Text S1

Model and Algorithm Details for iMAP

1. Basic model setup and prior specifications

We first consider the case where the two traits of interest come from a common GWAS consortium and are measured on the same set of individuals with the same sample size n. Extending iMAP to the case that the two traits are measured in different GWAS consortia with different sample sizes will be discussed in Section 6. In iMAP, we consider the following multivariate linear model for each SNP j in turn:

$$\begin{aligned} &\mathbf{y}_i = \boldsymbol{\beta}_j g_{ij} + \boldsymbol{e}_{ij},\ \ i = 1, 2, \cdots, n,\ \ j = 1, 2, \cdots, m, \\ &\boldsymbol{e}_{ij} \sim \mathrm{MVN}_2(0, \Sigma_j), \end{aligned} \tag{1}$$

where n is the number of individuals; m is the number of SNPs; $\mathbf{y}_i$ is a two by one phenotype vector that consists of y_{i1} and y_{i2} measured on the same individual i; g_{ij} is the genotype for SNP j of individual i; $\boldsymbol{\beta}_j$ is the corresponding two-vector of effect sizes of SNP j on the two phenotypes; $\boldsymbol{e}_{ij}$ is a two-vector of residual error with a covariance matrix $\boldsymbol{\Sigma}_j$ that accounts for phenotype correlation; and MVN_2 denotes a bivariate normal distribution. While Equation (1) is primarily aimed to deal with quantitative traits, we also use Equation (1) to model binary traits by treating binary data as continuous values following many previous studies (Moser, et al., 2015; Speed and Balding, 2014; Weissbrod, et al., 2016; Zhou, et al., 2013). Methodologically, modeling binary data with linear model can be justified by the fact that a linear model is a first order Taylor approximation to a generalized linear model; and the approximation is accurate when the SNP effect size is small (Zhou, et al., 2013) — a condition that generally holds as most complex phenotypes are polygenic and are affected by many SNPs with small effects (Visscher, et al., 2017).

The above model (1) is specified on SNP j. To infer genome-wide pleiotropic association pattern and improve association mapping power, we model all SNPs jointly by assuming that the joint likelihood for all SNPs is simply a product of the likelihood for each SNP. To facilitate information sharing across genome-wide SNPs, we assign a common prior on the effects and assume that each $\boldsymbol{\beta}_j$ *a priori* follows the same four-component Gaussian mixture distribution (1)

$$\boldsymbol{\beta}_j \sim \pi_{11}\mathrm{MVN}_2(0, \mathbf{V}_{11}) + \pi_{10}\mathrm{MVN}_2(0, \mathbf{V}_{10}) + \pi_{01}\mathrm{MVN}_2(0, \mathbf{V}_{01}) + \pi_{00}\delta_0, \qquad (2)$$

with a prior probabilities π_k (k = 11, 10, 01 and 00) that sum to one. Here π_{11} represents the prior probability that a SNP is associated with both traits; $V_{11} = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 \end{pmatrix}$ is a two by two covariance matrix modeling the covariance of SNP effects on the two traits. π_{10} represents the prior probability that a SNP is associated with the first trait but not the second; $V_{10} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & 0 \end{pmatrix}$ is a low-rank covariance matrix restricting that only the effect size for the first trait is nonzero. π_{01} represents the prior probability that a SNP is associated with the second trait but not the first; $V_{01} = \begin{pmatrix} 0 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$ is a low-rank covariance matrix restricting nonzero effect only for the second trait. Finally, π_{00} represents the prior probability that a SNP is not associated with any traits; δ_0 denotes a point mass at zero. We specify conjugate priors for the three hyper-parameters $\mathbf{V}_k$ (k = 11, 10, and 01) and we borrow information across genome-wide SNPs to infer these parameters. The estimated probability of a SNP having nonzero effects on any traits (i.e. $\pi_{11} + \pi_{10} + \pi_{01}$) is commonly referred to as the posterior inclusion probability (PIP), which represents association evidence for the given SNP. We can also use PIP to provide a conservative estimate of false discovery rate (FDR) (Benjamini and Hochberg, 1995) based on local false discovery rate (Efron, et al., 2001) via the direct posterior probability approach (Newton, et al., 2004). We impose the following priors for the unknown parameters in model (2)

$$
\begin{aligned}
p(\pi) &\propto \pi_{11}^{\alpha_{11}} \pi_{10}^{\alpha_{10}} \pi_{01}^{\alpha_{01}} \pi_{00}^{\alpha_{00}}, \\
p(\mathbf{V}_k) &\propto \text{Inv-Wishart}_{\nu_0}(\Lambda_0^{-1}),\ k = 11, 01, 10, \\
\text{Inv-Wishart}_{\nu_0}(\Lambda_0^{-1}) &\propto |\mathbf{V}|^{-\frac{\nu_0+d+1}{2}} e^{-\frac{1}{2}\mathrm{tr}(\nu_0\Lambda_0\mathbf{V}^{-1})},
\end{aligned}
\tag{3}
$$

where d is the dimension of $\mathbf{V}$. To ensure numeric stability, we follow (Chung, et al., 2015; Gelman, et al., 2008) and specify relatively informative priors for the hyper-parameters $\boldsymbol{\alpha} = (\alpha_{11}, \alpha_{10}, \alpha_{01}$ and $\alpha_{00})$, ν_0 and $\boldsymbol{\Lambda}_0$. Specifically, we set $\alpha_k = m/1000$, where m is the total number of SNPs. We set the mean of the inverse-Wishart $\boldsymbol{\Lambda}_0\nu_0 = \mathbf{V}_0 m/1000$, where $\mathbf{V}_0$ is the estimated phenotypic covariance matrix to be obtained from summary statistics (Pickrell, et al., 2016; Stephens, 2013).

2. Likelihood function and posterior distribution

To facilitate computation, for each SNP j we assign a 4-vector of membership indicator variables $\boldsymbol{\gamma}_j$, here each element $\gamma_{jk} = 1$ if $\boldsymbol{\beta}_j$ is from the k^{th} normal component and $\gamma_{jk} = 0$ otherwise, for k = 11, 01, 10 or 00. Let $\boldsymbol{\theta}_j = (\boldsymbol{\beta}_j, \boldsymbol{\gamma}_j, \Sigma_j)$ and $\Omega = \{11, 10, 01, 00\}$. We denote $\boldsymbol{\theta} = (\boldsymbol{\theta}_j, \cdots, \boldsymbol{\theta}_m, \mathrm{V}_{11}, \mathrm{V}_{01}, \mathrm{V}_{10}, \pi_{11}, \pi_{01}, \pi_{10})$ to include all the unknown parameters in the iMAP model. To write down the likelihood, instead of focusing on the original phenotypes $\mathbf{y}$, following (Zhu and Stephens, 2017) we focus on the marginal effect size estimates $\hat{\boldsymbol{\beta}}_j$. Each $\hat{\boldsymbol{\beta}}_j$ is a two-dimensional vector with each element obtained by fitting a generalized linear regression model between the genotype vector and the corresponding trait. For continuous traits, we use the original scale of $\hat{\boldsymbol{\beta}}_j$, while for binary traits, we use the log-scale of $\hat{\boldsymbol{\beta}}_j$ (i.e. log odd ratio). Then, under the model described in Equation (1), $\hat{\boldsymbol{\beta}}_j$ marginally follows a multivariate normal distribution. The effect size estimates $\hat{\boldsymbol{\beta}}_j$ are correlated among each other due to linkage disequilibrium. We ignore the correlation among SNPs and rely on independence marginal likelihood for inference. Independence marginal likelihood can be considered as a special case of composite likelihood that is commonly used for statistical estimation and inference of complex models, for which it is impossible or difficult to yield and estimate the full likelihood due to complicated dependence (Larribe and Fearnhead, 2011; Varin, 2008; Varin, et al., 2011). Under regulatory conditions, the point estimates obtained based on composite likelihood are consistent and asymptotically normally distributed (Kenne Pagui, et al., 2015; Varin, 2008; Varin, et al., 2011). The independent composite likelihood for all SNPs is a simple product of the likelihood for each SNP

$$\log p(\hat{\beta}_1, ..., \hat{\beta}_m \mid \theta) \sim \sum_{j=1}^{m} \log p(\hat{\beta}_j \mid \theta), \tag{3}$$

which, after ignoring the constant terms, is in turn equivalent to assuming that the joint likelihood based of θ on $\mathbf{y}$ is a simple product of the corresponding likelihood for each SNP

$$\begin{aligned}
\log p(\mathbf{y} \mid \theta) &= \sum_{j=1}^{m} \log p(\mathbf{y} \mid \theta_j) \\
&= \sum_{j=1}^{m} \left\{ \sum_{i=1}^{n} \log p(\mathbf{y}_i \mid \boldsymbol{\beta}_j, \Sigma_j) + \log p(\boldsymbol{\beta}_j \mid \boldsymbol{\gamma}_j, \mathbf{V}_k) + \log p(\boldsymbol{\gamma}_j) \right\}, \\
&= \sum_{j=1}^{m} \left\{ \sum_{i=1}^{n} (-\frac{1}{2} \log |\Sigma_j| - \frac{1}{2} (\mathbf{y}_i - \boldsymbol{\beta}_j g_{ij})^T \Sigma_j^{-1} (\mathbf{y}_i - \boldsymbol{\beta}_j g_{ij})) \right\} \\
&+ \sum_{j=1}^{m} \left\{ \sum_{k \in \Omega} \gamma_{jk} (-\frac{d_k}{2} \log 2\pi - \frac{1}{2} \log |\mathbf{V}_k| - \frac{1}{2} \beta_{jk}^T \mathbf{V}_k^{-1} \boldsymbol{\beta}_{jk}) \right\} \\
&+ \sum_{j=1}^{m} \left\{ \sum_{k \in \Omega} \gamma_{jk} \log(\pi_k) \right\},
\end{aligned} \tag{4}$$

with the log joint posterior as

$$
\begin{aligned}
\log p(\theta \mid \mathbf{y}) &\propto \sum_{j=1}^{m} \log p(\mathbf{y} \mid \theta_j) p(\theta) \\
&= \sum_{j=1}^{m} \left\{ \sum_{i=1}^{n} \log p(\mathbf{y}_i \mid \beta_j, \Sigma_j) + \log p(\beta_j \mid \boldsymbol{\gamma}_j, \mathbf{V}_k) + \log p(\boldsymbol{\gamma}_j) \right\} p(\theta), \\
&= \sum_{j=1}^{m} \left\{ \sum_{i=1}^{n} (-\frac{1}{2} \log |\Sigma_j| - \frac{1}{2} (\mathbf{y}_i - \beta_j g_{ij})^T \Sigma_j^{-1} (\mathbf{y}_i - \beta_j g_{ij})) \right\} \\
&+ \sum_{j=1}^{m} \left\{ \sum_{k \in \Omega} \gamma_{jk} (-\frac{d_k}{2} \log 2\pi - \frac{1}{2} \log |\mathbf{V}_k| - \frac{1}{2} \beta_{jk}^T \mathbf{V}_k^{-1} \beta_{jk}) \right\} \\
&+ \sum_{j=1}^{m} \left\{ \sum_{k \in \Omega} \gamma_{jk} \log(\pi_k) \right\} \\
&+ \sum_{k \in \Omega} \alpha_k \log(\pi_k) + \sum_{k \in \Omega} (-\frac{\nu_0 + d_k + 1}{2} \log |\mathbf{V}_k| - \frac{1}{2} \mathrm{tr}(\Lambda_0 \mathbf{V}_k^{-1}).
\end{aligned}
\tag{5}
$$

3. Expectation-Maximization algorithm

We use the Expectation-Maximization (EM) algorithm (Dempster, et al., 1977) to obtain the maximum a posterior (MAP) estimate of the parameters in model (5). Specifically, we view the mixture group assignment of each SNP $\boldsymbol{\gamma}_j$ as missing data and impute them in the expectation step. We then perform optimization and obtain parameter estimates in the maximization step. The detailed E and M steps are listed below.

E step

The log likelihood for γ_{jk} is

$$
\begin{aligned}
&\log p(\gamma_{jk} \mid \mathbf{y}) \propto \gamma_{jk} C_{jk}, \\
&\qquad C_{jk} = -\frac{1}{2} \log \left| \sum_{i=1}^{n} \Sigma_j^{-1} \mathbf{V}_k g_{ij}^2 + \mathbf{I}_2 \right| + \frac{1}{2} \hat{\beta}_{jk}^T (\sum_{i=1}^{n} \Sigma_j^{-1} g_{ij}^2 + \mathbf{V}_k^{-1}) \hat{\beta}_{jk} + \log(\pi_k),
\end{aligned}
\tag{6}
$$

where $\hat{\beta}_{jk}$ is given in Equation (9) below. We can obtain the conditional expectation for γ_{jk} as

$$
E(\gamma_{jk}) = \varphi_{jk} \propto e^{C_{jk}}. \tag{7}
$$

Replacing γ_{jk} in Equation (6) with its expectation in Equation (7) gives our target function for optimization.

M step

We optimize the expectation of Equation (7) to obtain estimates for each parameter in turn. The log likelihood and estimate for π_k are

$$\begin{gathered}\log p(\pi_k \mid \mathbf{y}) \propto \sum_{j=1}^{m} \gamma_{jk} \log(\pi_k) + \alpha_k \log(\pi_k), \\ \pi_k \propto (\sum_{j=1}^{m} \varphi_{jk} + \alpha_k).\end{gathered} \tag{8}$$

The log likelihood for β_{jk} is

$$\log p(\beta_{jk} \mid \mathbf{y}, \gamma_{jk}) \propto -\frac{1}{2} \gamma_{jk} \beta_{jk}^T (\sum_{i=1}^{n} \Sigma_j^{-1} g_{ij}^2 + \mathbf{V}_k^{-1}) \beta_{jk} + \gamma_{jk} \beta_{jk}^T (\sum_{i=1}^{n} \Sigma_j^{-1} \mathbf{y}_i g_{ij}). \tag{9}$$

And the estimate and its variance for β_{jk} are

$$\begin{gathered}\hat{\beta}_{jk} = (\sum_{i=1}^{n} \Sigma_j^{-1} g_{ij}^2 + \mathbf{V}_k^{-1})^{-1} (\sum_{i=1}^{n} \Sigma_j^{-1} \mathbf{y}_i g_{ij}), \\ Var(\hat{\beta}_{jk}) = (\sum_{i=1}^{n} \Sigma_j^{-1} g_{ij}^2 + \mathbf{V}_k^{-1})^{-1}.\end{gathered} \tag{10}$$

The log likelihood and estimate for $\mathbf{V}_k$ are

$$\begin{aligned}\log p(\mathbf{V}_k \mid \mathbf{y}) \propto & -\frac{1}{2} (\sum_{j=1}^{m} \gamma_{jk} + \nu_0 + d_k + 1) \log |\mathbf{V}_k| \\ & -\frac{1}{2} \mathrm{tr}(\Lambda_0 \mathbf{V}_k^{-1}) + \sum_{j=1}^{m} \gamma_{jk} (-\frac{1}{2} \beta_{jk}^T \mathbf{V}_k^{-1} \beta_{jk}), \\ \mathbf{V}_k = & \frac{1}{m_k} (\sum_{j=1}^{m} \varphi_{jk} \beta_{jk} \beta_{jk}^T + \Lambda_0),\ m_k = \sum_{j=1}^{m} \varphi_{jk} + \nu_0 + d_k + 1.\end{aligned} \tag{11}$$

The log likelihood and estimate for Σ_j are

$$
\begin{aligned}
\log p(\Sigma_j \mid \mathbf{y}) &\propto \sum_{i=1}^{n}\Big(-\frac{1}{2}\log|\Sigma_j| - \frac{1}{2}(\mathbf{y}_i - (\sum_{k=1}^{4}\varphi_{jk}\beta_{jk})\mathbf{g}_{ij})^T \Sigma_j^{-1} (\mathbf{y}_i - (\sum_{k=1}^{4}\varphi_{jk}\beta_{jk})\mathbf{g}_{ij})\Big),\\
\Sigma_j &= \frac{1}{n}\sum_{i=1}^{n}(\mathbf{y}_i - (\sum_{k=1}^{4}\varphi_{jk}\beta_{jk})\mathbf{g}_{ij})(\mathbf{y}_i - (\sum_{k=1}^{4}\varphi_{jk}\beta_{jk})\mathbf{g}_{ij})^T.
\end{aligned}
\tag{12}
$$

We summarize the EM algorithm for iMAP in Algorithm 1, where we denote Δ and Δ_0 as the pre-specified convergence criterion and threshold, respectively. In particular, we define Δ as the difference between two consecutive parameter estimates. In the present study, we set $\Delta = \max_k |m_k^{\ell+1} - m_k^{\ell}|$ (see Equation (13) for the computation of m) and Δ_0 = 5.

4. Use of Summary Statistics

We summarize the estimation procedure above in Algorithm 1, which assumes the availability of individual-level genotypes and phenotypes. However, Algorithm 1 can be easily modified when only summary statistics in terms of marginal z scores are available. To do so, we make two assumptions. First, we assume that both the genotypes and phenotypes are standardized to have mean zero and standard deviation one. This assumption is only for convenience. The algorithm can be easily modified to cases where we know the phenotypic variance and the genotype variance for each SNP – the later can be obtained from a reference panel in practice. Second, we assume that the effect size for each SNP is small, such that the phenotypic variation and the residual error variance are approximately equal to each other. The second assumption is valid for almost all complex traits. The two assumptions we make are commonly employed in previous summary statistics methods (Bulik-Sullivan, et al., 2015; Finucane, et al., 2015; Pasaniuc and Price, 2017; Shi, et al., 2016; Vilhjálmsson, et al., 2015). With these assumptions, we can convert Algorithm 1 to Algorithm 2 that uses only summary statistics in terms of marginal z scores.

Algorithm 1: EM algorithm for iMAP using individual-level data

---

Input: *individual-level data genotypes* $\mathbf{G}$ *and traits* $\mathbf{y}$*; and initial values* $\boldsymbol{\beta}_{jk}^{0}, \mathbf{V}_{k}^{0}$ and π_{k}^{0}

Let $\ell = 0$

while $\Delta > \Delta_0$ **do**

    $\ell = \ell + 1$

    For $j \in \{1, 2, ..., m\}$ **do**

        For $k \in \Omega$ **do**

            Compute $\varphi_{jk}^{\ell+1}, m_k^{\ell+1}, \pi_k^{\ell+1}, \Sigma_j^{\ell+1}, \hat{\boldsymbol{\beta}}_{jk}^{\ell+1}, \mathbf{V}_k^{\ell+1}$ in terms of Equations (6)-(12);

        End

    End

End

Output: $\hat{\varphi}_{jk}, \hat{m}_k, \hat{\pi}_k, \hat{\Sigma}_j, \hat{\boldsymbol{\beta}}_{jk}, \hat{\mathbf{V}}_k$.

Algorithm 2: EM algorithm for iMAP using marginal z scores

Input: summary-level z score values, sample size n, initial values $\boldsymbol{\beta}_{jk}^{0}, \mathbf{V}_{k}^{0}$ and π_{k}^{0}; and the estimated phenotypic covariance matrix Σ directly using marginal z scores from the null SNPs.

Let $\ell = 0$

while $\Delta > \Delta_0$ **do**

    $\ell = \ell + 1$

    For $j \in \{1, 2, ..., m\}$ **do**

        For $k \in \Omega$ **do**

            Compute $\varphi_{jk}^{\ell+1}, m_k^{\ell+1}, \pi_k^{\ell+1}, \hat{\boldsymbol{\beta}}_{jk}^{\ell+1}, \mathbf{V}_k^{\ell+1}$ in terms of Equation (13);

        End

    End

End

Output: $\hat{\varphi}_{jk}, \hat{m}_k, \hat{\pi}_k, \hat{\boldsymbol{\beta}}_{jk}, \hat{\mathbf{V}}_k$.

$$
\begin{aligned}
\varphi_{jk}^{\ell+1} &\propto e^{-\frac{1}{2}\log|n\boldsymbol{\Sigma}^{-1}+(\mathbf{V}_k^{\ell})^{-1}|+\frac{1}{2}(\boldsymbol{\beta}_{jk}^{\ell})^T(n\boldsymbol{\Sigma}^{-1}+(\mathbf{V}_k^{\ell})^{-1})(\boldsymbol{\beta}_{jk}^{\ell})^T-\frac{1}{2}\log|\mathbf{V}_k^{\ell}|+\log(\pi_k^{\ell})}, \\
m_k^{\ell+1} &= \sum_{j=1}^{m}\varphi_{jk}^{\ell+1}, \\
\pi_k^{\ell+1} &\propto (\sum_{j=1}^{m}\varphi_{jk}^{\ell+1}+\alpha_k), \\
\boldsymbol{\beta}_{jk}^{\ell+1} &= (n\Sigma^{-1}+(\mathbf{V}_k^{\ell})^{-1})^{-1}(n^{1/2}\Sigma^{-1}\mathbf{z}_j), \\
\mathbf{V}_k^{\ell+1} &= \frac{1}{m_k^{\ell+1}}(\sum_{j=1}^{m}\varphi_{jk}^{\ell+1}\boldsymbol{\beta}_{jk}^{\ell+1}(\boldsymbol{\beta}_{jk}^{\ell+1})^T+\nu_0\Lambda_0),
\end{aligned}
\tag{13}
$$

5. Estimate phenotypic covariance using summary statistics

iMAP relies on the phenotype covariance matrix $\boldsymbol{\Sigma}$ to account for phenotypic correlation between traits (see Algorithm 2). Here, we show how to estimate $\boldsymbol{\Sigma}$ using only summary statistics. Let $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ be the n by one centered phenotypic vectors for the first and the second traits, respectively. Let $\boldsymbol{G}_j$ be the n by one standardized genotype vector for SNP j (i.e. $\boldsymbol{G}_j^T\boldsymbol{G}_j = n-1$ and $\sum_{i=1}^{n} G_{ij} = 0$) for j = 1, 2, …, m. We fit a standard linear regression on SNP j for each trait separately and we obtain the marginal z scores approximately as

$$
\begin{aligned}
z_{1j} &= \frac{1}{\sqrt{n-1}}\boldsymbol{G}_j^T\boldsymbol{y}_1, \\
z_{2j} &= \frac{1}{\sqrt{n-1}}\boldsymbol{G}_j^T\boldsymbol{y}_2.
\end{aligned}
\tag{14}
$$

As shown in previous studies (Liu and Lin, 2017; Stephens, 2013; Zhu, et al., 2015), under the null that SNP j have zero effects on both traits, the covariance between the marginal z scores is

$$
\begin{aligned}
\mathrm{Cov}(z_{1j}, z_{2j}) &= E(z_{1j}-E(z_{1j}))^T(z_{2j}-E(z_{2j})) \\
&= \frac{1}{n-1}\boldsymbol{G}_j^T\boldsymbol{G}_j E(\boldsymbol{y}_1^T\boldsymbol{y}_2) \\
&= E(\boldsymbol{y}_1^T\boldsymbol{y}_2) \\
&= \mathrm{Cov}(\boldsymbol{y}_1, \boldsymbol{y}_2),
\end{aligned}
\tag{15}
$$

where the second equation holds because the genotype is assumed to be standardized, the phenotypes are assumed to be centered, and both z_{1j} and z_{2j} asymptotically follow a standard normal distribution under the null such that the expectations of z_{1j} and z_{2j} are zero. Therefore, the covariance between the marginal z scores under the null equals to the covariance between the two phenotypes. Relying on the relationship in Equation (15), we first obtain a set of null SNPs that have marginal z scores for both traits whose absolute values are below a threshold of two following (Stephens, 2013). We then compute the sample covariance between the z scores for the two traits across the null SNPs and it as an estimate of $\mathbf{\Sigma}$.

6. Extension to two traits from different consortia

We now extend iMAP to situations where the two traits of interest come from two different GWAS consortia with different sample sizes. In particular, we consider two common scenarios. In the first scenario, the two consortia have no overlapping individuals. In this case, we can use independent linear regression models to model the two traits separately

$$\begin{aligned} \mathbf{y}_{i1} &= \boldsymbol{\beta}_{j1} g_{ij} + \boldsymbol{e}_{ij1}, \quad \boldsymbol{e}_{ij1} \sim N_{n_1}(0, \sigma_1^2), \\ \mathbf{y}_{i2} &= \boldsymbol{\beta}_{j2} g_{ij} + \boldsymbol{e}_{ij2}, \quad \boldsymbol{e}_{ij2} \sim N_{n_2}(0, \sigma_2^2), \end{aligned} \tag{16}$$

which is somewhat equivalent to Equation (1) with $\mathbf{\Sigma}$ in the specific form of $\Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$. We denote n_1 and n_2 as the sample sizes of the two traits. Modifying our algorithms for the likelihood defined in Equation (16) is straightforward and involves only a simple replacement of $n\Sigma^{-1}$ in Algorithms 1 and 2 with $\begin{bmatrix} \sigma_1^{-2} & 0 \\ 0 & \sigma_2^{-2} \end{bmatrix} \times \begin{bmatrix} n_1 & 0 \\ 0 & n_2 \end{bmatrix}$ and a replacement of $n^{1/2}\mathbf{z}_{m\times 2}$ with $\mathbf{z}_{m\times 2} \times \begin{bmatrix} \sqrt{n_1} & 0 \\ 0 & \sqrt{n_2} \end{bmatrix}$. In the second scenario, the two consortia have partially overlapping individuals. Under this situation, we will rely on the model in Equation (1) and follow the strategy described in Section 5 to estimate the variance $\mathbf{\Sigma}$ of the two traits. Intuitively, if there is no sample overlap, then the off-diagonal element of $\mathbf{\Sigma}$ will be estimated to be zero, as the covariance between the marginal z scores from the two traits is expected to be zero. If there is a partial sample overlap, then the off-diagonal element will reflect the part of trait covariance retained due to the partial sample overlap. Therefore, in both scenarios, we can directly follow the procedure described in Section 5 to estimate the covariance of the two traits and then follow the algorithm described in Section 4 for estimation.

7. Functional annotations and penalized selection

We integrate SNP functional annotations into the basic model to facilitate SNP prioritization. Let $\mathbf{X}$ be an m by $(q + 1)$ matrix that contains q functional annotations for m SNPs, where the first column of $\mathbf{X}$ is a column of ones that represent the intercept. We adopt the multinomial logit (mlogit) regression model to relate $\mathbf{X}$ to the mixture prior probabilities π_k ($k = 11$, 10, 01 and 00)

$$\pi_{jk} = \frac{\exp(\sum_{d=0}^{q+1} x_{jd} b_{kd})}{\sum_{k \in \Omega} \exp(\sum_{d=0}^{q+1} x_{jd} b_{kd})} = \frac{\exp(x_j \boldsymbol{b}_k)}{\sum_{k \in \Omega} \exp(x_j \boldsymbol{b}_k)}, \tag{17}$$

where $\boldsymbol{b}_k$ are the annotation coefficients. We consider $k = 00$ as the reference and set $\boldsymbol{b}_{00} = 0$ to ensure model identifiability. Note that, in the cases where SNPs belong to only two categories (e.g. causal vs non-causal), the mlogit model reduces to a logistic regression model that is commonly used to link functional annotations to SNP causality (Carbonetto and Stephens, 2013; Wen, et al., 2016; Wen, et al., 2015). We use a mlogit model here because it naturally extends the logistic model to cases where SNPs can belong to more than two categories (e.g. four categories here: $k = 11$, 10, 01 and 00). The corresponding log likelihood of (17) for $\boldsymbol{b}$ becomes

$$L(\boldsymbol{b}) = \sum_{j=1}^{m} \left(\sum_{k \in \Omega} \log E(\gamma_{jk}) + x_j \boldsymbol{b}_k - \log \sum_{k \in \Omega} \exp(x_j \boldsymbol{b}_k) \right), \tag{18}$$

from which we obtain the estimates $\hat{\boldsymbol{b}}_k = \arg\max\{L(\boldsymbol{b})\}$. The Equation (18) can be viewed as the log likelihood function of a variation of the mlogit model, in which the usual 0/1 binary responses are replaced by the continuous responses $\varphi_{jk} = E(\gamma_{jk})$. Therefore, the optimization of Equation (18) can be carried out based a mlogit model that treats the conditional expectation φ_{jk} as responses.

The classical iteratively re-weighted least squares (IRLS) method commonly used for mlogit model inference (McCullagh and Nelder, 1989) can be used to estimate the annotation coefficients $\boldsymbol{b}_k$. However, IRLS is computational inefficient for large m. Therefore, we instead use the Newton-Raphson algorithm based on the computational strategy presented in (Hasan, et al., 2016) for optimization of Equation (18). This Newton-Raphson algorithm takes full advantage of the sparse structure of the Hessian matrix in the mlogit model, and is thus computationally efficient even when m is in the order of a million. In addition, the Newton-Raphson algorithm naturally provides the observed Fisher information matrix (i.e. the negative Hessian matrix) that can be used to further compute the standard errors for the annotation coefficient estimates.

With the growing number of SNP annotations nowadays, however, it becomes increasingly challenging to model all annotations in the above mlogit model. Examining one annotation at a time (Kichaev, et al., 2014; Pickrell, 2014) does not take full advantage of the correlation structure among annotations and may fail to properly account for multiple testing issue (Chen, et al., 2016). While including all annotations jointly without any prior assumption may incur heavy computational burden, reduce the degrees of freedom, and lead to a potential loss of power. Here, to handle a large number of annotations, we hypothesize that only a fraction of these annotations are likely informative. To select the small set of informative annotations, we impose a Lasso penalty (Tibshirani, 1996) on the log likelihood given in (18)

$$Q(\boldsymbol{b}) = -L(\boldsymbol{b}) + \lambda \sum_{k=\{11,10,01\}} \sum_{d=1}^{q} |b_{kd}|, \tag{19}$$

where λ is the tuning parameter. Optimizing the penalized log likelihood (19) is computationally challenging both due to the complicated mlogit likelihood and a large number of SNPs. To speed up computation, instead of using the usual gradient descent algorithm in Lasso for mlogit, we first apply the least squares approximation (LSA) to the mlogit likelihood (Wang and Leng, 2007; Zeng, et al., 2014) to obtain an easy to evaluate likelihood function. In particular, we first approximate $L(\boldsymbol{b})$ based on a Taylor expansion at the maximum likelihood estimator $\hat{\boldsymbol{b}}_{\mathrm{ml}}$ up to the second order (Hasan, et al., 2016)

$$-L(\boldsymbol{b}) \approx -L(\hat{\boldsymbol{b}}_{\mathrm{ml}}) - (\boldsymbol{b}-\hat{\boldsymbol{b}}_{\mathrm{ml}})^T L'(\hat{\boldsymbol{b}}_{\mathrm{ml}}) - \frac{1}{2}(\boldsymbol{b}-\hat{\boldsymbol{b}}_{\mathrm{ml}})^T L''(\hat{\boldsymbol{b}}_{\mathrm{ml}})(\boldsymbol{b}-\hat{\boldsymbol{b}}_{\mathrm{ml}}), \tag{20}$$

where L' and L'' are the first and second derivatives of the log likelihood function of $L(\boldsymbol{b})$, respectively. Because $\hat{\boldsymbol{b}}_{\mathrm{ml}}$ is the maximum likelihood estimator, $L(\hat{\boldsymbol{b}}_{\mathrm{ml}})$ is a constant and $L'(\hat{\boldsymbol{b}}_{\mathrm{ml}})$ equals zero. Therefore, the penalized log likelihood function Q in (19) is approximated by

$$-\frac{1}{2}(\boldsymbol{b}-\hat{\boldsymbol{b}}_{\mathrm{ml}})^T L''(\hat{\boldsymbol{b}}_{\mathrm{ml}})(\boldsymbol{b}-\hat{\boldsymbol{b}}_{\mathrm{ml}}) + \lambda \sum_{k=\{11,10,01\}} \sum_{d=1}^{q+1} |b_{kd}|. \tag{21}$$

The above approximation is referred to the least squares approximation (Wang and Leng, 2007) or the Laplace approximation in general context. Let $\mathrm{M} = \{-L''(\hat{\boldsymbol{b}}_{\mathrm{ml}})\}^{-1}$ be the observed variance-covariance matrix of $\hat{\boldsymbol{b}}_{\mathrm{ml}}$. We denote $\mathbf{X}^* = \mathrm{M}^{-1/2}$, $\mathbf{Y}^* = \mathrm{M}^{-1/2}\hat{\boldsymbol{b}}_{\mathrm{ml}}$, and re-express Equation (21) as

$$Q(\boldsymbol{b}) \approx \frac{1}{2}(\mathbf{Y}^* - \mathbf{X}^*\boldsymbol{b})'(\mathbf{Y}^* - \mathbf{X}^*\boldsymbol{b}) + \lambda \sum_{k=\{11,10,01\}} \sum_{d=1}^{q+1} |b_{kd}|, \tag{22}$$

which is in the form of the familiar standard Lasso regression. Therefore, we can use standard algorithms, such as the least angle regression algorithm (Efron, et al., 2004) or the coordinate descent algorithm (Friedman, et al., 2007; Friedman, et al., 2010), to obtain $\hat{\boldsymbol{b}}(\lambda) = \arg\min\{Q(\boldsymbol{b})\}$. Afterwards, following (Wang and Leng, 2007), we use a BIC-type criterion to select the optimal tuning parameter λ as in (22).

8. False discovery rate control

After we obtain the estimated parameters from iMAP, for every SNP in turn, we compute the local false discovery rate (locfdr) following (Efron, 2007; Efron, 2008; Efron, et al., 2001). locfdr is closely related to the mixture probability φ, and we use locfdr both to prioritize SNP associations, and in the real data, to identify a genome-wide significance threshold for declaring significance. To do so, we first rely on the definition of locfdr (Efron, et al., 2001) and compute four quantities: (i) locfdr_{10}, which represents the evidence that the SNP j is associated with the first trait; (ii) locfdr_{01}, which represents the evidence that the SNP j is associated with the second trait; (iii) locfdr_{11}, which represents the evidence that the SNP j is associated with both traits; and (iv) locfdr_{00}, which represents the evidence that the SNP j is not associated with any trait. The four quantities are computed as follows:

$$\begin{aligned}
\mathrm{locfdr}_{j10} &= \frac{\pi_{00}f_{00}+\pi_{01}f_{01}}{\pi_{11}f_{11}+\pi_{10}f_{10}+\pi_{01}f_{01}+\pi_{00}f_{00}} = \varphi_{j00}+\varphi_{j01},\\
\mathrm{locfdr}_{j01} &= \frac{\pi_{00}f_{00}+\pi_{10}f_{10}}{\pi_{11}f_{11}+\pi_{10}f_{10}+\pi_{01}f_{01}+\pi_{00}f_{00}} = \varphi_{j00}+\varphi_{j10},\\
\mathrm{locfdr}_{j11} &= \frac{\pi_{00}f_{00}+\pi_{10}f_{10}+\pi_{01}f_{01}}{\pi_{11}f_{11}+\pi_{10}f_{10}+\pi_{01}f_{01}+\pi_{00}f_{00}} = \varphi_{j00}+\varphi_{j10}+\varphi_{j01},\\
\mathrm{locfdr}_{j00} &= \frac{\pi_{00}f_{00}}{\pi_{11}f_{11}+\pi_{10}f_{10}+\pi_{01}f_{01}+\pi_{00}f_{00}} = \varphi_{j00},
\end{aligned} \tag{23}$$

where f_k is the probability density of the distribution component in the Gaussian mixture prior given in Equation (2). The densities f_k and the mixture proportions π_k can be estimated as by-products from our Algorithms 1 and 2. We can plug-in the estimates for f_k and π_k to Equation (23) to obtain estimates for the four locfdr.

We then use locfdr_k (k = 11, 10 and 01) to estimate the false discovery rate (FDR) (Benjamini and Hochberg, 1995) by using the direct posterior probability approach (Newton, et al., 2004). To do so, for each k, we first sort locfdr from the smallest to the largest, where the jth ordered value is $\mathrm{locfdr}_k^{(j)}$ (j = 1, …, m). We then fixed an FDR threshold of α_k (e.g. 0.1%) and identified the cutoff value $\hat{c}_k$ for locfdr that leads to the desired FDR, or

$$\hat{c}_k = \arg\max_{c_k}\{\frac{1}{L}\sum_{j=1}^{L}(\mathrm{I}(\mathrm{locfdr}_k^{(j)} \le c_k)\,\mathrm{locfdr}_k^{(j)}) \le \alpha_k\}, \tag{24}$$

where I is an indicator variable that equals to one when the condition is true and zero otherwise. Following (Chung, et al., 2014), we declare SNPs with at least one locfdr_k that is less than the identified threshold $\tilde{c}_k$ to be significant associations. We also declare SNPs with an locfdr_{11} that is less than the identified threshold $\tilde{c}_{11}$ to be significant pleiotropic associations. Certainly, for a given cutoff value $\tilde{c}_k$, we can also compute the corresponding FDR as

$$\mathrm{FDR}(\tilde{c}_k) = \frac{1}{R}\sum_{j=1}^{R}\mathrm{locfdr}_k^{(j)}\ \mathrm{I}(\mathrm{locfdr}_k^{(j)} \le \tilde{c}_k),\ R = \sum_{j=1}^{m}\mathrm{I}(\mathrm{locfdr}_k^{(j)} \le \tilde{c}_k)\,. \tag{25}$$

9. Compare with other existing methods

Besides iMAP, we also examine the following methods:

(i) univariate analysis, denoted as *univariate*, where we analyzed one SNP at a time. For each SNP, we obtained the marginal z scores for the two traits and then applied locfdr (R package version 1.1-8) (Efron, 2007) on the z scores across SNPs to identify associations at a fixed FDR. The univariate analysis does not incorporate SNP annotation patterns nor account for phenotypic correlation between the two traits.

(ii) gwas-pw (Pickrell, et al., 2016), denoted as *gwas-pw*, where we applied the software gwas-pw (https://github.com/joepickrell/gwas-pw) to analyze marginal variances of effects and z scores from the two traits. gwas-pw examines independent linkage disequilibrium (LD) blocks (Berisa and Pickrell, 2016) and classifies them into five groups based on SNP association pattern: (a) none of the SNPs inside the block are associated with either trait; (b) one SNP inside the block is associated with the first trait but not the second; (c) one SNP inside the block is associated with the second trait but not the first; (d) one SNP inside the block is associated with both traits; and (e) two SNPs in LD are each associated with one trait. Note that each block contains one SNP in our GPA based simulations, while contains approximately 30 SNPs in our gwas-pw based simulations. We used the posterior probability output from gwas-pw to estimate FDR.

(iii) GPA (Chung, et al., 2014), where we used the software GPA (https://github.com/dongjunchung/GPA) to analyze marginal p values from the two traits. We used the estimated local false discover rate values from GPA to prioritize SNPs and compute power and FDR. However, unlike iMAP, GPA does not output the estimated annotation coefficients that represent the importance of annotations.

Text S2

Simulations and Real Data Processing

1. Simulation Designs

We conducted a range of simulations to evaluate the performance of iMAP and compared it with existing methods. For simulations, we obtained genotypes from the Kaiser Permanente/UCSF Genetic Epidemiology Research Study on Adult Health and Aging (GERA; dbGaP accession number phs000674.v2.p2) (Banda, et al., 2015; Kvale, et al., 2015). The raw data of the GERA study consists of 62,313 individuals and 675,367 SNPs. We filtered out SNPs that had a missingness percentage greater than 0.95 across individuals, genotype calling rate below 0.95, minor allele frequency (MAF) greater than 0.01, or Hardy-Weinberg equilibrium test p value smaller than 10^{-4}. A total of 487,609 SNPs retained after filtering. The missing values of SNP were imputed with the mean of that SNP. Following (Chung, et al., 2014), we used plink (Purcell, et al., 2007) (version 1.90b3.38) to prune the genotypes (with plink command "*--indep-pairwise 100 100 0.01*") and obtained m=15,495 approximately independent SNPs. We also randomly selected n=10,000 individuals and used genotypes for this set of individuals to simulate pairs of phenotypes. We considered a range of simulation settings described below and performed 100 simulation replicates in each simulation setting.

In the simulations, we selected 500 SNPs to be causal for each trait. To allow for pleiotropic effects, these causal SNPs were selected in a way that ensures a desired proportion of pleiotropic SNPs (0%, 20%, 40%, 60%, 80% or 100%). For example, in the case of 20% pleiotropic SNPs, we first randomly selected 100 SNPs to have nonzero effects for both traits. We then randomly selected 400 to have nonzero effects on the first trait and another 400 to have nonzero effects on the second trait; thus, a total of 900 SNPs had nonzero effects on at least one trait. We simulated each nonzero effect of the causal SNPs independently from a standard normal distribution. Therefore, the effects of pleiotropic SNPs on the two traits are also independent as in previous studies (Chung, et al., 2014). While we only present results based on independent effects, we have examined scenarios where the pleiotropic SNPs have correlated effects but found effect correlation to have a minor influence on method comparison. After simulating the causal effects, we scaled the simulated effects further so that the proportion of phenotypic variance explained (PVE) by all causal SNPs for each trait was 0.5 (i.e. heritability of each trait equals 0.5). We then simulated residual errors from a bivariate normal distribution with mean zero, variance one, and covariance varied in the range of -0.8 to 0.8 (including 0) to evaluate the influence of phenotypic correlation on the performance of various methods. We summed genetic effects with residual errors to form the simulated phenotypes. We finally quantile-normalized the phenotypes for each trait to a standard normal distribution before analysis.

After we simulated phenotypes, we paired them with genotypes to obtain summary statistics. Specifically, for each trait in turn, we examined one SNP at a time and used a linear regression model to obtain a set of marginal statistics that include variances of the effect size estimates, z scores and p values. With these summary statistics, we fitted the following methods (details in the next section): (i) univariate analysis that relies on marginal z scores; (ii) gwas-pw that uses marginal z scores and variances of estimated effect sizes; (iii) GPA that uses marginal p values; (iv) iMAP that uses marginal z scores. Note that only GPA and iMAP model SNP annotations, and only iMAP models phenotypic correlation. Because iMAP relies on the phenotype covariance matrix $\mathbf{\Sigma}$ to account for phenotypic correlation, it also requires an estimate of $\mathbf{\Sigma}$. To obtain such an estimate, we followed previous studies (Liu and Lin, 2017; Stephens, 2013; Zhu, et al., 2015) to obtain a set of null SNPs that have marginal z scores (in absolute value) for both traits below two. We then computed the sample covariance matrix of the z scores for the two traits across the null SNPs as an estimate of $\mathbf{\Sigma}$.

To examine the benefits of incorporating annotations, following (Chung, et al., 2014), we simulated two sets of SNP annotations that include an informative set and an noninformative set. The values for the informative annotations are dependent on the causality of SNPs, while the values of the noninformative annotations are independent of SNP causality. For the tth informative annotation, we simulated the SNP annotation values from a normal distribution with variance one and a mean that depends on SNP causality: mean = m_t for causal SNPs and mean = 0 for noncausal SNPs, where m_t is the same across all SNPs and is randomly (with respect to t) set to be either 0.5 or -0.5. The mean values of 0.5 or -0.5 were selected to ensure a reasonably high power in the simulations. For noninformative annotations, we directly simulated annotations from a standard normal distribution.

We performed two primary sets of simulations with regard to the use of annotations. The first set involves a small number of annotations. Here, we considered four sub-scenarios that include different numbers of informative annotations: (i) no annotations; (ii) one annotation; (iii) two annotations; (iv) four annotations. Because GPA can only handle binary annotations, we transformed continuous annotations into binary annotations with a cutoff value of zero (which is optimal under our simulation setting). For iMAP, we performed analyses with either continuous annotations or the transformed binary annotations. The second set involves a large number of annotations. Here, we simulated 4 informative annotations and 100 noninformative annotations. We considered three sub-analyses with iMAP: (i) iMAP, which is a standard analysis that uses iMAP with Lasso-based selection to analyze all 104 annotations; (ii) iMAP-oracle, which is an oracle analysis that uses iMAP to analyze only the four informative annotations without Lasso-based selection; (iii) iMAP-full, which is a full analysis that uses iMAP to analyze all 104 annotations without selection.

While our main simulations followed that described in GPA (Chung, et al., 2014), we also performed an alternative set of simulations that follow the simulation setting described in gwas-pw (Pickrell, et al., 2016). Specifically, we divided 15,495 SNPs into 500 equal-size regions with approximately 30 SNPs in each region. We randomly selected 60% of the regions to be causal. In each causal region, we randomly selected two causal SNPs to have nonzero effects. These two causal SNPs have equal probability (= 1/3) to affect only the first trait, or the second trait, or both traits. We simulated causal effects and residual errors in the same way as above and summed genetic effects with residuals to form the simulated phenotypes.

2. Evaluating the influence of LD

Besides simulations with independent SNPs, we also performed a series of simulations using correlated SNPs to evaluate the robustness of various methods under linkage disequilibrium (LD). To do so, we used the same 10,000 individuals from GERA and obtained m = 79,979 genotypes from the first two chromosomes without pruning. We used simulation settings described in the main text and simulated phenotypes. Like in the main text, we performed 100 simulation replicates for each simulation setting. A key challenge to evaluate the performance of different methods in terms of association mapping power in the presence of correlated SNPs is the definition of "causal" SNPs. In particular, because SNPs are in LD, neighboring SNPs of the truly causal SNPs would also display association signal and identifying these neighboring SNPs can help pinpoint causal locus. Therefore, for power comparison, we define SNPs as "causal" if they are within a given distance of the truly causal SNPs. We examine a range of distance cutoffs (0 to 1000kb) in the simulations.

Regardless of the distance cutoff, the power comparison results with correlated SNPs are qualitatively similar compared to the early simulations with independent SNPs, though the power gain of iMAP over GPA becomes smaller (Figure S16). However, the estimated FDR from all methods depends heavily on the distance cutoff we use (Figure S17): the estimated FDR is overly liberal when the distance cutoff is within 300 kb, is approximately calibrated when the distance cutoff is in the range of 300 - 500 kb, and is overly conservative when the distance cutoff is beyond 500 kb. LD also influences causal proportion estimation. In particular, all three methods (gwas-pw, GPA and iMAP) overestimate π_{11} and π_{10} (and π_{01}) in the presence of LD (Figure S18). In addition, depending on the relative overestimation on these proportions, different methods can show bias and variance for estimating $\pi_{11}/(\pi_{10}+\pi_{11})$ (Figure S18). Specifically, gwas-pw and iMAP produce highly variable estimates that are sometimes underestimated and sometimes overestimated. In contrast, GPA tends to overestimate $\pi_{11}/(\pi_{10}+\pi_{11})$ when the proportion of pleiotropic SNPs is small (e.g. less than 20%) and underestimate it when the proportion of pleiotropic SNPs is large (e.g. greater than 20%). Overall, estimating proportion of causal SNPs is challenging with existing mixture models in the presence of correlated SNPs.

3. Data Processing for Real Data

We applied the above methods to analyze 48 traits from 31 GWAS consortium studies. These traits span a wide range of phenotypes that include anthropometric traits (e.g. height and BMI), hematopoietic traits (e.g. MCHC and RBC), immune diseases (e.g. CD and IBD) and neurological diseases (e.g. Alzheimer's disease and schizophrenia). We obtained these GWAS data from public websites (Table S1) and used marginal p values for GPA, variances of effect size estimates and z scores for gwas-pw, or z scores for iMAP and *univariate* analysis. We retained SNPs that have a MAF larger than 0.05 in the 503 individuals of European ancestry from the 1000 Genomes Project (The 1000 Genomes Project Consortium, 2015). We further removed the major histocompatibility complex (MHC, Chr6: 26 ~ 34Mb) region following (Finucane, et al., 2015). Our final analyses are focused on a set of 652,356 SNPs that are shared across all data sets. Table S1 lists all GWAS data sets used in the present study.

Specifically, GWAS data for FG, H2G and HOMA B are from the Meta-Analyses of Glucose and Insulin-related traits (MAGIC) Consortium (www.magicinvestigators.org). Data for Height and BMI2 are from the Genetic Investigation of ANthropometric Traits (GIANT) consortium (https://portals.broadinstitute.org). Data for T2D are from the DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) consortium (http://diagram-consortium.org.html). Data for HDL, LDL, TC and TG are from the Global Lipids Genetics Consortium (http://csg.sph.umich.edu). Data for CAD are from the CARDIoGRAMplusC4D consortium (www.cardiogramplusc4d.org). Data for heart rate are from the HRgene consortium (https://walker05.u.hpc.mssm.edu/). Data for Menarche and Menopause are from the ReproGen consortium (http://www.reprogen.org). Data for BMI1, BW2, Growth 10, Growth PG and Obesity are from the Early Growth Genetics (EGG) consortium (http://egg-consortium.org/). Data for Alzheimer's disease are from the International Genomics of Alzheimer's Project (http://web.pasteur-lille.fr). Data for Anorexia, DS, SCZ, BIPSCZ, BIP, Autism, CPD and Ever Smoked are from the

Psychiatric Genomics Consortium (https://www.med.unc.edu/pgc). Data for Neuroticism, YE1 and YE2 are from the Social Science Genetic Association Consortium (https://www.thessgac.org/). Data for UC, IBD and CD are from the International Inflammatory Bowel Disease Genetics Consortium (https://www.ibdgenetics.org/). Data for FNBMD and LSBMD are from the GEnetic Factors for OSteoporosis Consortium (http://www.gefos.org/). Data for MCHC, MCH, HB, MCV, MPV, PCV, PLT and RBC are from the Blood Cell Consortium (http://www.mhi-humangenetics.org/).

We extracted SNP annotations based on genome-wide histone occupancy of four histone marks (*H3K27me3*, *H3K36me3*, *H3K4me1* and *H3K4me3*) from 105 tissues in the Roadmap Epigenomics Project (Roadmap Epigenomics Consortium, et al., 2015). These four histone marks were selected instead of the original six because the four have few missing values across tissues. We grouped the 105 tissues into 10 tissue categories (i.e. Blood/Immune, Adipose, Adrenal/Pancreas, Bone/Connective, Cardiovascular, central nervous system (CNS), Gastrointestinal, Liver, Muscle and Other) based on anatomy type following previous studies (Finucane, et al., 2015; Lu, et al., 2016; Roadmap Epigenomics Consortium, et al., 2015). For each tissue type and each histone mark, we created a binary annotation to indicate whether a given SNP resides inside the peak regions of the histone mark. In addition, for each tissue group and each histone mark in turn, we averaged the binary annotations across all tissue types within that tissue group and generated a new, continuous histone annotation at the tissue group level. We used these 40-tissue group level histone annotations in the present study. We applied the same four methods described in the previous section to analyze the data.

References

Banda, Y., *et al.* (2015) Characterizing race/ethnicity and genetic ancestry for 100,000 subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort, *Genetics*, **200**, 1285-1295.

Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing, *J. R. Stat. Soc. Ser. B.*, **57**, 289-300.

Berisa, T. and Pickrell, J.K. (2016) Approximately independent linkage disequilibrium blocks in human populations, *Bioinformatics*, **32**, 283-285.

Bulik-Sullivan, B., *et al.* (2015) An atlas of genetic correlations across human diseases and traits, *Nat. Genet.*, **47**, 1236-1241.

Carbonetto, P. and Stephens, M. (2013) Integrated enrichment analysis of variants and pathways in genome-wide association studies indicates central role for IL-2 signaling genes in type 1 diabetes, and cytokine signaling genes in Crohn's disease, *PLoS Genet.*, **9**, e1003770.

Chen, W., *et al.* (2016) Incorporating Functional Annotations for Fine-Mapping Causal Variants in a Bayesian Framework Using Summary Statistics, *Genetics*, **204**, 933-958.

Chung, D., *et al.* (2014) GPA: a statistical approach to prioritizing GWAS results by integrating pleiotropy and annotation, *PLoS Genet.*, **10**, e1004787.

Chung, Y., *et al.* (2015) Weakly informative prior for point estimation of covariance matrices in hierarchical models, *Journal of Educational and Behavioral Statistics*, **40**, 136-157.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) Maximum Likelihood from Incomplete Data via the EM Algorithm, *J. R. Stat. Soc. Ser. B.*, **39**, 1-38.

Efron, B. (2007) Correlation and Large-Scale Simultaneous Significance Testing, *J. Am. Stat. Assoc.*, **102**, 93-103.

Efron, B. (2007) Size, power and false discovery rates, *Ann. Stat*, **35**, 1351-1377.

Efron, B. (2008) Microarrays, empirical Bayes and the two-groups model, *Stat. Sci.*, **23**, 1-22.

Efron, B., *et al.* (2004) Least angle regression, *Ann. Stat*, **32**, 407-499.

Efron, B., *et al.* (2001) Empirical Bayes analysis of a microarray experiment, *J. Am. Stat. Assoc.*, **96**, 1151-1160.

Finucane, H.K., *et al.* (2015) Partitioning heritability by functional annotation using genome-wide association summary statistics, *Nat. Genet.*, **47**, 1228-1235.

Friedman, J., *et al.* (2007) Pathwise coordinate optimization, *Ann. Appl. Stat.*, **1**, 302-332.

Friedman, J., Hastie, T. and Tibshirani, R. (2010) Regularization Paths for Generalized Linear Models via Coordinate Descent, *J. Stat. Softw.*, **33**, 1-22.

Gelman, A., *et al.* (2008) A weakly informative default prior distribution for logistic and other regression models, *Ann. Appl. Stat.*, **2**, 1360-1383.

Hasan, A., Zhiyu, W. and Mahani, A.S. (2016) Fast Estimation of Multinomial Logit Models: R Package mnlogit, *J. Stat. Softw.*, **75**, 1-24.

Kenne Pagui, E.C., Salvan, A. and Sartori, N. (2015) On full efficiency of the maximum composite likelihood estimator, *Statistics & Probability Letters*, **97**, 120-124.

Kichaev, G., *et al.* (2014) Integrating Functional Data to Prioritize Causal Variants in Statistical Fine-Mapping Studies, *PLoS Genet.*, **10**, e1004722.

Kvale, M.N., *et al.* (2015) Genotyping informatics and quality control for 100,000 subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort, *Genetics*, **200**, 1051-1060.

Larribe, F. and Fearnhead, P. (2011) On composite likelihoods in statistical genetics, *Stat. Sinica*, **21**, 43-69.

Liu, Z. and Lin, X. (2017) Multiple phenotype association tests using summary statistics in genome-wide association studies, *Biometrics*, 10.1111/biom.12735.

Lu, Q., *et al.* (2016) Integrative tissue-specific functional annotations in the human genome provide novel insights on many complex traits and improve signal prioritization in genome wide association studies, *PLoS Genet.*, **12**, e1005947.

McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models*. Chapman and Hall, London.

Moser, G., *et al.* (2015) Simultaneous Discovery, Estimation and Prediction Analysis of Complex Traits Using a Bayesian Mixture Model, *PLoS Genet.*, **11**, e1004969.

Newton, M.A., *et al.* (2004) Detecting differential gene expression with a semiparametric hierarchical mixture method, *Biostatistics*, **5**, 155-176.

Pasaniuc, B. and Price, A.L. (2017) Dissecting the genetics of complex traits using summary association statistics, *Nat. Rev. Genet.*, **18**, 117-127.

Pickrell, Joseph K. (2014) Joint Analysis of Functional Genomic Data and Genome-wide Association Studies of 18 Human Traits, *Am. J. Hum. Genet.*, **94**, 559-573.

Pickrell, J.K., *et al.* (2016) Detection and interpretation of shared genetic influences on 42 human traits, *Nat. Genet.*, **48**, 709-717.

Purcell, S., *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses, *Am. J. Hum. Genet.*, **81**, 559-575.

Roadmap Epigenomics Consortium, *et al.* (2015) Integrative analysis of 111 reference human epigenomes, *Nature*, **518**, 317-330.

Shi, H., Kichaev, G. and Pasaniuc, B. (2016) Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data, *Am. J. Hum. Genet.*, **99**, 139-153.

Speed, D. and Balding, D.J. (2014) MultiBLUP: improved SNP-based prediction for complex traits, *Genome Res.*, **24**, 1550-1557.

Stephens, M. (2013) A unified framework for association analysis with multiple related phenotypes, *PLoS ONE*, **8**, e65245.

The 1000 Genomes Project Consortium (2015) A global reference for human genetic variation, *Nature*, **526**, 68-74.

Tibshirani, R. (1996) Regression shrinkage and selection via the LASSO, *J. R. Stat. Soc. Ser. B.*, **58**, 267-288.

Varin, C. (2008) On composite marginal likelihoods, *AStA Adv Stat Anal*, **92**, 1.

Varin, C., Reid, N. and Firth, D. (2011) An overview of composite likelihood methods, *Stat. Sinica*, **21**, 5-42.

Vilhjálmsson, B.J., *et al.* (2015) Modeling linkage disequilibrium increases accuracy of polygenic risk scores, *Am. J. Hum. Genet.*, **97**, 576-592.

Visscher, P.M., *et al.* (2017) 10 Years of GWAS Discovery: Biology, Function, and Translation, *Am. J. Hum. Genet.*, **101**, 5-22.

Wang, H. and Leng, C. (2007) Unified LASSO estimation by least squares approximation, *J. Am. Stat. Assoc.*, **102**, 1039-1048.

Weissbrod, O., Geiger, D. and Rosset, S. (2016) Multikernel: linear mixed models for complex phenotype prediction, *Genome Res.*, **26**, 969-979.

Wen, X.*, et al.* (2016) Efficient Integrative Multi-SNP Association Analysis via Deterministic Approximation of Posteriors, *Am. J. Hum. Genet.*, **98**, 1114-1129.

Wen, X., Luca, F. and Pique-Regi, R. (2015) Cross-Population Joint Analysis of eQTLs: Fine Mapping and Functional Annotation, *PLoS Genet.*, **11**, e1005176.

Zeng, P.*, et al.* (2014) Variable selection approach for zero-inflated count data via adaptive lasso, *Journal of Applied Statistics*, **41**, 879-894.

Zhou, X., Carbonetto, P. and Stephens, M. (2013) Polygenic modeling with Bayesian sparse linear mixed models, *PLoS Genet.*, **9**, e1003264.

Zhu, X.*, et al.* (2015) Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension, *Am. J. Hum. Genet.*, **96**, 21-36.

Zhu, X. and Stephens, M. (2017) Bayesian large-scale multiple regression with summary statistics from genome-wide association studies, *Ann. Appl. Stat.*, **11**, 1561-1592.