, where $\operatorname{ratio}_{j}=\left(\frac{q}{1-q}\right) \times \frac{\sigma_{j}}{b} \times \exp \left[\frac{\hat{\beta}_{j}^{2} \sigma_{j}^{2}}{2 s_{j}^{4}}\right]$, and $\frac{1}{\sigma_{j}^{2}}=\frac{1}{s_{j}^{2}}+\frac{1}{b^{2}}$; if $\beta_{j}$ is selected to be non-zero, simulate it from $\beta_{j} \mid \boldsymbol{\beta}_{-j}, q, b, \hat{\boldsymbol{\beta}} \sim \mathrm{~N}\left(\frac{\sigma_{j}^{2}}{s_{j}^{2}} \hat{\beta}_{j}, \sigma_{j}^{2}\right)$. All the other steps of the algorithm remain the same.

## 9 Outline of mathematical derivation of the full conditional posterior distributions for the Dirac spike

### 9.1 Correlated case

$$
\hat{\beta}_{j} \mid \beta_{j} \sim \mathrm{~N}\left(\beta_{j}, s_{j}^{2}\right), \text { and } \hat{\boldsymbol{\beta}} \mid \boldsymbol{\beta}, S \sim \operatorname{MVN}(\boldsymbol{\beta}, S)
$$

$$
\beta_{j}\left|q, b \stackrel{i . i . d .}{ }(1-q) \times \delta_{\{0\}}\left(\beta_{j}\right)+q \times \mathrm{N}\left(0, b^{2}\right) ; q\right| c_{1}, c_{2} \sim \operatorname{Beta}\left(c_{1}, c_{2}\right), 0<q<1 .
$$

$\delta_{\{0\}}\left(\beta_{j}\right)$ is defined as: $\delta_{\{0\}}\left(\beta_{j}\right)=1$ when $\beta_{j}=0$, and $\delta_{\{0\}}\left(\beta_{j}\right)=0$ when $\beta_{j} \neq 0$. The full likelihood of the model is given by: $[\hat{\boldsymbol{\beta}} \mid \boldsymbol{\beta}] \times[\boldsymbol{\beta} \mid q, b] \times[q] \times[b]$, where $[\boldsymbol{\beta} \mid q, b]=\prod_{j=1}^{K}\left[\beta_{j} \mid q, b\right]$. Next we derive the full conditional posterior distributions of different parameters. Let $\hat{\boldsymbol{\beta}}_{-j}=\left\{\hat{\beta}_{1}, \ldots, \hat{\beta}_{j-1}, \hat{\beta}_{j+1}, \ldots, \hat{\beta}_{K}\right\}$ and $\boldsymbol{\beta}_{-j}=\left\{\beta_{1}, \ldots, \beta_{j-1}, \beta_{j+1}, \ldots, \beta_{K}\right\}$.

### 9.1.1 Full conditional posterior distribution of $\boldsymbol{\beta}$

For $j=1$, it can be shown that:

$$
\begin{align*}
{\left[\beta_{1} \mid \boldsymbol{\beta}_{-1}, q, b, \hat{\boldsymbol{\beta}}\right] } & \propto\left[\hat{\boldsymbol{\beta}}, \beta_{1}, \boldsymbol{\beta}_{-1}, q, b\right] \\
& \propto\left[\hat{\beta}_{1} \mid \hat{\boldsymbol{\beta}}_{-1}, \beta_{1}, \boldsymbol{\beta}_{-1}, q, b\right] \times\left[\hat{\boldsymbol{\beta}}_{-1}, \beta_{1}, \boldsymbol{\beta}_{-1}, q, b\right]  \tag{15}\\
& \propto\left[\hat{\boldsymbol{\beta}}_{1} \mid \hat{\boldsymbol{\beta}}_{-1}, \boldsymbol{\beta}\right] \times\left[\beta_{1} \mid q, b\right]
\end{align*}
$$

A general result: Suppose, $Y \sim \operatorname{MVN}(\mu, \Sigma)$ and $\left(Y_{1}, Y_{2}\right)$ is a partition of $Y$ with the corresponding partition of the mean vector and the covariance matrix as: $\mu=\left(\mu_{1}, \mu_{2}\right)$ and

$$
\Sigma=\left[\begin{array}{ll}
\Sigma_{11} & \Sigma_{12} \\
\Sigma_{21} & \Sigma_{22}
\end{array}\right]
$$

Then, $Y_{1} \mid Y_{2}=y_{2} \sim \operatorname{MVN}(\bar{\mu}, \bar{\Sigma})$, where $\bar{\mu}=\mu_{1}+\Sigma_{12} \Sigma_{22}^{-1}\left(y_{2}-\mu_{2}\right)$ and $\bar{\Sigma}=\Sigma_{11}-\Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$.

Using the above general result, we can obtain $\left[\hat{\beta}_{1} \mid \hat{\boldsymbol{\beta}}_{-1}, \boldsymbol{\beta}\right]$. Let the partition of $S$ according to the partition of $\boldsymbol{\beta}=\left(\beta_{1}, \boldsymbol{\beta}_{-1}\right)$ be given by:

$$
S=\left[\begin{array}{cc}
S_{11} & S_{1,-1} \\
S_{-1,1} & S_{-1,-1}
\end{array}\right]
$$

Here, $S_{11}=s_{1}^{2}$ is a scalar, $S_{1,-1}\left(=\left(s_{2}^{2}, \ldots, s_{K}^{2}\right)\right)$ is a vector of length $(K-1), S_{-1,-1}$ is a matrix of order $(K-1) \times(K-1)$. Thus, $\hat{\beta}_{1} \mid \hat{\boldsymbol{\beta}}_{-1}, \boldsymbol{\beta} \sim \operatorname{MVN}\left(\bar{\mu}_{1}, \bar{\Sigma}_{1}\right)$, where $\bar{\mu}_{1}=\beta_{1}+S_{1,-1} \times S_{-1,-1}^{-1} \times\left(\hat{\boldsymbol{\beta}}_{-1}-\beta_{-1}\right)=$ $\beta_{1}+\Sigma_{1,-1} \times\left(\hat{\boldsymbol{\beta}}_{-1}-\beta_{-1}\right) ; \Sigma_{1,-1}=S_{1,-1} \times S_{-1,-1}^{-1}$. And, $\bar{\Sigma}_{1}=S_{11}-S_{1,-1} \times S_{-1,-1}^{-1} \times S_{-1,1}=\bar{\sigma}_{1}^{2}$, say. Thus,

$$
\begin{equation*}
\left[\hat{\beta}_{1} \mid \hat{\boldsymbol{\beta}}_{-1}, \boldsymbol{\beta}\right]=\frac{1}{\sqrt{2 \pi} \bar{\sigma}_{1}} e^{-\frac{1}{2 \bar{\sigma}_{1}^{2}}\left(\beta_{1}-m_{1,-1}\right)^{2}} ; \text { where } m_{1,-1}=\hat{\beta}_{1}-\Sigma_{1,-1}\left(\hat{\boldsymbol{\beta}}_{-1}-\boldsymbol{\beta}_{-1}\right) \tag{16}
\end{equation*}
$$

Hence,

$$
\begin{align*}
{\left[\beta_{1} \mid \boldsymbol{\beta}_{-1}, q, b, \hat{\boldsymbol{\beta}}\right] } & =\left[\hat{\beta}_{1} \mid \hat{\boldsymbol{\beta}}_{-1}, \boldsymbol{\beta}\right] \times\left[\beta_{1} \mid q, b\right] \\
& =\frac{1}{\sqrt{2 \pi} \bar{\sigma}_{1}} e^{-\frac{1}{2 \bar{\sigma}_{1}^{2}}\left(\beta_{1}-m_{1,-1}\right)^{2}} \times\left[(1-q) \times \delta_{\{0\}}\left(\beta_{1}\right)+q \times \frac{1}{\sqrt{2 \pi} b} e^{\left.-\frac{\beta_{1}^{2}}{2 b^{2}}\right]}\right. \tag{17}
\end{align*}
$$

It is straightforward to derive from the above equation that:

$$
\begin{align*}
{\left[\beta_{1} \mid \boldsymbol{\beta}_{-1}, q, b, \hat{\boldsymbol{\beta}}\right] } & =0, \text { with probability } \propto(1-q) \times \frac{1}{\sqrt{2 \pi} \bar{\sigma}_{1}} e^{-\frac{m_{1,-1}^{2}}{2 \bar{\sigma}_{1}^{2}}} \\
& =\mathrm{N}\left(\frac{\sigma_{1}^{2} m_{1,-1}}{\bar{\sigma}_{1}^{2}}, \sigma_{1}^{2}\right), \text { with probability } \propto q \times \frac{\sigma_{1}}{\sqrt{2 \pi} \bar{\sigma}_{1} b} e^{-\frac{m_{1,-1}^{2}}{2 \bar{\sigma}_{1}^{2}}\left(1-\frac{\sigma_{1}^{2}}{\bar{\sigma}_{1}^{2}}\right)}, \tag{18}
\end{align*}
$$

where $\frac{1}{\sigma_{1}^{2}}=\frac{1}{\bar{\sigma}_{1}^{2}}+\frac{1}{b^{2}}$. Hence, $\sigma_{1}^{2}=\frac{1}{\frac{1}{\bar{\sigma}_{1}^{2}}+\frac{1}{b^{2}}}$.

More explicitly,

$$
\begin{equation*}
\left[\beta_{1} \mid \boldsymbol{\beta}_{-1}, q, b, \hat{\boldsymbol{\beta}}\right]=\operatorname{pr}_{1} \times \delta_{\{0\}}\left(\beta_{1}\right)+\left(1-\operatorname{pr}_{1}\right) \times \mathrm{N}\left(\frac{\sigma_{1}^{2} m_{1,-1}}{\bar{\sigma}_{1}^{2}}, \sigma_{1}^{2}\right), \tag{19}
\end{equation*}
$$

where $\operatorname{pr}_{1}=\frac{1}{1+\text { ratio }_{1}}$, and ratio $_{1}=\frac{q}{(1-q)} \frac{\sigma_{1}}{b} e^{\frac{m_{1,-1}^{2} \times \sigma_{1}^{2}}{2 \bar{\sigma}_{1}^{4}}}$

### 9.1.2 Full conditional posterior distribution of $q$

Next we derive the full conditional distribution of $q$. Let $k_{0}$ be the number of zeros in $\boldsymbol{\beta}$, and $k_{1}\left(=K-k_{0}\right)$ be the number of non-zero elements in $\boldsymbol{\beta}$. Let $\operatorname{dnorm}(x, \mu, \sigma)$ denote the probability density function of a normal distribution at $x$ with mean $\mu$ and variance $\sigma^{2}$. Under the Dirac spike, since $\beta_{j}$ has a positive mass $(1-q)$ at $0,\left[\beta_{j}=0 \mid q, b\right]=(1-q)+q \times \operatorname{dnorm}(0,0, b)$. Let $\boldsymbol{\beta}_{\text {obs }}$ denote an observed value of $\boldsymbol{\beta}$ in a MCMC iteration.

$$
\begin{align*}
{\left[q \mid \boldsymbol{\beta}=\boldsymbol{\beta}_{o b s}, b, \hat{\boldsymbol{\beta}}\right] } & \propto\left[\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}=\boldsymbol{\beta}_{o b s}, q, b\right] \\
& \propto\left[\boldsymbol{\beta}=\boldsymbol{\beta}_{o b s} \mid q, b\right] \times[q] \\
& =\prod_{j=1}^{K}\left[\beta_{j}=\beta_{j, o b s} \mid q, b\right] \times[q] \\
& =\prod_{i: \beta_{i, o b s}=0}\left[\beta_{i}=\beta_{i, o b s} \mid q, b\right] \times \prod_{i: \beta_{i, o b s} \neq 0}\left[\beta_{i}=\beta_{i, o b s} \mid q, b\right] \times[q] \\
& =\{(1-q)+q \times d n o r m(0,0, b)\}^{k_{0}} \times \prod_{i: \beta_{i, o b s} \neq 0} q \times d n o r m\left(\beta_{i, o b s}, 0, b\right) \times[q] \\
& =\left\{(1-q)+q \times \frac{1}{\sqrt{2 \pi} b}\right\}^{k_{0}} \times \prod_{i: \beta_{i, o b s} \neq 0} q \times \frac{1}{\sqrt{2 \pi} b} e^{-\frac{\beta_{i, o b s}^{2}}{2 b^{2}}} \times[q] \\
& =\left\{(1-q)+q \times \frac{1}{\sqrt{2 \pi} b}\right\}^{k_{0}} \times q^{k_{1}} \times\left(\frac{1}{\sqrt{2 \pi} b}\right)^{k_{1}} e^{-\frac{\sum_{i: \beta_{i, o b s} \neq 0}^{2 b^{2}}}{\beta_{i, o b s}^{2}} \times[q]} \\
& =\sum_{j=0}^{k_{0}}\binom{k_{0}}{j} \times(1-q)^{j} \times q^{k_{0}-j} \times\left(\frac{1}{\sqrt{2 \pi} b}\right)^{k_{0}-j} \times q^{k_{1}} \times\left(\frac{1}{\sqrt{2 \pi} b}\right)^{k_{1}} \times e^{-\frac{c o n s t}{b^{2}}} \times[q] ; \text { const }=\frac{\sum_{i=1}^{K} \beta_{i, o b s}^{2}}{2} \\
& \propto \sum_{j=0}^{k_{0}}\binom{k_{0}}{j} \times(1-q)^{j} \times q^{k_{0}-j+k_{1}} \times\left(\frac{1}{\sqrt{2 \pi b}}\right)^{k_{0}-j+k_{1}} \times \frac{1}{B e t a\left(c_{1}, c_{2}\right)} q^{c_{1}-1}(1-q)^{c_{2}-1} \\
& \propto \sum_{j=0}^{k_{0}}\binom{k_{0}}{j}\left(\frac{1}{\sqrt{2 \pi} b}\right)^{K-j} \times \operatorname{Beta}\left(c_{1}+K-j, c_{2}+j\right) \times\left[q \sim \operatorname{Beta}\left(c_{1}+K-j, c_{2}+j\right)\right] \tag{20}
\end{align*}
$$

Thus, the full conditional posterior distribution of $q$ is a mixture of $\left(k_{0}+1\right)$ Beta distributions as follows: for $j=0,1, \ldots, k_{0}, q \mid \boldsymbol{\beta}, b, \hat{\boldsymbol{\beta}} \sim \operatorname{Beta}\left(c_{1}+K-j, c_{2}+j\right)$ with probability $\propto\binom{k_{0}}{j}\left(\frac{1}{\sqrt{2 \pi b}}\right)^{K-j} \times \operatorname{Beta}\left(c_{1}+K-\right.$ $\left.j, c_{2}+j\right)$. Here $\operatorname{Beta}\left(r_{1}, r_{2}\right)$ denotes the normalizing constant of the $\operatorname{Beta}\left(r_{1}, r_{2}\right)$ distribution.

### 9.1.3 Full conditional posterior distribution of $b$

Since $b>0$, let $b=\frac{1}{v}$, where $v>0$. Thus, for $j=1, \ldots, K, \beta_{j} \mid q, v \xrightarrow{i . i . d .}(1-q) \times \delta_{\{0\}}\left(\beta_{j}\right)+q \times \mathrm{N}\left(0, \frac{1}{v^{2}}\right)$. Suppose, we want to update $b$ in a given range $\left(b_{0}, b_{1}\right)$. Let the corresponding range of $v$ be given by $\left(v_{0}, v_{1}\right)$. We assume a uniform prior on $v$. So, $[v]=\frac{1}{v_{1}-v_{0}}$, where $v_{0}<v<v_{1}$. Suppose, $\left(\beta_{1, o b s}, \ldots, \beta_{K, o b s}\right)$ denote an observed value of $\left(\beta_{1}, \ldots, \beta_{K}\right)$ in a MCMC iteration. Let $k_{1}=\#\left\{\beta_{j} \neq 0: j=1, \ldots, K\right\}$.

Case 1: $k_{1}>0$

$$
\begin{align*}
& {[v \mid \boldsymbol{\beta}, q, \hat{\boldsymbol{\beta}}] \propto[\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}, q, v]} \\
& \propto[\boldsymbol{\beta} \mid q, v] \times[v] \\
& =\prod_{j=1}^{K}\left[\beta_{j} \mid q, v\right] \times[v] \\
& =\prod_{i: \beta_{i, o b s}=0}\left[\beta_{i}=\beta_{i, o b s} \mid q, v\right] \prod_{i: \beta_{i, o b s} \neq 0}\left[\beta_{i}=\beta_{i, o b s} \mid q, v\right] \times[v] \\
& =\prod_{i: \beta_{i, o b s}=0}\left\{(1-q)+q \times \frac{v}{\sqrt{2 \pi}}\right\} \prod_{i: \beta_{i, o b s} \neq 0} q \times \frac{v}{\sqrt{2 \pi}} e^{-\frac{v^{2} \beta_{i, o b s}^{2}}{2}} \times[v]  \tag{21}\\
& =\left\{(1-q)+q \times \frac{v}{\sqrt{2 \pi}}\right\}^{k_{0}} \times q^{k_{1}} \times \frac{v^{k_{1}}}{(\sqrt{2 \pi})^{k_{1}}} \times e^{-\frac{v^{2}}{2} C} \times[v], \text { where } C=\sum_{i=1}^{K} \beta_{i, o b s}^{2} \\
& =\sum_{j=0}^{k_{0}}\binom{k_{0}}{j} \times(1-q)^{j} \times q^{K-j} \times \frac{1}{(\sqrt{2 \pi})^{K-j}} \times v^{K-j} \times e^{-\frac{v^{2}}{2} C} \times[v] \\
& =\sum_{j=0}^{k_{0}} w_{j}^{1} \times v^{K-j} \times e^{-\frac{v^{2}}{2} C} \times[v], \text { where } w_{j}^{1}=\binom{k_{0}}{j} \times(1-q)^{j} \times q^{K-j} \times \frac{1}{(\sqrt{2 \pi})^{K-j}} \\
& \propto \sum_{j=0}^{k_{0}} w_{j}^{1} \times \frac{1}{v_{1}-v_{0}} v^{K-j} \times e^{-\frac{v^{2}}{2} C}, v_{0}<v<v_{1} .
\end{align*}
$$

Now we consider the transformation: $y=v^{2} C \Rightarrow v=\frac{\sqrt{y}}{\sqrt{C}}$, and $v_{0}^{2} C<y<v_{1}^{2} C$. Using the above equation, we obtain that $y$ follows a mixture of $\left(k_{0}+1\right)$ truncated central chi-square distributions as follows:
for $j=0, \ldots, k_{0}, y \sim \operatorname{truncated}$ central $\chi_{t_{j}}^{2}$ with probability $w_{j}$, where $v_{0}^{2} C<y<v_{1}^{2} C$ and $t_{j}=K-j+1$. The mixture weight $w_{j}$ is given by: $w_{j} \propto w_{j}^{(1)} w_{j}^{(2)} w_{j}^{(3)} w_{j}^{(4)} ; w_{j}^{(1)}=\frac{1}{(2 \pi)^{\frac{K-j}{2}}}\binom{k_{0}}{j}(1-q)^{j} q^{K-j}, w_{j}^{(2)}=$ $\frac{1}{2\left(v_{1}-v_{0}\right)} \frac{1}{C^{\frac{t_{j}}{2}}}, w_{j}^{(3)}=2^{\frac{t_{j}}{2}} \times \Gamma\left(\frac{t_{j}}{2}\right), w_{j}^{(4)}=P\left(v_{0}^{2} C<\chi_{t_{j}}^{2}<v_{1}^{2} C\right)$. The updated $v$ is obtained from updated $y$ using the transformation: $v=\frac{\sqrt{y}}{\sqrt{C}}$.

Case 2: $k_{1}=0$
Similarly, if $k_{1}=0$, we can derive the full conditional posterior distribution of $v$ which appears to be a mixture of $(K+1)$ distributions. For $j=0,1, \ldots, K$, the $j^{t h}$ distribution of the mixture is given by: $t_{j} \times \frac{v^{K-j}}{v_{1}^{t_{j}}-v_{0}^{t_{j}}}, v_{0}<v<v_{1}$. Here $t_{j}=K-j+1$. The probability of the $j^{t h}$ mixture component is given by:
bioRxiv preprint doi: https://doi.org/10.1101/101543; this version posted January 19, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Bayesian meta-analysis for studying cross-phenotype genetic associations

$$
w_{j} \propto w_{j}^{(1)} \times w_{j}^{(2)} ; w_{j}^{(1)}=\frac{1}{v_{1}-v_{0}} \frac{1}{(2 \pi)^{\frac{K-j}{2}}}\binom{K}{j}(1-q)^{j} q^{K-j} \text { and } w_{j}^{(2)}=\frac{1}{t_{j}}\left(v_{1}^{t_{j}}-v_{0}^{t_{j}}\right)
$$

### 9.2 Uncorrelated case

For uncorrelated summary statistics, the full conditional posterior distribution of all the parameters except $\boldsymbol{\beta}$ remain the same. The derivation of full conditional posterior distribution of $\boldsymbol{\beta}$ for uncorrelated summary statistics is straightforward and will easily follow from the derivation for correlated summary statistics.

Figure S1: A diagram presenting the continuous spike and slab prior modeling pleiotropy with the spike variance $\tau^{2}=10^{-4}$.

Continuous spike and slab prior modeling pleiotropy


Bayesian meta-analysis for studying cross-phenotype genetic associations

Table S1: Main features of CPBayes and ASSET

| Features | CPBayes | ASSET |
| :--- | :--- | :--- |
| Paradigm | Bayesian | Frequentist |
| Measure of overall <br> pleiotropic association | Bayes factor <br> PPNA | P-value |
| Simultaneous selection <br> of non-null traits | Yes | Yes |
| Specificity of selection | High | Low-moderate |
| Sensitivity of selection | Moderate-high | Low-high |
| Heterogeneity in <br> direction of effects | Yes | Yes |
| Heterogeneity in <br> size of effects | Yes | No |

Table S2: Summary of measures for the evidence of the overall pleiotropic association under the global null hypothesis of no association when a cohort study with 15000 individuals is considered.

| K | m |  | mean | sd | min | quantile |  |  |  |  |  | max |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  | 5\% | 25\% | 50\% | 75\% | 95\% | 99\% |  |
| 5 | 0.3 | $\log _{10} \mathrm{BF}$ | -2.55 | 0.28 | -2.83 | -2.80 | -2.72 | -2.62 | -2.47 | -2.10 | -1.55 | -0.003 |
|  | 0.1 | $-\log _{10}$ PPNA | 0.05 | 0.09 | 0.02 | 0.02 | 0.02 | 0.03 | 0.04 | 0.09 | 0.27 | 1.50 |
|  |  | $-\log _{10}$ ASTpv | 0.36 | 0.53 | 0.00 | 0.004 | 0.06 | 0.20 | 0.45 | 1.21 | 2.32 | 6.20 |
|  |  | $\log _{10} \mathrm{BF}$ | -2.30 | 0.34 | -2.68 | -2.61 | -2.52 | -2.40 | -2.21 | -1.78 | -0.66 | -0.35 |
|  |  | $-\log _{10}$ PPNA | 0.08 | 0.13 | 0.03 | 0.03 | 0.04 | 0.05 | 0.08 | 0.18 | 0.89 | 1.17 |
|  |  | $-\log _{10}$ ASTpv | 0.36 | 0.45 | 0.00 | 0.002 | 0.06 | 0.20 | 0.50 | 1.28 | 2.13 | 3.48 |
| 10 | 0.3 | $\log _{10} \mathrm{BF}$ | -4.07 | 0.24 | -4.37 | -4.31 | -4.23 | -4.14 | -3.99 | -3.67 | -3.14 | -2.58 |
|  |  | $-\log _{10}$ PPNA | 0.04 | 0.05 | 0.02 | 0.02 | 0.03 | 0.03 | 0.04 | 0.09 | 0.24 | 0.57 |
|  |  | $-\log _{10}$ ASTpv | 0.18 | 0.50 | 0.00 | 0.00 | 0.00 | 0.02 | 0.14 | 0.80 | 2.34 | 4.72 |
|  | 0.1 | $\log _{10} \mathrm{BF}$ | -3.87 | 0.26 | -4.21 | -4.13 | -4.04 | -3.94 | -3.78 | -3.32 | -2.89 | -2.57 |
|  |  | $-\log _{10} \mathrm{PPNA}$ | 0.07 | 0.06 | 0.03 | 0.03 | 0.04 | 0.05 | 0.07 | 0.17 | 0.36 | 0.57 |
|  |  | $-\log _{10}$ ASTpv | 0.15 | 0.31 | 0.00 | 0.00 | 0.00 | 0.02 | 0.15 | 0.72 | 1.69 | 2.12 |
| 15 | 0.3 | $\log _{10} \mathrm{BF}$ | -5.54 | 0.31 | -5.82 | -5.78 | -5.69 | -5.60 | -5.48 | -5.13 | -4.88 | -0.14 |
|  |  | $-\log _{10}$ PPNA | 0.05 | 0.20 | 0.02 | 0.02 | 0.03 | 0.03 | 0.05 | 0.09 | 0.16 | 4.37 |
|  |  | $-\log _{10}$ ASTpv | 0.07 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.19 | 1.38 | 8.27 |
|  | 0.1 | $\log _{10} \mathrm{BF}$ | -5.30 | 0.27 | -5.66 | -5.58 | -5.49 | -5.37 | -5.19 | -4.77 | -4.27 | -4.10 |
|  |  | $-\log _{10}$ PPNA | 0.08 | 0.07 | 0.03 | 0.04 | 0.04 | 0.06 | 0.08 | 0.19 | 0.44 | 0.56 |
|  |  | $-\log _{10}$ ASTpv | 0.06 | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.27 | 1.11 | 3.28 |

Legend: $K$ - total number of phenotypes, $m$ - allele frequency at the marker $\mathrm{SNP} ; K=5,10,15$, and $m=0.3$, 0.1 . The following abbreviations denote $-\log _{10} \mathrm{BF}: \log _{10}$ (Bayes factor), $-\log _{10}$ PPNA: $-\log _{10}$ (posterior probability of null association), $-\log _{10}$ ASTpv: $-\log _{10}$ (ASSET p-value). For a cohort study, since the summary statistics are correlated, the combined strategy of CPBayes is implemented. Different summary measures obtained across 500 replications are provided: mean, standard deviation (sd), minimum (min), $5 \%$ quantile, $25 \%$ quantile, $50 \%$ quantile, $75 \%$ quantile, $95 \%$ quantile, $99 \%$ quantile, and maximum (max).

Table S3: Summary of measures for the evidence of the overall pleiotropic association when a subset of traits are associated for 5 non-overlapping case-control studies. Here 1 and $\mathbf{2}$ among 5 traits are associated.

|  | $m$ |  |  |  |  |  |  | quantil |  |  |  | max |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $K_{1}^{+}, K_{1}^{-}$ |  |  | mean | sd | min | 5\% | 25\% | 50\% | 75\% | 95\% | 99\% |  |
| 1,0 | 0.3 | $\log _{10} \mathrm{BF}$ | 11.05 | 15.21 | -2.98 | -2.54 | -0.77 | 5.12 | 19.30 | 42.71 | 61.24 | 70.26 |
|  |  | $-\log _{10}$ PPNA | 12.71 | 15.06 | 0.01 | 0.04 | 0.80 | 6.61 | 20.80 | 44.20 | 62.73 | 71.76 |
|  |  | - $\log _{10}$ ASTpv | 7.73 | 5.74 | 0.06 | 0.82 | 2.97 | 6.44 | 11.68 | 18.36 | 23.98 | 26.34 |
|  | 0.1 | $\log _{10} \mathrm{BF}$ | 4.22 | 10.17 | -2.81 | -2.65 | -2.00 | -0.51 | 6.95 | 24.89 | 45.04 | 56.89 |
|  |  | $-\log _{10}$ PPNA | 6.02 | 9.97 | 0.02 | 0.03 | 0.12 | 1.03 | 8.44 | 26.38 | 46.54 | 58.38 |
|  |  | - $\log _{10}$ ASTpv | 3.48 | 2.83 | 0.01 | 0.25 | 1.13 | 2.75 | 5.16 | 8.73 | 11.92 | 14.65 |
| 2,0 | 0.3 | $\log _{10} \mathrm{BF}$ | 31.70 | 26.39 | -2.78 | -1.03 | 9.11 | 27.99 | 48.92 | 80.29 | 102.88 | 126.71 |
|  |  | - $\log _{10}$ PPNA | 33.22 | 26.36 | 0.02 | 0.59 | 10.60 | 29.48 | 50.41 | 81.79 | 104.37 | 128.21 |
|  |  | - $\log _{10}$ ASTpv | 15.27 | 7.92 | 0.23 | 3.44 | 9.27 | 14.67 | 20.05 | 29.14 | 34.48 | 43.27 |
|  | 0.1 | $\log _{10} \mathrm{BF}$ | 15.00 | 19.73 | -2.74 | -1.99 | -0.03 | 6.88 | 24.90 | 57.87 | 78.48 | 90.53 |
|  |  | - $\log _{10}$ PPNA | 16.58 | 19.65 | 0.02 | 0.12 | 1.48 | 8.37 | 26.39 | 59.36 | 79.97 | 92.02 |
|  |  | - $\log _{10}{ }^{\text {ASTpv }}$ | 6.54 | 3.84 | 0.07 | 1.28 | 3.24 | 6.08 | 9.06 | 13.53 | 16.44 | 19.03 |
| 1,1 | 0.3 | $\log _{10} \mathrm{BF}$ | 28.15 | 23.99 | -2.75 | -0.70 | 7.78 | 24.71 | 42.89 | 72.80 | 98.50 | 117.92 |
|  |  | $-\log _{10}$ PPNA | 29.67 | 23.96 | 0.02 | 0.86 | 9.28 | 26.20 | 44.38 | 74.29 | 100.00 | 119.41 |
|  |  | - $\log _{10}$ ASTpv | 15.41 | 7.92 | 0.57 | 3.82 | 9.20 | 14.96 | 20.57 | 29.46 | 36.45 | 41.72 |
|  | 0.1 | $\log _{10} \mathrm{BF}$ | 11.18 | 17.06 | -2.61 | -2.09 | -0.16 | 3.28 | 17.60 | 44.54 | 70.04 | 141.15 |
|  |  | $-\log _{10}$ PPNA | 12.76 | 16.99 | 0.03 | 0.10 | 1.35 | 4.77 | 19.09 | 46.03 | 71.53 | 142.64 |
|  |  | - $\log _{10}$ ASTpv | 6.32 | 3.63 | 0.19 | 1.31 | 3.76 | 5.61 | 8.63 | 12.94 | 16.11 | 24.97 |

Legend: $m$ - allele frequency at the risk SNP; $m=0.3,0.1$. The number of positively and negatively associated traits are denoted by $K_{1}^{+}$and $K_{1}^{-}$, respectively. Hence the total number of associated traits is $K_{1}=K_{1}^{+}+K_{1}^{-}$. The following abbreviations denote $-\log _{10} \mathrm{BF}: \log _{10}$ (Bayes factor), $-\log _{10} \mathrm{PPNA}:-\log _{10}$ (posterior probability of null association), $-\log _{10}(\mathrm{ASTpv}):-\log _{10}(\mathrm{ASSET}$ p-value). For multiple studies with no overlapping subjects, the uncorrelated version of CPBayes is implemented. Different summary measures obtained across 500 replications are provided: mean, standard deviation (sd), minimum (min), $5 \%$ quantile, $25 \%$ quantile, $50 \%$ quantile, $75 \%$ quantile, $95 \%$ quantile, $99 \%$ quantile, and maximum (max).

Table S4: Summary of measures for the evidence of the overall pleiotropic association when a subset of traits are associated for $\mathbf{5}$ non-overlapping case-control studies. Here $\mathbf{3}$ and $\mathbf{4}$ among $\mathbf{5}$ traits are associated.

|  | $m$ |  |  |  |  |  |  | quant |  |  |  | max |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $K_{1}^{+}, K_{1}^{-}$ |  |  | mean | sd | min | 5\% | 25\% | 50\% | 75\% | 95\% | 99\% |  |
| 3,0 | 0.3 | $\log _{10} \mathrm{BF}$ | 61.61 | 40.01 | -2.44 | 4.61 | 30.29 | 58.08 | 85.48 | 134.62 | 175.19 | 197.55 |
|  |  | - $\log _{10}$ PPNA | 63.11 | 40.00 | 0.05 | 6.11 | 31.78 | 59.57 | 86.97 | 136.11 | 176.68 | 199.04 |
|  |  | - $\log _{10}$ ASTpv | 22.88 | 10.03 | 1.13 | 8.78 | 15.36 | 21.96 | 28.93 | 41.63 | 50.08 | 53.57 |
|  | 0.1 | $\log _{10} \mathrm{BF}$ | 33.38 | 34.50 | -2.37 | -0.84 | 3.62 | 24.34 | 52.62 | 99.66 | 142.34 | 171.82 |
|  |  | - $\log _{10}$ PPNA | 34.89 | 34.48 | 0.05 | 0.74 | 5.11 | 25.83 | 54.11 | 101.15 | 143.83 | 173.31 |
|  |  | - $\log _{10}$ ASTpv | 9.94 | 4.87 | 0.38 | 3.28 | 6.09 | 9.52 | 12.93 | 18.91 | 23.68 | 25.43 |
| 2,1 | 0.3 | $\log _{10} \mathrm{BF}$ | 60.31 | 39.41 | -2.07 | 4.80 | 29.29 | 56.15 | 84.89 | 131.91 | 170.04 | 201.46 |
|  |  | $-\log _{10}$ PPNA | 61.80 | 39.41 | 0.10 | 6.29 | 30.79 | 57.64 | 86.38 | 133.40 | 171.53 | 202.95 |
|  |  | - $\log _{10}$ ASTpv | 23.01 | 9.65 | 1.95 | 8.48 | 16.03 | 22.08 | 29.42 | 40.06 | 47.69 | 56.17 |
|  | 0.1 | $\log _{10} \mathrm{BF}$ | 30.24 | 31.41 | -2.60 | -1.22 | 3.58 | 21.99 | 47.77 | 90.33 | 129.23 | 188.29 |
|  |  | $-\log _{10}$ PPNA | 31.76 | 31.39 | 0.03 | 0.46 | 5.07 | 23.48 | 49.26 | 91.82 | 130.72 | 189.78 |
|  |  | $-\log _{10}$ ASTpv | 9.67 | 4.47 | 0.51 | 3.36 | 6.25 | 9.22 | 12.39 | 17.44 | 21.33 | 27.02 |
| 4,0 | 0.3 | $\log _{10} \mathrm{BF}$ | 95.17 | 48.22 | -0.33 | 24.89 | 59.37 | 91.99 | 128.33 | 179.22 | 214.40 | 245.43 |
|  |  | $-\log _{10}$ PPNA | 96.66 | 48.22 | 1.19 | 26.38 | 60.86 | 93.48 | 129.82 | 180.71 | 215.89 | 246.92 |
|  |  | $-\log _{10}$ ASTpv | 31.10 | 11.45 | 3.93 | 13.73 | 22.87 | 30.59 | 39.11 | 50.95 | 58.48 | 67.47 |
|  | 0.1 | $\log _{10} \mathrm{BF}$ | 56.83 | 46.68 | -2.16 | 0.80 | 19.87 | 48.05 | 80.55 | 146.62 | 213.33 | 234.68 |
|  |  | $-\log _{10} \mathrm{PPNA}$ | 58.32 | 46.68 | 0.08 | 2.30 | 21.36 | 49.54 | 82.04 | 148.12 | 214.82 | 236.17 |
|  |  | $-\log _{10}$ ASTpv | 13.50 | 5.62 | 1.01 | 5.64 | 9.36 | 12.71 | 16.70 | 24.19 | 28.88 | 33.48 |
| 2,2 | 0.3 | $\log _{10} \mathrm{BF}$ | 88.32 | 44.33 | 0.31 | 23.17 | 55.83 | 86.42 | 114.73 | 168.46 | 192.95 | 224.88 |
|  |  | $-\log _{10}$ PPNA | 89.81 | 44.33 | 1.81 | 24.66 | 57.32 | 87.91 | 116.22 | 169.95 | 194.44 | 226.37 |
|  |  | $-\log _{10}$ ASTpv | 30.90 | 11.07 | 5.90 | 14.24 | 22.88 | 30.45 | 37.58 | 51.09 | 55.65 | 65.78 |
|  | 0.1 | $\log _{10} \mathrm{BF}$ | 51.67 | 41.90 | -1.73 | 0.39 | 17.85 | 43.79 | 77.10 | 127.42 | 176.04 | 257.33 |
|  |  | $-\log _{10}$ PPNA | 53.16 | 41.89 | 0.20 | 1.88 | 19.34 | 45.28 | 78.59 | 128.91 | 177.53 | 258.83 |
|  |  | $-\log _{10}$ ASTpv | 12.96 | 5.06 | 1.70 | 5.38 | 9.37 | 12.67 | 16.02 | 21.63 | 25.04 | 37.98 |

Legend: $m$ - allele frequency at the risk SNP; $m=0.3,0.1$. The number of positively and negatively associated traits are denoted by $K_{1}^{+}$and $K_{1}^{-}$, respectively. Hence the total number of associated traits is $K_{1}=K_{1}^{+}+K_{1}^{-}$. The following abbreviations denote $-\log _{10} \mathrm{BF}: \log _{10}$ (Bayes factor), $-\log _{10} \mathrm{PPNA}:-\log _{10}$ (posterior probability of null association), $-\log _{10}(\mathrm{ASTpv}):-\log _{10}($ ASSET p-value). For multiple studies with no overlapping subjects, the uncorrelated version of CPBayes is implemented. Different summary measures obtained across 500 replications are provided: mean, standard deviation (sd), minimum (min), $5 \%$ quantile, $25 \%$ quantile, $50 \%$ quantile, $75 \%$ quantile, $95 \%$ quantile, $99 \%$ quantile, and maximum (max).

Table S5: Summary of measures for the evidence of the overall pleiotropic association when a subset of traits are associated for $\mathbf{1 0}$ non-overlapping case-control studies. Here $\mathbf{2}$ and $\mathbf{4}$ among $\mathbf{1 0}$ traits are associated.

| $K_{1}^{+}, K_{1}^{-}$ | $m$ |  | quantile |  |  |  |  |  |  |  |  | max |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | mean | sd | min | 5\% | 25\% | 50\% | 75\% | 95\% | 99\% |  |
| 2,0 | 0.3 | $\log _{10} \mathrm{BF}$ | 27.15 | 25.74 | -4.30 | -2.69 | 5.25 | 21.92 | 43.93 | 75.80 | 101.52 | 129.36 |
|  |  | - $\log _{10}$ PPNA | 30.19 | 25.70 | 0.02 | 0.49 | 8.26 | 24.93 | 46.94 | 78.81 | 104.53 | 132.37 |
|  |  | $-\log _{10}$ ASTpv | 13.69 | 7.69 | 0.30 | 2.76 | 7.79 | 12.76 | 18.96 | 27.45 | 33.79 | 41.97 |
|  | 0.1 | $\log _{10} \mathrm{BF}$ | 11.17 | 21.28 | -4.34 | -3.79 | -2.10 | 1.72 | 15.69 | 63.79 | 86.82 | 112.94 |
|  |  | - $\log _{10}$ PPNA | 14.30 | 21.20 | 0.02 | 0.07 | 0.96 | 4.73 | 18.70 | 66.80 | 89.83 | 115.95 |
|  |  | $-\log _{10}$ ASTpv | 5.54 | 3.90 | 0.00 | 0.79 | 2.57 | 4.72 | 7.63 | 13.72 | 16.25 | 19.99 |
| 1,1 | 0.3 | $\log _{10} \mathrm{BF}$ | 29.01 | 27.54 | -4.18 | -2.55 | 6.54 | 24.33 | 46.85 | 81.66 | 117.15 | 138.98 |
|  |  | $-\log _{10}$ PPNA | 32.06 | 27.50 | 0.03 | 0.59 | 9.55 | 27.34 | 49.86 | 84.67 | 120.16 | 141.99 |
|  |  | $-\log _{10} \mathrm{ASTpv}$ | 13.48 | 7.47 | 0.56 | 2.75 | 7.71 | 13.09 | 18.61 | 26.26 | 34.03 | 40.35 |
|  | 0.1 | $\log _{10} \mathrm{BF}$ | 11.73 | 21.45 | -4.25 | -3.85 | -2.22 | 2.19 | 18.41 | 63.34 | 88.81 | 122.15 |
|  |  | - $\log _{10}$ PPNA | 14.88 | 21.35 | 0.02 | 0.06 | 0.85 | 5.20 | 21.42 | 66.35 | 91.82 | 125.16 |
|  |  | $-\log _{10}$ ASTpv | 5.25 | 3.54 | 0.05 | 0.61 | 2.67 | 4.77 | 6.98 | 12.29 | 15.35 | 19.13 |
| 4,0 | 0.3 | $\log _{10} \mathrm{BF}$ | 89.89 | 47.02 | -3.04 | 17.36 | 56.24 | 88.18 | 117.85 | 174.97 | 214.41 | 257.88 |
|  |  | $-\log _{10}$ PPNA | 92.90 | 47.02 | 0.29 | 20.37 | 59.25 | 91.19 | 120.86 | 177.98 | 217.41 | 260.89 |
|  |  | $-\log _{10}$ ASTpv | 28.90 | 10.95 | 2.45 | 12.23 | 21.61 | 28.34 | 35.62 | 48.58 | 59.27 | 64.09 |
|  | 0.1 | $\log _{10} \mathrm{BF}$ | 48.12 | 44.63 | -3.91 | -2.02 | 9.87 | 38.46 | 72.24 | 133.95 | 162.01 | 208.74 |
|  |  | - $\log _{10}$ PPNA | 51.15 | 44.62 | 0.05 | 1.03 | 12.88 | 41.47 | 75.25 | 136.96 | 165.02 | 211.75 |
|  |  | $-\log _{10}$ ASTpv | 11.54 | 5.42 | 0.99 | 3.44 | 7.58 | 11.01 | 15.03 | 21.23 | 24.74 | 28.48 |
| 2,2 | 0.3 | $\log _{10} \mathrm{BF}$ | 85.65 | 45.24 | -3.28 | 14.94 | 52.11 | 84.36 | 117.50 | 161.07 | 188.45 | 227.38 |
|  |  | - $\log _{10}$ PPNA | 88.66 | 45.24 | 0.19 | 17.95 | 55.12 | 87.37 | 120.51 | 164.08 | 191.46 | 230.39 |
|  |  | $-\log _{10}$ ASTpv | 28.33 | 10.51 | 2.58 | 12.07 | 20.62 | 27.62 | 34.88 | 46.53 | 52.30 | 58.24 |
|  | 0.1 | $\log _{10} \mathrm{BF}$ | 43.43 | $44.04$ | $-4.13$ | $-2.32$ | $7.47$ | $32.08$ | 67.89 | $136.26$ | $172.62$ | $215.32$ |
|  |  | $-\log _{10} \mathrm{PPNA}$ | $46.45$ | 44.02 | 0.03 | 0.77 | 10.48 | 35.09 | 70.90 | 139.27 | 175.63 | 218.33 |
|  |  | $-\log _{10}$ ASTpv | 10.80 | 5.12 | 0.09 | 3.64 | 7.04 | 10.20 | 14.12 | 20.77 | 24.02 | 27.71 |

Legend: $m$ - allele frequency at the risk SNP; $m=0.3,0.1$. The number of positively and negatively associated traits are denoted by $K_{1}^{+}$and $K_{1}^{-}$, respectively. Hence the total number of associated traits is $K_{1}=K_{1}^{+}+K_{1}^{-}$. The following abbreviations denote $-\log _{10} \mathrm{BF}: \log _{10}$ (Bayes factor), $-\log _{10} \mathrm{PPNA}:-\log _{10}$ (posterior probability of null association), $-\log _{10}(\mathrm{ASTpv}):-\log _{10}($ ASSET p-value). For multiple studies with no overlapping subjects, the uncorrelated version of CPBayes is implemented. Different summary measures obtained across 500 replications are provided: mean, standard deviation (sd), minimum (min), $5 \%$ quantile, $25 \%$ quantile, $50 \%$ quantile, $75 \%$ quantile, $95 \%$ quantile, $99 \%$ quantile, and maximum (max).

Table S6: Summary of measures for the evidence of the overall pleiotropic association when a subset of traits are associated for $\mathbf{1 0}$ non-overlapping case-control studies. Here $\mathbf{6}$ and $\mathbf{8}$ among $\mathbf{1 0}$ traits are associated.


Legend: $m$ - allele frequency at the risk SNP; $m=0.3,0.1$. The number of positively and negatively associated traits are denoted by $K_{1}^{+}$and $K_{1}^{-}$, respectively. Hence the total number of associated traits is $K_{1}=K_{1}^{+}+K_{1}^{-}$. The following abbreviations denote $-\log _{10} \mathrm{BF}: \log _{10}$ (Bayes factor), $-\log _{10}$ PPNA: $-\log _{10}$ (posterior probability of null association), $-\log _{10}$ (ASTpv): - $\log _{10}$ (ASSET p-value). For multiple studies with no overlapping subjects, the uncorrelated version of CPBayes is implemented. Different summary measures obtained across 500 replications are provided: mean, standard deviation (sd), minimum (min), $5 \%$ quantile, $25 \%$ quantile, $50 \%$ quantile, $75 \%$ quantile, $95 \%$ quantile, $99 \%$ quantile, and maximum (max).

Table S7: Summary of measures for the evidence of the overall pleiotropic association when a subset of traits are associated for 15 non-overlapping case-control studies. Here 3, 6, and 9 among 15 traits are associated.

|  |  |  |  |  |  |  | quantile |  |  |  |  |
| :---: | :---: | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $K_{1}^{+}, K_{1}^{-}$ | $m$ |  |  |  |  |  |  |  |  |  |  |
| 3,0 | 0.3 | $\log _{10}$ BF | 55.17 | 41.68 | -5.35 | -0.51 | 23.33 | 51.13 | 81.53 | 132.41 | 168.25 |

Legend: $m$ - allele frequency at the risk SNP; $m=0.3,0.1$. The number of positively and negatively associated traits are denoted by $K_{1}^{+}$and $K_{1}^{-}$, respectively. Hence the total number of associated traits is $K_{1}=K_{1}^{+}+K_{1}^{-}$. The following abbreviations denote $-\log _{10} \mathrm{BF}: \log _{10}$ (Bayes factor), $-\log _{10} \mathrm{PPNA}:-\log _{10}$ (posterior probability of null association), $-\log _{10}(\mathrm{ASTpv}):-\log _{10}(\mathrm{ASSET}$ p-value). For multiple studies with no overlapping subjects, the uncorrelated version of CPBayes is implemented. Different summary measures obtained across 500 replications are provided: mean, standard deviation (sd), minimum (min), $5 \%$ quantile, $25 \%$ quantile, $50 \%$ quantile, $75 \%$ quantile, $95 \%$ quantile, $99 \%$ quantile, and maximum (max).

Table S8: Summary of measures for the evidence of the overall pleiotropic association when a subset of traits are associated for $\mathbf{5}$ overlapping case-control studies. Here $\mathbf{1}$ and $\mathbf{2}$ among $\mathbf{5}$ traits are associated.

| $K_{1}^{+}, K_{1}^{-}$ | $m$ | uncor\% |  | mean sd |  |  | quantile |  |  |  |  |  | max |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  | min | 5\% | 25\% | 50\% | 75\% | 95\% | 99\% |  |
| 1,0 | 0.3 | 0.4\% | $\log _{10} \mathrm{BF}$ | 6.95 | 10.32 | -3.10 | -2.61 | -0.84 | 3.21 | 11.81 | 28.04 | 37.84 | 58.48 |
|  |  |  | - $\log _{10}$ PPNA | 8.62 | 10.16 | 0.01 | 0.03 | 0.74 | 4.70 | 13.30 | 29.53 | 39.33 | 59.97 |
|  |  |  | - $\log _{10}$ ASTpv | 6.10 | 4.83 | 0.00 | 0.36 | 2.14 | 4.96 | 9.13 | 15.12 | 19.45 | 25.38 |
|  | 0.1 | 0.8\% | $\log _{10} \mathrm{BF}$ | 2.02 | 7.16 | -2.90 | -2.76 | -2.03 | -0.52 | 2.19 | 17.23 | 32.78 | 43.58 |
|  |  |  | - $\log _{10}$ PPNA | 3.86 | 6.96 | 0.02 | 0.02 | 0.11 | 1.02 | 3.68 | 18.72 | 34.27 | 45.07 |
|  |  |  | $-\log _{10}$ ASTpv | 2.58 | 2.39 | 0.00 | 0.11 | 0.69 | 1.87 | 3.75 | 7.36 | 9.46 | 14.57 |
| 2,0 | 0.3 | 0.6\% | $\log _{10} \mathrm{BF}$ | 21.94 | 21.25 | -2.68 | -1.15 | 4.03 | 16.66 | 34.84 | 61.40 | 89.90 | 101.77 |
|  |  |  | $-\log _{10}$ PPNA | 23.46 | 21.22 | 0.03 | 0.50 | 5.52 | 18.15 | 36.33 | 62.90 | 91.40 | 103.26 |
|  |  |  | - $\log _{10}$ ASTpv | 9.24 | 5.13 | 0.29 | 2.15 | 5.27 | 8.86 | 12.38 | 19.15 | 22.07 | 26.72 |
|  | 0.1 | 0.8\% | $\log _{10} \mathrm{BF}$ | 9.77 | 16.91 | -2.83 | -2.23 | -0.54 | 1.63 | 14.19 | 47.50 | 70.04 | 129.58 |
|  |  |  | $-\log _{10} \mathrm{PPNA}$ | 11.39 | 16.82 | 0.02 | 0.07 | 1.00 | 3.12 | 15.68 | 48.99 | 71.53 | 131.07 |
|  |  |  | - $\log _{10}$ ASTpv | 4.07 | 2.76 | 0.03 | 0.58 | 2.02 | 3.51 | 5.63 | 9.26 | 11.86 | 19.98 |
| 1,1 | 0.3 | 0.6\% | $\log _{10} \mathrm{BF}$ | 29.20 | $23.09$ | -2.27 | 0.44 | 11.06 | 25.75 | 43.96 | 74.31 | 96.00 | 115.74 |
|  |  |  | $-\log _{10} \mathrm{PPNA}$ | 30.70 | 23.07 | 0.07 | 1.94 | 12.55 | 27.24 | 45.45 | 75.80 | 97.49 | 117.23 |
|  |  |  | - $\log _{10}$ ASTpv | 11.97 | 6.28 | 0.67 | 3.05 | 7.22 | 11.58 | 16.14 | 23.73 | 27.73 | 32.93 |
|  | 0.1 | 4.6\% | $\log _{10} \mathrm{BF}$ | 16.22 | 18.96 | -2.66 | -1.77 | 0.50 | 9.03 | 26.47 | 51.73 | 75.19 | 95.58 |
|  |  |  | - $\log _{10}$ PPNA | 17.77 | 18.91 | 0.03 | 0.18 | 1.99 | 10.52 | 27.96 | 53.22 | 76.68 | 97.07 |
|  |  |  | $-\log _{10}$ ASTpv | 4.86 | 2.86 | 0.04 | 0.89 | 2.57 | 4.49 | 6.91 | 10.03 | 12.34 | 13.67 |

Legend: $m$ - allele frequency at the risk SNP; $m=0.3,0.1$. The number of positively and negatively associated traits are denoted by $K_{1}^{+}$and $K_{1}^{-}$, respectively. Hence the total number of associated traits is $K_{1}=K_{1}^{+}+K_{1}^{-}$. The following abbreviations denote $-\log _{10} \mathrm{BF}: \log _{10}$ (Bayes factor), $-\log _{10} \mathrm{PPNA}:-\log _{10}$ (posterior probability of null association), $-\log _{10}(\mathrm{ASTpv}):-\log _{10}$ (ASSET p-value). For multiple studies with overlapping subjects, the combined strategy of CPBayes is implemented. Of note, uncor\% denotes the percentage of replications in which the combined strategy of CPBayes used the uncorrelated version of CPBayes. Different summary measures obtained across 500 replications are provided: mean, standard deviation (sd), minimum (min), $5 \%$ quantile, $25 \%$ quantile, $50 \%$ quantile, $75 \%$ quantile, $95 \%$ quantile, $99 \%$ quantile, and maximum (max).

Table S9: Summary of measures for the evidence of the overall pleiotropic association when a subset of traits are associated for $\mathbf{5}$ overlapping case-control studies. Here $\mathbf{3}$ and $\mathbf{4}$ among $\mathbf{5}$ traits are associated.

| $K_{1}^{+}, K_{1}^{-}$ | $m$ | uncor\% |  | quantile |  |  |  |  |  |  |  |  | max |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  | mean | sd | min | 5\% | 25\% | 50\% | 75\% | 95\% | 99\% |  |
| 3,0 | 0.3 | 0.4\% | $\log _{10} \mathrm{BF}$ | 34.94 | 28.81 | -2.55 | 0.14 | 12.46 | 28.52 | 51.75 | 87.75 | 119.47 | 161.65 |
|  |  |  | - $\log _{10}$ PPNA | 36.45 | 28.80 | 0.04 | 1.64 | 13.95 | 30.01 | 53.24 | 89.24 | 120.96 | 163.15 |
|  |  |  | $-\log _{10}$ ASTpv | 11.64 | 5.40 | 0.78 | 3.71 | 7.81 | 10.96 | 15.17 | 21.38 | 25.35 | 31.84 |
|  | 0.1 | 0.8\% | $\log _{10} \mathrm{BF}$ | 15.32 | 22.10 | -2.73 | -2.04 | -0.10 | 5.92 | 23.74 | 60.50 | 88.52 | 139.69 |
|  |  |  | - $\log _{10}$ PPNA | 16.90 | 22.03 | 0.02 | 0.11 | 1.41 | 7.41 | 25.23 | 61.99 | 90.01 | 141.18 |
|  |  |  | $-\log _{10}$ ASTpv | 5.08 | 2.98 | 0.03 | 1.05 | 2.94 | 4.59 | 6.77 | 10.77 | 13.30 | 16.52 |
| 2,1 | 0.3 | 2.2\% | $\log _{10} \mathrm{BF}$ | 52.74 | 31.83 | -2.01 | 7.77 | 28.42 | 49.47 | 72.46 | 112.17 | 141.18 | 156.62 |
|  |  |  | $-\log _{10}$ PPNA | 54.24 | 31.83 | 0.12 | 9.27 | 29.91 | 50.96 | 73.95 | 113.66 | 142.67 | 158.11 |
|  |  |  | $-\log _{10}$ ASTpv | 15.68 | 6.63 | 1.21 | 5.67 | 10.92 | 15.00 | 20.00 | 27.83 | 33.05 | 35.93 |
|  | 0.1 | 5.8\% | $\log _{10} \mathrm{BF}$ | 32.79 | 29.50 | -2.78 | -0.93 | 6.50 | 26.35 | 52.58 | 88.01 | 113.65 | 122.83 |
|  |  |  | - $\log _{10}$ PPNA | 34.30 | 29.48 | 0.02 | 0.66 | 8.00 | 27.84 | 54.08 | 89.50 | 115.14 | 124.33 |
|  |  |  | $-\log _{10}$ ASTpv | 6.45 | 3.21 | 0.02 | 2.00 | 3.93 | 6.16 | 8.59 | 11.99 | 14.84 | 16.27 |
| 4,0 | 0.3 | 1 | $\log _{10} \mathrm{BF}$ | 40.37 | 32.14 | -2.32 | 0.74 | 14.01 | 35.31 | 59.80 | 98.97 | 127.94 | 169.53 |
|  |  |  | $-\log _{10}$ PPNA | 41.87 | 32.13 | 0.06 | 2.24 | 15.50 | 36.80 | 61.29 | 100.46 | 129.43 | 171.03 |
|  |  |  | $-\log _{10}$ ASTpv | 13.67 | 5.53 | 0.76 | 5.38 | 9.76 | 12.96 | 17.49 | 23.38 | 28.23 | 31.19 |
|  | 0.1 | 3.6 | $\log _{10} \mathrm{BF}$ | 14.98 | 24.90 | -2.65 | -1.94 | -0.29 | 3.55 | 21.02 | 69.20 | 105.87 | 151.60 |
|  |  |  | $-\log _{10}$ PPNA | 16.56 | 24.84 | 0.03 | 0.13 | 1.23 | 5.05 | 22.51 | 70.69 | 107.36 | 153.09 |
|  |  |  | $-\log _{10} \mathrm{ASTpv}$ | 5.94 | 3.19 | 0.35 | 1.54 | 3.61 | 5.56 | 7.85 | 12.20 | 14.84 | 17.87 |
| 2,2 | 0.3 | 1 | $\log _{10} \mathrm{BF}$ | 82.12 | 37.92 | -0.81 | 24.63 | 53.20 | 78.45 | 109.34 | 143.83 | 178.97 | 195.74 |
|  |  |  | $-\log _{10}$ PPNA | 83.61 | 37.92 | 0.76 | 26.12 | 54.70 | 79.95 | 110.83 | 145.32 | 180.46 | 197.24 |
|  |  |  | $-\log _{10}$ ASTpv | 19.07 | 6.56 | 2.14 | 8.70 | 14.05 | 18.80 | 23.69 | 30.08 | 33.76 | 37.17 |
|  | 0.1 | 6.8 | $\log _{10} \mathrm{BF}$ | 53.77 | 35.96 | -1.82 | 0.89 | 25.71 | 49.63 | 78.00 | 123.01 | 135.74 | 151.22 |
|  |  |  | $-\log _{10} \mathrm{PPNA}$ | 55.26 | 35.95 | 0.17 | 2.39 | 27.20 | 51.12 | 79.49 | 124.50 | 137.23 | 152.71 |
|  |  |  | $-\log _{10} \mathrm{ASTpv}$ | 7.74 | 3.03 | 0.80 | 3.17 | 5.54 | 7.29 | 9.91 | 12.92 | 15.16 | 16.92 |

Legend: $m$ - allele frequency at the risk SNP; $m=0.3,0.1$. The number of positively and negatively associated traits are denoted by $K_{1}^{+}$and $K_{1}^{-}$, respectively. Hence the total number of associated traits is $K_{1}=K_{1}^{+}+K_{1}^{-}$. The following abbreviations denote $-\log _{10} \mathrm{BF}: \log _{10}$ (Bayes factor), $-\log _{10} \mathrm{PPNA}:-\log _{10}$ (posterior probability of null association), $-\log _{10}(\mathrm{ASTpv}):-\log _{10}(A S S E T$ p-value). For multiple studies with overlapping subjects, the combined strategy of CPBayes is implemented. Of note, uncor\% denotes the percentage of replications in which the combined strategy of CPBayes used the uncorrelated version of CPBayes. Different summary measures obtained across 500 replications are provided: mean, standard deviation (sd), minimum (min), $5 \%$ quantile, $25 \%$ quantile, $50 \%$ quantile, $75 \%$ quantile, $95 \%$ quantile, $99 \%$ quantile, and maximum (max).

Table S10: Summary of measures for the evidence of the overall pleiotropic association when a subset of traits are associated for $\mathbf{1 0}$ overlapping case-control studies. Here $\mathbf{2}$ and $\mathbf{4}$ among 10 traits are associated.

| $K_{1}^{+}, K_{1}^{-}$ | $m$ | uncor\% |  | quantile |  |  |  |  |  |  |  |  | max |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  | mean | sd | min | 5\% | 25\% | 50\% | 75\% | 95\% | 99\% |  |
| 2,0 | 0.3 | 0.2\% | $\log _{10} \mathrm{BF}$ | 22.76 | 22.21 | -4.25 | -2.54 | 4.48 | 17.93 | 37.13 | 65.19 | 87.68 | 113.37 |
|  |  |  | $-\log _{10}$ PPNA | 25.80 | 22.17 | 0.02 | 0.60 | 7.49 | 20.94 | 40.14 | 68.20 | 90.69 | 116.38 |
|  |  |  | $-\log _{10}$ ASTpv | 7.93 | 4.91 | 0.00 | 0.99 | 4.12 | 7.38 | 11.04 | 16.83 | 21.79 | 26.42 |
|  | 0.1 | 2.4\% | $\log _{10} \mathrm{BF}$ | 9.70 | 17.03 | -4.32 | -3.86 | -2.07 | 3.30 | 15.21 | 47.44 | 72.03 | 93.28 |
|  |  |  | - $\log _{10}$ PPNA | 12.84 | 16.92 | 0.02 | 0.06 | 0.99 | 6.31 | 18.22 | 50.45 | 75.04 | 96.29 |
|  |  |  | $-\log _{10}$ ASTpv | 2.92 | 2.57 | 0.00 | 0.03 | 0.91 | 2.38 | 4.24 | 7.53 | 12.19 | 14.01 |
| 1,1 | 0.3 | $1 \%$ | $\log _{10} \mathrm{BF}$ | 27.39 | 22.87 | -4.08 | -2.22 | 8.91 | 23.86 | 41.70 | 71.44 | 87.53 | 117.80 |
|  |  |  | $-\log _{10}$ PPNA | 30.43 | 22.84 | 0.04 | 0.85 | 11.92 | 26.87 | 44.71 | 74.45 | 90.54 | 120.81 |
|  |  |  | - $\log _{10}{ }_{0}$ ASTpv | 10.07 | 6.08 | 0.02 | 1.18 | 5.12 | 9.55 | 13.97 | 21.60 | 24.71 | 29.81 |
|  | 0.1 | 3.8\% | $\log _{10} \mathrm{BF}$ | 13.86 | 19.58 | -4.39 | -3.75 | -1.16 | 6.53 | 22.99 | 52.39 | 82.73 | 105.61 |
|  |  |  | $-\log _{10}$ PPNA | 16.97 | 19.49 | 0.02 | 0.07 | 1.86 | 9.54 | 26.00 | 55.40 | 85.74 | 108.62 |
|  |  |  | $-\log _{10}$ ASTpv | 3.28 | 2.58 | 0.00 | 0.07 | 1.22 | 2.78 | 4.92 | 8.47 | 10.73 | 12.49 |
| 4,0 | 0.3 | 0.6\% | $\log _{10} \mathrm{BF}$ | 57.66 | 38.31 | -3.66 | 3.74 | 28.53 | 53.68 | 81.07 | 132.47 | 157.92 | 183.85 |
|  |  |  | $-\log _{10}$ PPNA | 60.67 | 38.31 | 0.09 | 6.75 | 31.54 | 56.69 | 84.08 | 135.48 | 160.93 | 186.85 |
|  |  |  | - $\log _{10}{ }^{\text {ASTpv }}$ | 11.73 | 5.22 | 0.34 | 4.46 | 8.02 | 11.03 | 14.89 | 21.50 | 25.96 | 30.74 |
|  | 0.1 | 3.6\% | $\log _{10} \mathrm{BF}$ | 25.97 | 31.75 | -3.81 | -2.76 | 0.36 | 14.82 | 40.68 | 93.18 | 122.71 | 161.92 |
|  |  |  | $-\log _{10}$ PPNA | 29.01 | 31.73 | 0.06 | 0.44 | 3.37 | 17.83 | 43.69 | 96.19 | 125.72 | 164.93 |
|  |  |  | $-\log _{10}$ ASTpv | 4.41 | 2.84 | 0.02 | 0.72 | 2.28 | 4.03 | 5.88 | 9.77 | 13.47 | 15.66 |
| 2,2 | 0.3 | 4.8\% | $\log _{10} \mathrm{BF}$ | 77.81 | 40.39 | -1.34 | 18.29 | 50.34 | 72.52 | 100.73 | 152.82 | 172.32 | 199.91 |
|  |  |  | $-\log _{10}$ PPNA | 80.82 | 40.39 | 1.68 | 21.30 | 53.35 | 75.53 | 103.74 | 155.83 | 175.33 | 202.92 |
|  |  |  | - $\log _{10}{ }^{\text {ASTpv }}$ | 16.80 | 6.35 | 1.63 | 6.71 | 12.27 | 16.17 | 21.39 | 28.00 | 30.75 | 32.92 |
|  | 0.1 | 12.4\% | $\log _{10} \mathrm{BF}$ | 46.82 | 36.71 | -3.32 | -1.54 | 18.64 | 39.66 | 70.42 | 114.56 | 141.42 | 179.66 |
|  |  |  | $-\log _{10} \mathrm{PPNA}$ | 49.84 | 36.69 | 0.17 | 1.48 | 21.65 | 42.67 | 73.43 | 117.57 | 144.43 | 182.67 |
|  |  |  | $-\log _{10}$ ASTpv | 5.88 | 2.76 | 0.08 | 1.56 | 3.76 | 5.71 | 7.80 | 10.59 | 12.33 | 14.16 |

Legend: $m$ - allele frequency at the risk SNP; $m=0.3,0.1$. The number of positively and negatively associated traits are denoted by $K_{1}^{+}$and $K_{1}^{-}$, respectively. Hence the total number of associated traits is $K_{1}=K_{1}^{+}+K_{1}^{-}$. The following abbreviations denote $-\log _{10} \mathrm{BF}: \log _{10}$ (Bayes factor), $-\log _{10} \mathrm{PPNA}:-\log _{10}$ (posterior probability of null association), $-\log _{10}(\mathrm{ASTpv}):-\log _{10}(\mathrm{ASSET}$ p-value). For multiple studies with overlapping subjects, the combined strategy of CPBayes is implemented. Of note, uncor\% denotes the percentage of replications in which the combined strategy of CPBayes used the uncorrelated version of CPBayes. Different summary measures obtained across 500 replications are provided: mean, standard deviation (sd), minimum (min), $5 \%$ quantile, $25 \%$ quantile, $50 \%$ quantile, $75 \%$ quantile, $95 \%$ quantile, $99 \%$ quantile, and maximum (max).

Table S11: Summary of measures for the evidence of the overall pleiotropic association when a subset of traits are associated for $\mathbf{1 0}$ overlapping case-control studies. Here $\mathbf{6}$ and $\mathbf{8}$ among $\mathbf{1 0}$ traits are associated.

| $K_{1}^{+}, K_{1}^{-}$ | $m$ | uncor\% |  | mean | sd | min | quantile |  |  |  |  |  | max |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  | 5\% |  |  | 25\% | 50\% | 75\% | 95\% | 99\% |  |
| 6,0 | 0.3 | 3\% | $\log _{10} \mathrm{BF}$ |  | 83.92 | 49.08 | -3.71 | 12.94 | 48.13 | 79.33 | 116.43 | 170.26 | 224.05 | 265.78 |
|  |  |  | $-\log _{10} \mathrm{PPNA}$ | 86.93 | 49.08 | 0.08 | 15.95 | 51.14 | 82.34 | 119.44 | 173.27 | 227.06 | 268.79 |
|  |  |  | $-\log _{10}$ ASTpv | 14.06 | 5.37 | 0.94 | 6.58 | 10.02 | 13.72 | 17.25 | 24.07 | 27.59 | 34.60 |
|  | 0.1 | 8.6\% | $\log _{10} \mathrm{BF}$ | 35.46 | 42.51 | -3.52 | -2.38 | 3.21 | 21.12 | 53.48 | 122.70 | 162.39 | 261.30 |
|  |  |  | $-\log _{10}$ PPNA | 38.48 | 42.50 | 0.12 | 0.72 | 6.22 | 24.13 | 56.49 | 125.71 | 165.40 | 264.31 |
|  |  |  | $-\log _{10}$ ASTpv | 5.41 | 3.03 | 0.03 | 1.52 | 3.07 | 4.91 | 7.07 | 11.41 | 14.02 | 17.92 |
| 3,3 | 0.3 | 6.8\% | $\log _{10} \mathrm{BF}$ | 139.07 | 53.94 | 8.81 | 55.16 | 101.18 | 135.69 | 173.47 | 230.87 | 289.54 | 300.51 |
|  |  |  | $-\log _{10}$ PPNA | 142.07 | 53.90 | 11.82 | 58.17 | 104.19 | 138.70 | 176.48 | 233.88 | 292.55 | 303.52 |
|  |  |  | $-\log _{10}$ ASTpv | 21.21 | 6.49 | 5.19 | 11.23 | 16.48 | 20.85 | 25.65 | 32.16 | 36.14 | 39.79 |
|  | 0.1 | 14.8\% | $\log _{10} \mathrm{BF}$ | 97.04 | 57.63 | -2.57 | 4.98 | 53.47 | 92.07 | 136.05 | 200.53 | 234.69 | 280.41 |
|  |  |  | $-\log _{10} \mathrm{PPNA}$ | 100.05 | 57.63 | 0.58 | 7.99 | 56.48 | 95.08 | 139.06 | 203.54 | 237.70 | 283.42 |
|  |  |  | $-\log _{10}$ ASTpv | 7.84 | 3.02 | 0.92 | 3.21 | 5.69 | 7.69 | 9.84 | 12.86 | 15.05 | 18.50 |
| 8,0 | 0.3 | 6.4\% | $\log _{10} \mathrm{BF}$ | 90.62 | 59.51 | -3.79 | 11.84 | 45.35 | 79.68 | 123.46 | 205.05 | 266.41 | 300.00 |
|  |  |  | - $\log _{10}$ PPNA | 93.63 | 59.49 | 0.07 | 14.85 | 48.36 | 82.69 | 126.47 | 208.06 | 269.42 | 300.00 |
|  |  |  | - $\log _{10}$ ASTpv | 16.28 | 5.76 | 0.77 | 8.38 | 12.13 | 15.56 | 19.73 | 26.55 | 32.47 | 34.89 |
|  | 0.1 | 21.8\% | $\log _{10} \mathrm{BF}$ | 39.87 | 52.80 | -3.99 | -2.85 | -0.56 | 19.04 | 65.18 | 143.11 | 222.08 | 301.12 |
|  |  |  | $-\log _{10}$ PPNA | 42.91 | 52.78 | 0.04 | 0.39 | 2.45 | 22.05 | 68.19 | 146.12 | 225.09 | 304.13 |
|  |  |  | $-\log _{10}$ ASTpv | 6.41 | 3.40 | 0.05 | 1.88 | 3.81 | 5.84 | 8.43 | 12.73 | 15.66 | 18.32 |
| 4,4 | 0.3 | 7\% | $\log _{10} \mathrm{BF}$ | 201.68 | 59.53 | 30.89 | 103.30 | 159.75 | 200.35 | 245.62 | 300.00 | 303.97 | 308.11 |
|  |  |  | $-\log _{10}$ PPNA | 204.45 | 59.13 | 33.90 | 106.31 | 162.76 | 203.36 | 248.63 | 300.00 | 303.90 | 307.17 |
|  |  |  | $-\log _{10} \mathrm{ASTpv}$ | 24.99 | 6.41 | 5.69 | 14.63 | 20.24 | 25.42 | 29.21 | 35.53 | 39.73 | 43.02 |
|  | 0.1 | 23.4\% | $\log _{10} \mathrm{BF}$ | 153.58 | 75.61 | -2.99 | 25.13 | 98.67 | 157.58 | 208.83 | 279.24 | 300.00 | 304.04 |
|  |  |  | $-\log _{10}$ PPNA | 156.52 | 75.48 | 0.31 | 28.14 | 101.68 | 160.59 | 211.84 | 282.25 | 300.00 | 307.05 |
|  |  |  | - $\log _{10}$ ASTpv | 9.49 | 3.05 | 1.02 | 4.55 | 7.33 | 9.46 | 11.69 | 14.54 | 16.51 | 17.63 |

Legend: $m$ - allele frequency at the risk SNP; $m=0.3,0.1$. The number of positively and negatively associated traits are denoted by $K_{1}^{+}$and $K_{1}^{-}$, respectively. Hence the total number of associated traits is $K_{1}=K_{1}^{+}+K_{1}^{-}$. The following abbreviations denote $-\log _{10} \mathrm{BF}: \log _{10}$ (Bayes factor), $-\log _{10} \mathrm{PPNA}:-\log _{10}$ (posterior probability of null association), $-\log _{10}(\mathrm{ASTpv}):-\log _{10}(\mathrm{ASSET}$ p-value). For multiple studies with overlapping subjects, the combined strategy of CPBayes is implemented. Of note, uncor\% denotes the percentage of replications in which the combined strategy of CPBayes used the uncorrelated version of CPBayes. Different summary measures obtained across 500 replications are provided: mean, standard deviation (sd), minimum (min), $5 \%$ quantile, $25 \%$ quantile, $50 \%$ quantile, $75 \%$ quantile, $95 \%$ quantile, $99 \%$ quantile, and maximum (max).

Table S12: Summary of measures for the evidence of the overall pleiotropic association when a subset of traits are associated for $\mathbf{1 5}$ overlapping case-control studies. Here 3,6, and 9 among 15 traits are associated.


Legend: $m$ - allele frequency at the risk SNP; $m=0.3,0.1$. The number of positively and negatively associated traits are denoted by $K_{1}^{+}$and $K_{1}^{-}$, respectively. Hence the total number of associated traits is $K_{1}=K_{1}^{+}+K_{1}^{-}$. The following abbreviations denote $-\log _{10} \mathrm{BF}: \log _{10}$ (Bayes factor), $-\log _{10} \mathrm{PPNA}:-\log _{10}$ (posterior probability of null association), $-\log _{10}(\mathrm{ASTpv}):-\log _{10}($ ASSET p-value). For multiple studies with overlapping subjects, the combined strategy of CPBayes is implemented. Of note, uncor\% denotes the percentage of replications in which the combined strategy of CPBayes used the uncorrelated version of CPBayes. Different summary measures obtained across 500 replications are provided: mean, standard deviation (sd), minimum (min), $5 \%$ quantile, $25 \%$ quantile, $50 \%$ quantile, $75 \%$ quantile, $95 \%$ quantile, $99 \%$ quantile, and maximum (max).

Table S13: Summary of measures for the evidence of the overall pleiotropic association for a cohort study with 5 binary phenotypes. Here 1, 2, and 3 among 5 traits are associated.


Legend: $m$ - allele frequency at the risk SNP; $m=0.3,0.1$. The number of positively and negatively associated traits are denoted by $K_{1}^{+}$and $K_{1}^{-}$, respectively. Hence the total number of associated traits is $K_{1}=K_{1}^{+}+K_{1}^{-}$. The following abbreviations denote $-\log _{10} \mathrm{BF}: \log _{10}$ (Bayes factor), $-\log _{10}$ PPNA: $-\log _{10}$ (posterior probability of null association), $-\log _{10}(\mathrm{ASTpv}):-\log _{10}$ (ASSET p-value). For a cohort study, since the summary statistics are correlated, the combined strategy of CPBayes is implemented. Of note, uncor\% denotes the percentage of replications in which the combined strategy of CPBayes used the uncorrelated version of CPBayes. Different summary measures obtained across 500 replications are provided: mean, standard deviation (sd), minimum (min), $5 \%$ quantile, $25 \%$ quantile, $50 \%$ quantile, $75 \%$ quantile, $95 \%$ quantile, $99 \%$ quantile, and maximum (max).

Table S14: Summary of measures for the evidence of the overall pleiotropic association for a cohort study with 10 binary phenotypes. Here 2, 4, and 6 among 10 traits are associated.

| $K_{1}^{+}, K_{1}^{-}$ | $m$ | uncor\% |  | mean | sd | min | quantile |  |  |  |  |  | max |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  |  | 5\% | 25\% | 50\% | 75\% | 95\% | 99\% |  |
| 2,0 | 0.3 | 1.4\% | $\log _{10} \mathrm{BF}$ | 15.43 | 22.86 | -3.84 | -2.83 | -0.96 | 6.94 | 23.85 | 64.26 | 94.62 | 141.25 |
|  |  |  | $-\log _{10}$ PPNA | 18.47 | 22.83 | 0.06 | 0.40 | 2.05 | 9.95 | 26.86 | 67.27 | 97.63 | 144.26 |
|  |  |  | - $\log _{10}$ ASTpv | 4.42 | 2.22 | 0.21 | 1.32 | 2.82 | 4.09 | 5.69 | 8.65 | 9.98 | 14.28 |
|  | 0.1 | 2\% | $\log _{10} \mathrm{BF}$ | 39.56 | 52.79 | -3.65 | -2.65 | -0.73 | 17.77 | 61.14 | 150.17 | 208.91 | 300.00 |
|  |  |  | - $\log _{10}$ PPNA | 42.59 | 52.74 | 0.09 | 0.52 | 2.28 | 20.78 | 64.15 | 153.18 | 211.92 | 300.00 |
|  |  |  | - $\log _{10}$ ASTpv | 3.83 | 1.94 | 0.32 | 0.90 | 2.47 | 3.54 | 5.04 | 7.19 | 9.34 | 10.46 |
| 1,1 | 0.3 | 2.4\% | $\log _{10} \mathrm{BF}$ | 25.82 | 24.81 | -3.56 | -1.85 | 5.20 | 19.64 | 39.79 | 74.28 | 98.89 | 120.00 |
|  |  |  | $-\log _{10}$ PPNA | 28.84 | 24.80 | 0.11 | 1.18 | 8.21 | 22.65 | 42.80 | 77.29 | 101.90 | 123.01 |
|  |  |  | $-\log _{10}$ ASTpv | 4.88 | 2.42 | 0.18 | 1.58 | 3.06 | 4.63 | 6.24 | 9.40 | 11.67 | 15.97 |
|  | 0.1 | 2.2\% | $\log _{10} \mathrm{BF}$ | 53.73 | 49.08 | -3.88 | -0.92 | 17.83 | 40.63 | 76.23 | 145.13 | 224.11 | 306.52 |
|  |  |  | $-\log _{10}$ PPNA | 56.72 | 48.95 | 0.06 | 2.10 | 20.84 | 43.64 | 79.24 | 148.14 | 227.12 | 305.48 |
|  |  |  | $-\log _{10}$ ASTpv | 4.93 | 2.41 | 0.06 | 1.74 | 3.25 | 4.56 | 6.36 | 9.42 | 11.40 | 17.14 |
| 4,0 | 0.3 | 2.2\% | $\log _{10} \mathrm{BF}$ | 56.33 | 54.33 | -3.51 | -1.89 | 6.75 | 46.66 | 86.96 | 169.80 | 215.64 | 244.32 |
|  |  |  | $-\log _{10}$ PPNA | 59.35 | 54.32 | 0.12 | 1.15 | 9.76 | 49.67 | 89.97 | 172.81 | 218.65 | 247.33 |
|  |  |  | $-\log _{10}$ ASTpv | 8.47 | 3.36 | 0.93 | 3.89 | 5.94 | 7.89 | 10.38 | 14.76 | 18.33 | 20.98 |
|  | 0.1 | 3.8\% | $\log _{10} \mathrm{BF}$ | 119.80 | 100.88 | -3.34 | -1.95 | 23.09 | 104.41 | 193.24 | 300.00 | 304.52 | 308.12 |
|  |  |  | $-\log _{10}$ PPNA | 122.52 | 100.33 | 0.17 | 1.09 | 26.10 | 107.42 | 196.25 | 300.00 | 300.01 | 308.03 |
|  |  |  | - $\log _{10}$ ASTpv | 7.23 | 2.92 | 1.29 | 3.07 | 5.24 | 6.92 | 8.79 | 12.70 | 15.48 | 17.78 |
| 2,2 | 0.3 | 3.8\% | $\log _{10} \mathrm{BF}$ | 99.93 | 46.20 | -2.36 | 26.66 | 66.67 | 99.40 | 130.90 | 176.83 | 216.07 | 254.43 |
|  |  |  | $-\log _{10}$ PPNA | 102.94 | 46.20 | 0.74 | 29.67 | 69.68 | 102.41 | 133.91 | 179.84 | 219.08 | 257.44 |
|  |  |  | - $\log _{10}$ ASTpv | 10.42 | 2.98 | 3.39 | 5.87 | 8.26 | 10.36 | 12.27 | 15.42 | 18.26 | 21.29 |
|  | 0.1 | 4.4\% | $\log _{10} \mathrm{BF}$ | 227.05 | 78.66 | -2.35 | 71.00 | 171.94 | 249.85 | 300.00 | 300.00 | 303.33 | 307.17 |
|  |  |  | $-\log _{10}$ PPNA | 229.08 | 77.75 | 0.75 | 74.01 | 174.95 | 252.86 | 300.00 | 300.00 | 302.47 | 306.32 |
|  |  |  | - $\log _{10}$ ASTpv | 10.63 | 3.30 | 1.40 | 5.53 | 8.25 | 10.43 | 12.64 | 16.28 | 18.50 | 26.30 |
| 6,0 | 0.3 | 5.6\% | $\log _{10} \mathrm{BF}$ | 88.10 | 77.83 | -3.57 | -1.81 | 17.54 | 73.16 | 144.29 | 232.21 | 300.00 | 302.61 |
|  |  |  | $-\log _{10}$ PPNA | 91.09 | 77.73 | 0.11 | 1.23 | 20.55 | 76.17 | 147.30 | 235.22 | 300.00 | 305.62 |
|  |  |  | - $\log _{10}{ }^{\text {ASTpv}}$ | 11.59 | 4.12 | 1.83 | 5.67 | 8.64 | 11.23 | 13.91 | 19.29 | 22.88 | 29.19 |
|  | 0.1 | 4.8\% | $\log _{10} \mathrm{BF}$ | 174.39 | 115.88 | -3.43 | -1.72 | 66.20 | 181.02 | 300.00 | 300.00 | 301.73 | 306.12 |
|  |  |  | $-\log _{10}$ PPNA | 176.49 | 114.88 | 0.14 | 1.31 | 69.20 | 184.03 | 300.00 | 300.00 | 303.51 | 307.87 |
|  |  |  | $-\log _{10}$ ASTpv | 10.22 | 3.52 | 1.83 | 5.29 | 7.76 | 9.89 | 12.26 | 16.74 | 20.79 | 23.54 |
| 3,3 | 0.3 | 5.2\% | $\log _{10} \mathrm{BF}$ | 202.10 | 57.94 | 32.55 | 100.05 | 165.63 | 204.42 | 245.35 | 300.00 | 302.67 | 305.65 |
|  |  |  | $-\log _{10}$ PPNA | 204.94 | 57.66 | 35.56 | 103.06 | 168.64 | 207.43 | 248.36 | 300.00 | 302.39 | 307.74 |
|  |  |  | - $\log _{10}$ ASTpv | 15.67 | 3.58 | 6.08 | 10.05 | 13.19 | 15.49 | 17.99 | 21.41 | 25.39 | 27.98 |
|  | 0.1 | 4\% | $\log _{10} \mathrm{BF}$ | 295.46 | 22.55 | 93.92 | 277.38 | 300.00 | 300.00 | 300.00 | 300.00 | 300.00 | 305.60 |
|  |  |  | $-\log _{10}$ PPNA | 295.64 | 21.98 | 96.93 | 280.39 | 300.00 | 300.00 | 300.00 | 300.00 | 300.00 | 304.24 |
|  |  |  | - $\log _{10}$ ASTpv | 15.89 | 3.57 | 7.36 | 10.46 | 13.71 | 15.67 | 18.22 | 21.72 | 24.68 | 32.26 |

Legend: $m$ - allele frequency at the risk SNP; $m=0.3,0.1$. The number of positively and negatively associated traits are denoted by $K_{1}^{+}$and $K_{1}^{-}$, respectively. Hence the total number of associated traits is $K_{1}=K_{1}^{+}+K_{1}^{-}$. The following abbreviations denote $-\log _{10} \mathrm{BF}: \log _{10}$ (Bayes factor), $-\log _{10} \mathrm{PPNA}:-\log _{10}$ (posterior probability of null association), $-\log _{10}(\mathrm{ASTpv}):-\log _{10}(\mathrm{ASSET}$ p-value). For a cohort study, since the summary statistics are correlated, the combined strategy of CPBayes is implemented. Of note, uncor\% denotes the percentage of replications in which the combined strategy of CPBayes used the uncorrelated version of CPBayes. Different summary measures obtained across 500 replications are provided: mean, standard deviation (sd), minimum (min), $5 \%$ quantile, $25 \%$ quantile, $50 \%$ quantile, $75 \%$ quantile, $95 \%$ quantile, $99 \%$ quantile, and maximum (max).

Table S15: Summary of measures for the evidence of the overall pleiotropic association for a cohort study with 15 binary phenotypes. Here 3, 6, and 9 among 15 traits are associated.

| $K_{1}^{+}, K_{1}^{-}$ | $m$ | uncor\% |  | mean | sd | min | quantile |  |  |  |  |  | max |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  |  | 5\% | 25\% | 50\% | 75\% | 95\% | 99\% |  |
| 3,0 | 0.3 | 2.8\% | $\log _{10} \mathrm{BF}$ | 35.52 | 38.84 | -5.03 | -3.62 | 3.13 | 23.84 | 59.72 | 109.93 | 156.95 | 206.75 |
|  |  |  | $-\log _{10}$ PPNA | 40.05 | 38.82 | 0.12 | 0.95 | 7.65 | 28.35 | 64.24 | 114.45 | 161.46 | 211.26 |
|  |  | 1.6\% | $-\log _{10}$ ASTpv | 5.42 | 2.53 | 0.36 | 1.89 | 3.65 | 5.07 | 6.95 | 9.68 | 12.78 | 16.67 |
|  | 0.1 |  | $\log _{10} \mathrm{BF}$ | 91.59 | 85.98 | -5.06 | -3.10 | 10.53 | 79.47 | 148.50 | 259.71 | 300.00 | 305.64 |
|  |  |  | - $\log _{10}$ PPNA | 96.01 | 85.70 | 0.11 | 1.43 | 15.05 | 83.99 | 153.02 | 264.22 | 300.00 | 307.76 |
|  |  |  | - $\log _{10}{ }_{0}$ ASTpv | 4.68 | 2.23 | 0.01 | 1.69 | 3.04 | 4.40 | 5.98 | 8.61 | 11.66 | 13.06 |
| 2,1 | 0.3 | 4.8\% | $\log _{10} \mathrm{BF}$ | 50.99 | 38.85 | -4.91 | -2.40 | 17.71 | 45.66 | 75.27 | 119.05 | 157.58 | 188.70 |
|  |  |  | $-\log _{10}$ PPNA | 55.51 | 38.85 | 0.15 | 2.12 | 22.23 | 50.17 | 79.78 | 123.56 | 162.09 | 193.22 |
|  |  |  | $-\log _{10}$ ASTpv | 5.58 | 2.48 | 0.41 | 2.03 | 3.79 | 5.37 | 6.98 | 10.15 | 12.31 | 14.73 |
|  | 0.1 | $4 \%$ | $\log _{10} \mathrm{BF}$ | 127.84 | 80.25 | -4.47 | 7.71 | 63.09 | 123.28 | 182.87 | 276.54 | 300.00 | 306.77 |
|  |  |  | $-\log _{10}$ PPNA | 132.20 | 79.92 | 0.33 | 12.23 | 67.60 | 127.79 | 187.39 | 281.05 | 300.00 | 305.14 |
|  |  |  | $-\log _{10}{ }^{\text {ASTpv }}$ | 5.40 | 2.37 | 0.01 | 1.74 | 3.64 | 5.31 | 6.91 | 9.23 | 11.28 | 15.27 |
| 6,0 | 0.3 | 4.8\% | $\log _{10} \mathrm{BF}$ | 103.39 | 80.43 | -4.47 | -2.86 | 36.27 | 95.49 | 154.95 | 250.77 | 294.16 | 300.00 |
|  |  |  | $-\log _{10}$ PPNA | 107.88 | 80.34 | 0.32 | 1.67 | 40.78 | 100.00 | 159.46 | 255.29 | 298.67 | 300.35 |
|  |  |  | - $\log _{10}{ }^{\text {ASTpv }}$ | 9.90 | 3.53 | 2.06 | 4.70 | 7.34 | 9.65 | 11.96 | 16.32 | 19.38 | 22.77 |
|  | 0.1 | 5.8\% | $\log _{10} \mathrm{BF}$ | 197.13 | 110.97 | -4.91 | -2.45 | 99.53 | 236.02 | 300.00 | 300.00 | 300.00 | 307.61 |
|  |  |  | - $\log _{10}$ PPNA | 199.83 | 109.28 | 0.15 | 2.07 | 104.04 | 240.53 | 300.00 | 300.00 | 302.37 | 304.28 |
|  |  |  | - $\log _{10}{ }_{0}$ ASTpv | 8.57 | 3.06 | 0.76 | 4.22 | 6.37 | 8.37 | 10.34 | 13.81 | 17.21 | 21.96 |
| 3,3 | 0.3 | 4\% | $\log _{10} \mathrm{BF}$ | 190.67 | 61.02 | 14.94 | 88.20 | 152.24 | 192.81 | 231.43 | 300.00 | 300.00 | 304.60 |
|  |  |  | $-\log _{10}$ PPNA | 194.93 | 60.57 | 19.45 | 92.71 | 156.75 | 197.33 | 235.94 | 300.00 | 300.00 | 306.52 |
|  |  |  | - $\log _{10}{ }^{\text {ASTpv }}$ | 13.37 | 3.51 | 4.33 | 7.93 | 11.07 | 13.20 | 15.49 | 19.59 | 22.10 | 25.72 |
|  | 0.1 | 7.6\% | $\log _{10} \mathrm{BF}$ | 292.98 | 28.66 | 10.06 | 249.35 | 300.00 | 300.00 | 300.00 | 300.00 | 301.12 | 307.81 |
|  |  |  | - $\log _{10}$ PPNA | 293.42 | 27.66 | 14.58 | 253.87 | 300.00 | 300.00 | 300.00 | 300.00 | 300.00 | 307.89 |
|  |  |  | $-\log _{10}{ }^{\text {ASTpv }}$ | 13.49 | 3.58 | 3.94 | 8.35 | 10.89 | 13.18 | 15.67 | 19.83 | 22.81 | 25.69 |
| 9,0 | 0.3 | 8.2\% | $\log _{10} \mathrm{BF}$ | 137.39 | 110.59 | -4.67 | -3.30 | 25.92 | 131.91 | 236.19 | 300.00 | 302.61 | 307.78 |
|  |  |  | $-\log _{10}$ PPNA | 141.24 | 109.60 | 0.23 | 1.25 | 30.44 | 136.43 | 240.70 | 300.00 | 303.51 | 308.00 |
|  |  |  | $-\log _{10} \mathrm{ASTpv}^{2}$ | 13.51 | 4.81 | 3.76 | 7.28 | 9.76 | 12.87 | 16.43 | 22.69 | 27.08 | 34.57 |
|  | 0.1 | 8.4\% | $\log _{10} \mathrm{BF}$ | 216.47 | 113.62 | -4.69 | -2.90 | 120.36 | 300.00 | 300.00 | 300.00 | 300.00 | 305.19 |
|  |  |  | $-\log _{10}$ PPNA | 218.47 | 111.75 | 0.22 | 1.62 | 124.87 | 300.00 | 300.00 | 300.00 | 300.01 | 301.90 |
|  |  |  | - $\log _{10}{ }_{0}$ ASTpv | 11.88 | 4.25 | 3.98 | 5.69 | 8.93 | 11.44 | 14.25 | 19.82 | 23.76 | 30.41 |
| 5,4 | 0.3 | 7.6\% | $\log _{10} \mathrm{BF}$ | 287.14 | 31.28 | 87.96 | 215.08 | 298.03 | 300.00 | 300.00 | 300.00 | 303.78 | 308.20 |
|  |  |  | - $\log _{10}$ PPNA | 288.29 | 29.91 | 92.48 | 219.60 | 300.00 | 300.00 | 300.00 | 300.00 | 304.74 | 307.64 |
|  |  |  | - $\log _{10}{ }_{0}$ ASTpv | 19.61 | 3.85 | 8.76 | 13.22 | 17.16 | 19.35 | 22.16 | 25.80 | 29.35 | 33.71 |
|  | 0.1 | 9.2\% | $\log _{10} \mathrm{BF}$ | 300.00 | 0.16 | 297.25 | 300.00 | 300.00 | 300.00 | 300.00 | 300.00 | 300.00 | 302.15 |
|  |  |  | $-\log _{10}$ PPNA | 300.02 | 0.35 | 300.00 | 300.00 | 300.00 | 300.00 | 300.00 | 300.00 | 300.00 | 306.66 |
|  |  |  | $-\log _{10}$ ASTpv | 19.41 | 3.96 | 9.27 | 13.56 | 16.56 | 19.19 | 21.81 | 26.19 | 30.74 | 31.58 |

Legend: $m$ - allele frequency at the risk SNP; $m=0.3,0.1$. The number of positively and negatively associated traits are denoted by $K_{1}^{+}$and $K_{1}^{-}$, respectively. Hence the total number of associated traits is $K_{1}=K_{1}^{+}+K_{1}^{-}$. The following abbreviations denote $-\log _{10} \mathrm{BF}: \log _{10}$ (Bayes factor), $-\log _{10} \mathrm{PPNA}:-\log _{10}$ (posterior probability of null association), $-\log _{10}(\mathrm{ASTpv}):-\log _{10}(\mathrm{ASSET}$ p-value). For a cohort study, since the summary statistics are correlated, the combined strategy of CPBayes is implemented. Of note, uncor\% denotes the percentage of replications in which the combined strategy of CPBayes used the uncorrelated version of CPBayes. Different summary measures obtained across 500 replications are provided: mean, standard deviation (sd), minimum (min), $5 \%$ quantile, $25 \%$ quantile, $50 \%$ quantile, $75 \%$ quantile, $95 \%$ quantile, $99 \%$ quantile, and maximum (max).

Figure S2: Comparison of the accuracy of selection of associated traits by CPBayes and ASSET for cohort study. The total number of phenotypes/studies is denoted by $K$ and $m$ denotes the minor allele frequency at the risk SNP.


Table S16: Simulation study for $\mathbf{5 0}$ traits. Summary of measures for the evidence of the overall pleiotropic association for 50 overlapping case-control studies. Here 0,5 , and 10 among 50 traits are associated.

| $K_{1}^{+}, K_{1}^{-}$ | $m$ |  | mean | sd | min | quantile |  |  |  |  |  | max |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | 5\% |  |  | 25\% | 50\% | 75\% | 95\% | 99\% |  |
| 0,0 | 0.3 | $\log _{10} \mathrm{BF}$ |  | -16.43 | 0.11 | -16.61 | -16.56 | -16.50 | -16.44 | -16.38 | -16.25 | -16.08 | -15.83 |
|  |  | $-\log _{10}$ PPNA | 0.02 | 0.006 | 0.01 | 0.01 | 0.02 | 0.02 | 0.02 | 0.03 | 0.04 | 0.07 |
|  | 0.1 | $\log _{10} \mathrm{BF}$ | -16.18 | 0.25 | -16.42 | -16.38 | -16.30 | -16.24 | -16.15 | -15.83 | -15.51 | -13.96 |
|  |  | $-\log _{10}$ PPNA | 0.04 | 0.08 | 0.02 | 0.02 | 0.02 | 0.03 | 0.03 | 0.07 | 0.13 | 1.12 |
| 5,0 | 0.3 | $\log _{10} \mathrm{BF}$ | 80.75 | 49.38 | -10.03 | 9.21 | 45.43 | 74.39 | 107.94 | 170.54 | 216.63 | 240.42 |
|  |  | $-\log _{10}$ PPNA | 95.80 | 49.38 | 5.02 | 24.26 | 60.48 | 89.44 | 123.00 | 185.60 | 231.68 | 255.47 |
|  | 0.1 | $\log _{10} \mathrm{BF}$ | 44.74 | 48.08 | -15.93 | -14.86 | 1.51 | 38.10 | 74.46 | 124.66 | 164.88 | 205.38 |
|  |  | $-\log _{10}$ PPNA | 59.82 | 48.05 | 0.05 | 0.41 | 16.56 | 53.15 | 89.51 | 139.71 | 179.94 | 220.43 |
| 3,2 | 0.3 | $\log _{10} \mathrm{BF}$ | 90.92 | 52.29 | -8.18 | 19.74 | 51.37 | 85.36 | 120.45 | 192.00 | 224.04 | 235.20 |
|  |  | $-\log _{10}$ PPNA | 105.97 | 52.29 | 6.87 | 34.80 | 66.42 | 100.41 | 135.50 | 207.05 | 239.10 | 250.25 |
|  | 0.1 | $\log _{10} \mathrm{BF}$ | 49.80 | $55.62$ | $-15.27$ | -14.33 | $3.45$ | $34.91$ | 81.20 | 161.52 | 196.32 | $221.93$ |
|  |  | $-\log _{10} \mathrm{PPNA}$ | $64.86$ | $55.60$ | $0.21$ | 0.80 | 18.51 | 49.96 | 96.25 | 176.58 | 211.37 | $236.98$ |
| 10,0 | 0.3 | $\log _{10} \mathrm{BF}$ | 199.00 | 67.19 | 25.03 | 96.56 | 146.74 | 201.41 | 249.66 | 300.00 | 300.01 | 307.05 |
|  |  | $-\log _{10} \mathrm{PPNA}$ | 212.35 | 64.74 | 40.08 | 111.61 | 161.79 | 216.46 | 264.71 | 300.00 | 305.53 | 306.50 |
|  | 0.1 | $\log _{10} \mathrm{BF}$ | 130.40 | 89.63 | -15.33 | -12.20 | 58.70 | 128.76 | 189.76 | 285.52 | 300.02 | 306.13 |
|  |  | $-\log _{10}$ PPNA | 144.74 | 88.29 | 0.18 | 2.85 | 73.75 | 143.81 | 204.81 | 300.00 | 300.16 | 305.42 |
| 5,5 | 0.3 | $\log _{10} \mathrm{BF}$ | 226.35 | 61.43 | 86.31 | 122.24 | 179.19 | 232.38 | 288.33 | 300.00 | 303.86 | 307.56 |
|  |  | $-\log _{10}$ PPNA | 237.90 | 57.39 | 101.36 | 137.29 | 194.24 | 247.43 | 300.00 | 300.00 | 305.03 | 306.94 |
|  | 0.1 | $\log _{10} \mathrm{BF}$ | 159.54 | 83.40 | -14.64 | 21.29 | 98.02 | 158.45 | 222.71 | 300.00 | 301.30 | 308.25 |
|  |  | $-\log _{10}$ PPNA | 173.33 | 81.33 | 0.56 | 36.34 | 113.07 | 173.50 | 237.76 | 300.00 | 302.45 | 305.23 |

Legend: $m$ - allele frequency at the risk SNP; $m=0.3,0.1$. The number of positively and negatively associated traits are denoted by $K_{1}^{+}$and $K_{1}^{-}$, respectively. Hence the total number of associated traits is $K_{1}=K_{1}^{+}+K_{1}^{-}$. The following abbreviations denote $-\log _{10} \mathrm{BF}: \log _{10}$ (Bayes factor), $-\log _{10}$ PPNA: $-\log _{10}$ (posterior probability of null association). For multiple studies with overlapping subjects, the combined strategy of CPBayes is implemented. Different summary measures obtained across 200 replications are provided: mean, standard deviation (sd), minimum (min), $5 \%$ quantile, $25 \%$ quantile, $50 \%$ quantile, $75 \%$ quantile, $95 \%$ quantile, $99 \%$ quantile, and maximum (max).

Bayesian meta-analysis for studying cross-phenotype genetic associations

Table S17: Simulation study for 50 traits. Accuracy in selection of associated traits by CPBayes for 50 overlapping case-control studies. The number of positively and negatively associated traits are denoted by $K_{1}^{+}$and $K_{1}^{-}$, respectively. And $m$ denotes the minor allele frequency at the risk SNP.

| $K_{1}^{+}, K_{1}^{-}$ | $m$ | specificity | sensitivity |
| :--- | :--- | :--- | :--- |
| 5,0 | 0.3 | 1.00 | 0.81 |
|  | 0.1 | 1.00 | 0.61 |
| 3,2 | 0.3 | 1.00 | 0.81 |
|  | 0.1 | 1.00 | 0.59 |
| 10,0 | 0.3 | 1.00 | 0.83 |
|  | 0.1 | 1.00 | 0.62 |
| 5,5 | 0.3 | 1.00 | 0.82 |
|  | 0.1 | 1.00 | 0.62 |

Bayesian meta-analysis for studying cross-phenotype genetic associations

Table S18: Name of 22 phenotypes in the Kaiser GERA cohort analyzed by CPBayes and ASSET.

|  | GERA cohort phenotypes |  |  |
| :--- | :--- | :--- | :--- |
| Asthma | Type 2 Diabetes | Iron Deficiency | Peptic Ulcer |
| Allergic Rhinitis | Dyslipidemia | Irritable Bowel Syndrome | Psychiatric disorders |
| Cardiovascular Disease | Hypertension | Macular Degeneration | Stress Disorders |
| Cancers | Hemorrhoids | Osteoarthritis | Varicose Veins |
| Depressive Disorder | Abdominal Hernia | Osteoporosis |  |
| Dermatophytosis | Insomnia | Peripheral Vascular Disease |  |

Table S19: Independent pleiotropic SNPs identified by CPBayes for which one phenotype was selected.

| rsID | chrom <br> band | CPBayes $\log _{10} \mathrm{BF}$ | $\begin{aligned} & \text { CPBayes } \\ & \text { PPNA } \end{aligned}$ | Associated trait | $\mathrm{PPA}_{j}$ | Direction | Univariate p-values |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| rs2300430 | 1q31.3 | 300.00 | $1.00 \mathrm{E}-300$ | Macular Degeneration | 100\% | negative | $2.57 \mathrm{E}-77$ |
| rs35505017 | 1 q 31.3 | 213.91 | $2.92 \mathrm{E}-221$ | Macular Degeneration | 100\% | negative | $3.14 \mathrm{E}-14$ |
| rs77394225 | 1 q 31.3 | 76.74 | $4.33 \mathrm{E}-84$ | Macular Degeneration | 100\% | negative | $6.04 \mathrm{E}-16$ |
| rs4658043 | 1 q 31.3 | 15.56 | $6.62 \mathrm{E}-23$ | Macular Degeneration | 100\% | negative | $4.61 \mathrm{E}-08$ |
| rs77795056 | 1p36.23 | 9.50 | $7.47 \mathrm{E}-17$ | Iron Deficiency | 100\% | negative | $5.13 \mathrm{E}-05$ |
| rs79032519 | 1q43 | 8.80 | $3.76 \mathrm{E}-16$ | Peptic Ulcer | 100\% | negative | $2.14 \mathrm{E}-05$ |
| rs481069 | 2p24.1 | 46.16 | $1.64 \mathrm{E}-53$ | Dyslipidemia | 100\% | positive | $2.89 \mathrm{E}-30$ |
| rs1367117 | 2p24.1 | 38.25 | $1.34 \mathrm{E}-45$ | Dyslipidemia | 100\% | negative | $5.26 \mathrm{E}-43$ |
| rs7604788 | 2p24.1 | 26.91 | 2.95E-34 | Dyslipidemia | 100\% | positive | $1.19 \mathrm{E}-13$ |
| rs17398765 | 2p24.1 | 17.46 | $8.33 \mathrm{E}-25$ | Dyslipidemia | 100\% | negative | $2.11 \mathrm{E}-15$ |
| rs11709077 | 3p25.2 | 9.86 | $3.29 \mathrm{E}-17$ | Type 2 Diabetes | 100\% | positive | $1.26 \mathrm{E}-09$ |
| rs115946033 | 3 q 25.32 | 8.59 | $6.07 \mathrm{E}-16$ | Type 2 Diabetes | 100\% | positive | $3.83 \mathrm{E}-07$ |
| rs75081018 | 5q34 | 16.24 | $1.37 \mathrm{E}-23$ | Iron Deficiency | 100\% | positive | $1.06 \mathrm{E}-05$ |
| rs183671 | 5p13.2 | 13.00 | $2.37 \mathrm{E}-20$ | Cancers | 100\% | positive | $4.13 \mathrm{E}-09$ |
| rs12916 | 5 q 13.3 | 8.32 | $1.15 \mathrm{E}-15$ | Dyslipidemia | 100\% | negative | $1.46 \mathrm{E}-22$ |
| rs12203592 | 6p25.3 | 77.46 | $8.27 \mathrm{E}-85$ | Cancers | 100\% | negative | $2.36 \mathrm{E}-48$ |
| rs4151672 | 6 p 21.33 | 40.39 | $9.64 \mathrm{E}-48$ | Macular Degeneration | 100\% | positive | $1.01 \mathrm{E}-08$ |
| rs2300051 | 7q31.1 | 113.86 | $3.28 \mathrm{E}-121$ | Peptic Ulcer | 100\% | negative | $3.93 \mathrm{E}-07$ |
| rs849135 | 7p15.1 | 7.66 | $5.22 \mathrm{E}-15$ | Type 2 Diabetes | 100\% | positive | $1.79 \mathrm{E}-14$ |
| rs17321515 | 8q24.13 | 10.81 | $3.73 \mathrm{E}-18$ | Dyslipidemia | 100\% | positive | $1.19 \mathrm{E}-24$ |
| rs74580577 | 9 q 22.33 | 14.09 | $1.95 \mathrm{E}-21$ | Peptic Ulcer | 100\% | negative | $1.81 \mathrm{E}-05$ |
| rs61871745 | 10q26.13 | 300.00 | $1.00 \mathrm{E}-300$ | Macular Degeneration | 100\% | negative | $1.96 \mathrm{E}-75$ |
| rs2292627 | 10q26.13 | 84.13 | $1.75 \mathrm{E}-91$ | Macular Degeneration | 100\% | negative | $8.23 \mathrm{E}-21$ |
| rs2253755 | 10q26.13 | 51.81 | $3.65 \mathrm{E}-59$ | Macular Degeneration | 100\% | negative | $2.94 \mathrm{E}-21$ |
| rs2672589 | 10q26.13 | 22.96 | $2.60 \mathrm{E}-30$ | Macular Degeneration | 100\% | positive | $3.36 \mathrm{E}-16$ |
| rs75799135 | 10q26.13 | 17.07 | $2.04 \mathrm{E}-24$ | Macular Degeneration | 100\% | negative | $1.95 \mathrm{E}-13$ |
| rs76517520 | 10p13 | 13.60 | $5.95 \mathrm{E}-21$ | Peptic Ulcer | 100\% | negative | $1.39 \mathrm{E}-05$ |
| rs7079711 | 10q25.2 | 8.22 | $1.45 \mathrm{E}-15$ | Type 2 Diabetes | 100\% | positive | $6.74 \mathrm{E}-12$ |
| rs1799963 | 11p11.2 | 41.52 | 7.13E-49 | Peripheral Vascular Disease | 100\% | negative | $1.19 \mathrm{E}-08$ |
| rs964184 | 11q23.3 | 32.15 | $1.68 \mathrm{E}-39$ | Dyslipidemia | 100\% | negative | $5.49 \mathrm{E}-28$ |
| rs55975204 | 11q13.2 | 11.43 | $8.91 \mathrm{E}-19$ | Osteoporosis | 100\% | negative | $2.00 \mathrm{E}-09$ |
| rs74836424 | 16q24.3 | 12.39 | $9.69 \mathrm{E}-20$ | Cancers | 100\% | negative | $1.14 \mathrm{E}-12$ |
| rs1801689 | 17q24.2 | 8.78 | $3.93 \mathrm{E}-16$ | Dyslipidemia | 100\% | negative | $4.80 \mathrm{E}-09$ |
| rs8092654 | 18p11.31 | 12.24 | $1.39 \mathrm{E}-19$ | Peripheral Vascular Disease | 100\% | negative | $4.18 \mathrm{E}-06$ |
| rs56289821 | 19p13.2 | 151.61 | $5.81 \mathrm{E}-159$ | Dyslipidemia | 100\% | positive | $2.09 \mathrm{E}-62$ |
| rs28399654 | 19q13.32 | 54.76 | $4.12 \mathrm{E}-62$ | Dyslipidemia | 100\% | positive | $2.79 \mathrm{E}-17$ |
| rs76075198 | 19q13.31 | 41.56 | $6.53 \mathrm{E}-49$ | Dyslipidemia | 100\% | positive | $5.29 \mathrm{E}-11$ |
| rs34095326 | 19q13.32 | 40.41 | $9.18 \mathrm{E}-48$ | Dyslipidemia | 100\% | negative | $4.03 \mathrm{E}-25$ |
| rs2927472 | 19q13.32 | 16.99 | $2.42 \mathrm{E}-24$ | Dyslipidemia | 100\% | positive | $1.56 \mathrm{E}-21$ |
| rs2230199 | 19p13.3 | 7.58 | $6.28 \mathrm{E}-15$ | Macular Degeneration | 100\% | negative | $4.42 \mathrm{E}-09$ |

bioRxiv preprint doi: https://doi.org/10.1101/101543; this version posted January 19, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Bayesian meta-analysis for studying cross-phenotype genetic associations

Table S20: Independent pleiotropic signals detected by ASSET for chromosome 1-2

| rsID | chrom <br> band | ASTpv | Subset of | Univariate |
| :--- | :--- | :--- | :--- | :--- |
| rs2300430 | 1q31.3 | $1.52 \mathrm{E}-72$ | Allergic Rhinitis | p-values |
|  |  |  | Depressive Disorder | 0.34 |
|  |  |  | Dermatophytosis | 0.18 |
|  |  |  | Hemorrhoids | 0.09 |
|  |  |  | Insomnia | 0.36 |
|  |  |  | Macular Degeneration | 0.47 |
|  |  |  | Osteoporosis | $2.57 \mathrm{E}-77$ |
|  |  |  |  | Peptic Ulcer |

Bayesian meta-analysis for studying cross-phenotype genetic associations

Table S21: Independent pleiotropic signals detected by ASSET for chromosome 3-7

| rsID | chrom band | ASTpv | Subset of associated traits | Univariate p-values |
| :---: | :---: | :---: | :---: | :---: |
| rs3846662 | 5q13.3 | 8.84E-16 | Depressive Disorder | 0.18 |
|  |  |  | Type 2 Diabetes | 0.02 |
|  |  |  | Dyslipidemia | $2.29 \mathrm{E}-22$ |
|  |  |  | Iron Deficiency | 0.48 |
|  |  |  | Irritable Bowel Syndrome | 0.49 |
|  |  |  | Macular Degeneration | 0.18 |
|  |  |  | Peptic Ulcer | 0.34 |
| rs12203592 | 6p25.3 | 7.70E-42 | Cancers | $2.36 \mathrm{E}-48$ |
|  |  |  | Dyslipidemia | 0.13 |
|  |  |  | Hypertension | 0.006 |
|  |  |  | Osteoarthritis | 0.10 |
|  |  |  | Peptic Ulcer | 0.08 |
|  |  |  | Psychiatric disorders | 0.02 |
| rs78825896 | 6 p 21.32 | 7.32E-13 | Asthma | 0.003 |
|  |  |  | Allergic Rhinitis | 0.06 |
|  |  |  | Cancers | $1.92 \mathrm{E}-07$ |
|  |  |  | Dyslipidemia | $2.20 \mathrm{E}-13$ |
|  |  |  | Hemorrhoids | 0.03 |
|  |  |  | Iron Deficiency | 0.46 |
|  |  |  | Irritable Bowel Syndrome | 0.43 |
|  |  |  | Peptic Ulcer | 0.45 |
| rs10455872 | 6q25.3 | $5.49 \mathrm{E}-11$ | Allergic Rhinitis | 0.07 |
|  |  |  | Cardiovascular Disease | $6.14 \mathrm{E}-05$ |
|  |  |  | Dyslipidemia | $6.97 \mathrm{E}-15$ |
|  |  |  | Insomnia | 0.37 |
|  |  |  | Iron Deficiency | 0.57 |
|  |  |  | Irritable Bowel Syndrome | 0.34 |
|  |  |  | Osteoarthritis | 0.12 |
|  |  |  | Peripheral Vascular Disease | 0.0002 |
| rs522162 | 6p21.33 | 3.16E-09 | Hypertension | 0.16 |
|  |  |  | Hemorrhoids | 0.13 |
|  |  |  | Macular Degeneration | $1.02 \mathrm{E}-15$ |
|  |  |  | Osteoarthritis | 0.26 |
|  |  |  | Peptic Ulcer | 0.62 |
|  |  |  | Varicose Veins | 0.07 |
| rs2395182 | 6p21.32 | $1.87 \mathrm{E}-08$ | Asthma | 0.009 |
|  |  |  | Allergic Rhinitis | 0.05 |
|  |  |  | Cancers | 0.001 |
|  |  |  | Dermatophytosis | 0.016 |
|  |  |  | Type 2 Diabetes | $1.36 \mathrm{E}-05$ |
|  |  |  | Dyslipidemia | $9.65 \mathrm{E}-06$ |
|  |  |  | Insomnia | 0.09 |
|  |  |  | Iron Deficiency | 0.49 |
|  |  |  | Macular Degeneration | 0.08 |
|  |  |  | Osteoporosis | 0.39 |
|  |  |  | Peripheral Vascular Disease | 0.13 |
|  |  |  | Peptic Ulcer | 0.30 |
|  |  |  | Psychiatric disorders | 0.36 |
| rs864745 | 7p15.1 | 2.52E-09 | Asthma | $1.25 \mathrm{E}-05$ |
|  |  |  | Cancers | 0.30 |
|  |  |  | Type 2 Diabetes | $4.19 \mathrm{E}-14$ |
|  |  |  | Hemorrhoids | 0.52 |
|  |  |  | Macular Degeneration | 0.33 |
|  |  |  | Peripheral Vascular Disease | 0.23 |

bioRxiv preprint doi: https://doi.org/10.1101/101543; this version posted January 19, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Bayesian meta-analysis for studying cross-phenotype genetic associations

Table S22: Independent pleiotropic signals detected by ASSET for chromosome 8-10

| rsID | chrom band | ASTpv | Subset of associated traits | Univariate p-values |
| :---: | :---: | :---: | :---: | :---: |
| rs2001945 | 8q24.13 | $4.01 \mathrm{E}-19$ | Asthma | 0.43 |
|  |  |  | Allergic Rhinitis | 0.26 |
|  |  |  | Depressive Disorder | 0.39 |
|  |  |  | Dyslipidemia | $1.86 \mathrm{E}-25$ |
|  |  |  | Abdominal Hernia | 0.37 |
|  |  |  | Iron Deficiency | 0.61 |
|  |  |  | Macular Degeneration | 0.07 |
|  |  |  | Osteoporosis | 0.39 |
|  |  |  | Stress Disorders | 0.30 |
| rs651007 | 9 q 34.2 | $3.09 \mathrm{E}-16$ | Depressive Disorder | 0.09 |
|  |  |  | Dyslipidemia | $2.89 \mathrm{E}-15$ |
|  |  |  | Hypertension | 0.03 |
|  |  |  | Hemorrhoids | 0.17 |
|  |  |  | Abdominal Hernia | 0.34 |
|  |  |  | Peripheral Vascular Disease | $9.55 \mathrm{E}-08$ |
|  |  |  | Peptic Ulcer | 0.03 |
|  |  |  | Varicose Veins | 0.0027 |
| rs61871745 | 10q26.13 | 6.99E-69 | Depressive Disorder | 0.04 |
|  |  |  | Hemorrhoids | 0.05 |
|  |  |  | Insomnia | 0.16 |
|  |  |  | Macular Degeneration | 1.96E-75 |
|  |  |  | Stress Disorders | 0.01 |
| rs4506565 | 10q25.2 | $1.48 \mathrm{E}-46$ | Asthma | 0.18 |
|  |  |  | Type 2 Diabetes | $2.02 \mathrm{E}-55$ |
|  |  |  | Irritable Bowel Syndrome | 0.37 |
|  |  |  | Peptic Ulcer | 0.63 |
|  |  |  | Varicose Veins | 0.36 |
| rs2292627 | 10q26.13 | $9.71 \mathrm{E}-15$ | Type 2 Diabetes | 0.06 |
|  |  |  | Hemorrhoids | 0.34 |
|  |  |  | Insomnia | 0.29 |
|  |  |  | Irritable Bowel Syndrome | 0.54 |
|  |  |  | Macular Degeneration | $8.23 \mathrm{E}-21$ |
|  |  |  | Peptic Ulcer | 0.54 |
|  |  |  | Stress Disorders | 0.36 |
|  |  |  | Varicose Veins | 0.65 |
| rs2253755 | 10q26.13 | $3.65 \mathrm{E}-14$ | Allergic Rhinitis | 0.18 |
|  |  |  | Cardiovascular Disease | 0.29 |
|  |  |  | Dermatophytosis | 0.37 |
|  |  |  | Type 2 Diabetes | 0.14 |
|  |  |  | Macular Degeneration | $2.94 \mathrm{E}-21$ |
|  |  |  | Psychiatric disorders | 0.26 |
|  |  |  | Stress Disorders | 0.39 |
| rs2672589 | 10q26.13 | $1.70 \mathrm{E}-10$ | Iron Deficiency | 0.17 |
|  |  |  | Irritable Bowel Syndrome | 0.29 |
|  |  |  | Macular Degeneration | $3.36 \mathrm{E}-16$ |
|  |  |  | Peptic Ulcer | 0.058 |
|  |  |  | Stress Disorders | 0.019 |
| rs75799135 | 10q26.13 | $2.45 \mathrm{E}-08$ | Depressive Disorder | 0.04 |
|  |  |  | Type 2 Diabetes | 0.029 |
|  |  |  | Hemorrhoids | 0.088 |
|  |  |  | Insomnia | 0.037 |
|  |  |  | Macular Degeneration | $1.95 \mathrm{E}-13$ |
|  |  |  | Stress Disorders | 0.19 |

bioRxiv preprint doi: https://doi.org/10.1101/101543; this version posted January 19, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Bayesian meta-analysis for studying cross-phenotype genetic associations

Table S23: Independent pleiotropic signals detected by ASSET for chromosome 11-22

| rsID | chrom band | ASTpv | Subset of associated traits | Univariate p-values |
| :---: | :---: | :---: | :---: | :---: |
| rs964184 | 11q23.3 | $5.36 \mathrm{E}-21$ | Asthma | 0.32 |
|  |  |  | Allergic Rhinitis | 0.16 |
|  |  |  | Dyslipidemia | $5.49 \mathrm{E}-28$ |
|  |  |  | Insomnia | 0.39 |
|  |  |  | Iron Deficiency | 0.11 |
|  |  |  | Irritable Bowel Syndrome | 0.42 |
|  |  |  | Macular Degeneration | 0.24 |
|  |  |  | Osteoarthritis | 0.23 |
|  |  |  | Osteoporosis | 0.077 |
|  |  |  | Peripheral Vascular Disease | 0.46 |
|  |  |  | Peptic Ulcer | 0.13 |
| rs3764261 | 16q13 | $1.32 \mathrm{E}-09$ | Depressive Disorder | 0.09 |
|  |  |  | Dyslipidemia | 7.51E-13 |
|  |  |  | Irritable Bowel Syndrome | 0.27 |
|  |  |  | Macular Degeneration | $8.58 \mathrm{E}-05$ |
|  |  |  | Osteoarthritis | 0.0029 |
|  |  |  | Osteoporosis | 0.023 |
|  |  |  | Peripheral Vascular Disease | 0.14 |
|  |  |  | Peptic Ulcer | 0.2 |
|  |  |  | Stress Disorders | 0.24 |
| rs56289821 | 19p13.2 | $1.17 \mathrm{E}-52$ | Asthma | 0.44 |
|  |  |  | Allergic Rhinitis | 0.35 |
|  |  |  | Dermatophytosis | 0.47 |
|  |  |  | Dyslipidemia | $2.09 \mathrm{E}-62$ |
|  |  |  | Osteoarthritis | 0.54 |
|  |  |  | Stress Disorders | 0.69 |
|  |  |  | Varicose Veins | 0.48 |
| rs34095326 | 19q13.32 | $7.53 \mathrm{E}-20$ | Cancers | 0.045 |
|  |  |  | Type 2 Diabetes | 0.018 |
|  |  |  | Dyslipidemia | $4.03 \mathrm{E}-25$ |
|  |  |  | Irritable Bowel Syndrome | 0.25 |
|  |  |  | Macular Degeneration | 0.096 |
|  |  |  | Osteoporosis | 0.016 |
|  |  |  | Peptic Ulcer | 0.60 |
| rs2927472 | 19q13.32 | $1.20 \mathrm{E}-13$ | Cancers | 0.14 |
|  |  |  | Dyslipidemia | $1.56 \mathrm{E}-21$ |
|  |  |  | Hemorrhoids | 0.055 |
|  |  |  | Macular Degeneration | 0.51 |
|  |  |  | Peptic Ulcer | 0.29 |
| rs28399654 | 19q13.32 | $1.37 \mathrm{E}-09$ | Allergic Rhinitis | 0.40 |
|  |  |  | Depressive Disorder | 0.43 |
|  |  |  | Dermatophytosis | 0.62 |
|  |  |  | Dyslipidemia | $2.79 \mathrm{E}-17$ |
|  |  |  | Macular Degeneration | 0.51 |
|  |  |  | Peripheral Vascular Disease | 0.65 |
|  |  |  | Peptic Ulcer | 0.65 |

Bayesian meta-analysis for studying cross-phenotype genetic associations
Table S24: Pair-wise trait-trait correlation matrix in the GERA cohort.

|  | astm | alrg | card | canc | dprs | drmt | dia2 | dysl | hypr | hemr | hern | insm | iron | irrb | mcdg | osta | ostp | pvd | pept | psyc | strs | vrvn |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| astm | 1.00 | 0.27 | 0.06 | 0.00 | 0.09 | 0.03 | 0.05 | 0.04 | 0.06 | 0.03 | 0.03 | 0.05 | 0.05 | 0.06 | 0.02 | 0.08 | 0.02 | 0.05 | 0.02 | 0.09 | 0.08 | 0.02 |
| alrg | 0.27 | 1.00 | 0.02 | 0.00 | 0.08 | 0.05 | 0.01 | 0.03 | 0.03 | 0.04 | 0.02 | 0.06 | 0.02 | 0.09 | 0.01 | 0.08 | 0.02 | 0.02 | 0.02 | 0.09 | 0.09 | 0.03 |
| card | 0.06 | 0.02 | 1.00 | 0.16 | 0.01 | 0.08 | 0.16 | 0.27 | 0.32 | 0.05 | 0.12 | 0.06 | 0.09 | 0.02 | 0.14 | 0.18 | 0.07 | 0.22 | 0.09 | 0.03 | 0.02 | 0.03 |
| canc | 0.00 | 0.00 | 0.16 | 1.00 | -0.02 | 0.07 | 0.05 | 0.10 | 0.14 | 0.06 | 0.10 | 0.03 | 0.05 | 0.01 | 0.10 | 0.11 | 0.07 | 0.14 | 0.04 | -0.01 | 0.00 | 0.03 |
| dprs | 0.09 | 0.08 | 0.01 | -0.02 | 1.00 | 0.03 | 0.03 | 0.02 | 0.01 | 0.03 | 0.00 | 0.13 | 0.04 | 0.09 | 0.00 | 0.06 | 0.01 | 0.03 | 0.02 | 0.35 | 0.17 | 0.01 |
| drmt | 0.03 | 0.05 | 0.08 | 0.07 | 0.03 | 1.00 | 0.08 | 0.06 | 0.07 | 0.04 | 0.06 | 0.04 | 0.03 | 0.02 | 0.04 | 0.10 | 0.01 | 0.07 | 0.02 | 0.03 | 0.04 | 0.05 |
| dia2 | 0.05 | 0.01 | 0.16 | 0.05 | 0.03 | 0.08 | 1.00 | 0.26 | 0.28 | 0.01 | 0.04 | 0.01 | 0.07 | -0.01 | 0.05 | 0.08 | -0.03 | 0.08 | 0.04 | 0.01 | 0.03 | 0.00 |
| dysl | 0.04 | 0.03 | 0.27 | 0.10 | 0.02 | 0.06 | 0.26 | 1.00 | 0.36 | 0.06 | 0.07 | 0.04 | 0.04 | 0.01 | 0.07 | 0.16 | 0.04 | 0.12 | 0.05 | 0.02 | 0.03 | 0.01 |
| hypr | 0.06 | 0.03 | 0.32 | 0.14 | 0.01 | 0.07 | 0.28 | 0.36 | 1.00 | 0.05 | 0.09 | 0.04 | 0.07 | 0.02 | 0.13 | 0.21 | 0.05 | 0.15 | 0.06 | 0.02 | 0.04 | 0.02 |
| hemr | 0.03 | 0.04 | 0.05 | 0.06 | 0.03 | 0.04 | 0.01 | 0.06 | 0.05 | 1.00 | 0.06 | 0.04 | 0.05 | 0.07 | 0.02 | 0.06 | 0.01 | 0.04 | 0.03 | 0.04 | 0.03 | 0.02 |
| her | 0.03 | 0.02 | 0.12 | 0.10 | 0.00 | 0.06 | 0.04 | 0.07 | 0.09 | 0.06 | 1.00 | 0.02 | 0.08 | 0.01 | 0.05 | 0.07 | 0.00 | 0.09 | 0.06 | 0.00 | -0.01 | 0.02 |
| ins | 0.05 | 0.06 | 0.06 | 0.03 | 0.13 | 0.04 | 0.01 | 0.04 | 0.04 | 0.04 | 0.02 | 1.00 | 0.04 | 0.08 | 0.03 | 0.08 | 0.04 | 0.06 | 0.03 | 0.15 | 0.10 | 0.02 |
| iro | 0.05 | 0.02 | 0.09 | 0.05 | 0.04 | 0.03 | 0.07 | 0.04 | 0.07 | 0.05 | 0.08 | 0.04 | 1.00 | 0.03 | 0.04 | 0.06 | 0.03 | 0.09 | 0.10 | 0.04 | 0.03 | 0.02 |
| irr | 0.06 | 0.09 | 0.02 | 0.01 | 0.09 | 0.02 | -0.01 | 0.01 | 0.02 | 0.07 | 0.01 | 0.08 | 0.03 | 1.00 | 0.02 | 0.07 | 0.05 | 0.01 | 0.03 | 0.14 | 0.11 | 0.02 |
| mcdg | 0.02 | 0.01 | 0.14 | 0.10 | 0.00 | 0.04 | 0.05 | 0.07 | 0.13 | 0.02 | 0.05 | 0.03 | 0.04 | 0.02 | 1.00 | 0.11 | 0.09 | 0.10 | 0.03 | 0.00 | 0.00 | 0.02 |
| osta | 0.08 | 0.08 | 0.18 | 0.11 | 0.06 | 0.10 | 0.08 | 0.16 | 0.21 | 0.06 | 0.07 | 0.08 | 0.06 | 0.07 | 0.11 | 1.00 | 0.11 | 0.13 | 0.06 | 0.04 | 0.07 | 0.07 |
| ostp | 0.02 | 0.02 | 0.07 | 0.07 | 0.01 | 0.01 | -0.03 | 0.04 | 0.05 | 0.01 | 0.00 | 0.04 | 0.03 | 0.05 | 0.09 | 0.11 | 1.00 | 0.08 | 0.02 | 0.03 | 0.03 | 0.04 |
| pvd | 0.05 | 0.02 | 0.22 | 0.14 | 0.03 | 0.07 | 0.08 | 0.12 | 0.15 | 0.04 | 0.09 | 0.06 | 0.09 | 0.01 | 0.10 | 0.13 | 0.08 | 1.00 | 0.06 | 0.03 | 0.02 | 0.04 |
| pept | 0.02 | 0.02 | 0.09 | 0.04 | 0.02 | 0.02 | 0.04 | 0.05 | 0.06 | 0.03 | 0.06 | 0.03 | 0.10 | 0.03 | 0.03 | 0.06 | 0.02 | 0.06 | 1.00 | 0.02 | 0.01 | 0.01 |
| psyc | 0.09 | 0.09 | 0.03 | -0.01 | 0.35 | 0.03 | 0.01 | 0.02 | 0.02 | 0.04 | 0.00 | 0.15 | 0.04 | 0.14 | 0.00 | 0.04 | 0.03 | 0.03 | 0.02 | 1.00 | 0.24 | 0.02 |
| strs | 0.08 | 0.09 | 0.02 | 0.00 | 0.17 | 0.04 | 0.03 | 0.03 | 0.04 | 0.03 | -0.01 | 0.10 | 0.03 | 0.11 | 0.00 | 0.07 | 0.03 | 0.02 | 0.01 | 0.24 | 1.00 | 0.03 |
| vrvn | 0.02 | 0.03 | 0.03 | 0.03 | 0.01 | 0.05 | 0.00 | 0.01 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.07 | 0.04 | 0.04 | 0.01 | 0.02 | 0.03 | 1.00 |

## 10 Comparison between the continuous spike and Dirac spike

We carried out simulation study to compare the selection accuracy of two different type of spikes. We chose the same set-up of multiple overlapping case-control studies considered while comparing CPBayes and ASSET in the main simulation study. We implemented the Gibbs sampling algorithm for the Dirac spike described in Algorithm 2. Since the summary statistics are correlated, we implemented the combined strategy for the Dirac spike as well as for the continuous spike. The slab variance for the Dirac spike $\left(b^{2}\right)$ is considered to vary in $0.8-1.2$ (the same as that for the continuous spike). We compute the mean specificity and sensitivity across 200 replications. The results are provided in Figure S3 (see the next page).

We observe that the Dirac spike produces less specificity than the continuous spike. The Dirac spike suffers from reduced specificity more for larger number of associated traits $\left(K_{1}\right)$. The continuous spike consistently yields very good level of specificity across different scenarios. The Dirac spike offers higher sensitivity, but at the expense of lower specificity compared to the continuous spike. For example, for $K=10, K_{1}^{+}=4 \& K_{1}^{-}=4\left(K_{1}=8\right)$, and $m=0.1$, the Dirac spike produced a mean specificity of $64 \%$ and sensitivity of $87 \%$, whereas the continuous spike gave a mean specificity of $97 \%$ and sensitivity of $72 \%$. Similarly, for $K=15, K_{1}^{+}=5 \& K_{1}^{-}=4\left(K_{1}=9\right)$, and $m=0.1$, the Dirac spike gave $65 \%$ specificity and $85 \%$ sensitivity, whereas the continuous spike produced $99 \%$ specificity and $69 \%$ sensitivity.

Figure S3: Comparison of the accuracy of selection of associated traits by the continuous and the Dirac spike for multiple overlapping case-control studies. The total number of phenotypes/studies is denoted by $K$ and $m$ denotes the minor allele frequency at the risk SNP.




