Bioinformatics, 34(16), 2018, 2797–2807
doi: 10.1093/bioinformatics/bty204
Advance Access Publication Date: 4 April 2018
Original Paper

OXFORD

Genetics and population analysis

Pleiotropic mapping and annotation selection in genome-wide association studies with penalized Gaussian mixture models

Ping Zeng[1,2,3], Xingjie Hao[2,3] and Xiang Zhou[2,3,*]

[1]Department of Epidemiology and Biostatistics, Xuzhou Medical University, Xuzhou, Jiangsu 221004, China, [2]Department of Biostatistics and [3]Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA

*To whom correspondence should be addressed.

Associate Editor: Oliver Stegle

Received on September 26, 2017; revised on January 30, 2018; editorial decision on March 22, 2018; accepted on April 2, 2018

Abstract

Motivation: Genome-wide association studies (GWASs) have identified many genetic loci associated with complex traits. A substantial fraction of these identified loci is associated with multiple traits—a phenomena known as pleiotropy. Identification of pleiotropic associations can help characterize the genetic relationship among complex traits and can facilitate our understanding of disease etiology. Effective pleiotropic association mapping requires the development of statistical methods that can jointly model multiple traits with genome-wide single nucleic polymorphisms (SNPs) together.

Results: We develop a joint modeling method, which we refer to as the **i**ntegrative **MA**pping of **P**leiotropic association (iMAP). iMAP models summary statistics from GWASs, uses a multivariate Gaussian distribution to account for phenotypic correlation, simultaneously infers genome-wide SNP association pattern using mixture modeling and has the potential to reveal causal relationship between traits. Importantly, iMAP integrates a large number of SNP functional annotations to substantially improve association mapping power, and, with a sparsity-inducing penalty, is capable of selecting informative annotations from a large, potentially non-informative set. To enable scalable inference of iMAP to association studies with hundreds of thousands of individuals and millions of SNPs, we develop an efficient expectation maximization algorithm based on an approximate penalized regression algorithm. With simulations and comparisons to existing methods, we illustrate the benefits of iMAP in terms of both high association mapping power and accurate estimation of genome-wide SNP association patterns. Finally, we apply iMAP to perform a joint analysis of 48 traits from 31 GWAS consortia together with 40 tissue-specific SNP annotations generated from the Roadmap Project.

Availability:
and implementation: iMAP is freely available at http://www.xzlab.org/software.html.
Contact: xzhousph@umich.edu
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Genome-wide association studies (GWASs) have identified thousands of genetic variants associated with many complex diseases and traits (MacArthur *et al.*, 2017). A substantial fraction of these identified loci often display association with more than one trait—a phenomenon known as pleiotropy (Solovieff *et al.*, 2013). Notable examples of pleiotropic genes include *PTPN22* and *HLA-DRB1* that are associated with several different autoimmune disorders (The Wellcome Trust Case Control Consortium, 2007); *CLPTM1L* with multiple types of cancers (Fletcher and Houlston, 2010); *CACNA1C* with bipolar disorder and schizophrenia (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013b); *ANGPTL3*, *TIMD4*, *PPP1R3B* and *TRIB1* with multiple plasma lipid traits (Teslovich *et al.*, 2010); *MYL2* with various metabolic traits (Kim *et al.*, 2011); *ZFAT* with total cholesterol, blood pressure and cardiovascular disease (Smith *et al.*, 2010; Warren *et al.*, 2017); and *ABO* with blood group, various blood measurements and coronary artery disease (Bulik-Sullivan *et al.*, 2015; Pickrell *et al.*, 2016). Overall, it has been estimated that 4.6% of the previously identified associated variants and 16.9% of the previously identified associated genes show pleiotropic effects (Sivakumaran *et al.*, 2011). The percentage of pleiotropic association is often much higher in biologically relevant traits. For example, about 44–70% of genetic variants that are associated with one autoimmune disease are also associated with another (Cotsapas *et al.*, 2011; Jostins *et al.*, 2012; Solovieff *et al.*, 2013). In addition, joint heritability analyses of multiple traits have also revealed substantial genetic correlation among various psychiatric disorders (Bulik-Sullivan *et al.*, 2015; Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013a), between schizophrenia and brain anatomic measurements (Lee *et al.*, 2016) or amyotrophic lateral sclerosis (McLaughlin *et al.*, 2017), between neuropsychiatric disorders and several metabolic traits (Lane *et al.*, 2017), as well as among multiple autoimmune diseases (Ji *et al.*, 2017; Zhernakova *et al.*, 2009).

Because of the prevalence and importance of pleiotropy, many statistical methods have been developed to identify genetic variants that are associated with multiple phenotypes (Li and Kellis, 2016; Liu *et al.*, 2016; Solovieff *et al.*, 2013; van der Sluis *et al.*, 2013; Zhou and Stephens 2014). Most of these methods rely on multivariate models to analyze multiple traits jointly. Unlike the commonly used univariate methods that examine one trait at a time, multivariate methods analyze multiple traits together while properly accounting for the phenotypic correlation among them. As a result, multivariate methods often substantially improve association mapping power compared with univariate methods. For example, compared with univariate analysis, multivariate analysis achieves ~50% power gain in mapping systolic blood pressure and other comorbid traits (Andreassen *et al.*, 2014) as well as in mapping multiple correlated blood lipid traits (Stephens, 2013; Zhou and Stephens, 2014). Applications of multivariate methods in GWASs have identified a large number of pleiotropic associations (He *et al.*, 2013; Solovieff *et al.*, 2013; Van der Sluis *et al.*, 2015; Zhu *et al.*, 2015).

Among the existing methods for pleiotropic association mapping, mixture models have attracted substantial recent attention. Mixture models attempt to classify genome-wide single nucleic polymorphisms (SNPs) into different categories based on their effect sizes on the jointly analyzed traits. For example, the Bayesian co-localization test (Giambartolomei *et al.*, 2014) and its recent extension and implementation in the computational tool, gwas-pw (Pickrell *et al.*, 2016) (pair-wise traits analysis of GWAS), analyzes pairs of traits jointly. In its simplified version, gwas-pw divides SNPs into those that are associated with both traits (i.e. pleiotropic associations), associated with only one trait and associated with neither. By partitioning SNPs into different association categories and inferring the genome-wide association pattern, mixture models can improve association mapping power (Andreassen *et al.*, 2014; Giambartolomei *et al.*, 2014). In addition, by computing the proportion of SNPs in each category, mixture models have the potential to provide evidence supporting potential directional causality between the two analyzed phenotypes (Pickrell *et al.*, 2016). Specifically, if SNPs associated with the first trait are also associated with the second trait, but not vice versa, then the first trait may causally influence the second trait. Therefore, mixture model methods can facilitate the identification and interpretation of pleiotropic associations and can help better characterize the genetic relationship among traits (Chen *et al.*, 2016a; Ernst and Kellis, 2010; Ernst *et al.*, 2011; Fu *et al.*, 2014; Ionita-Laza *et al.*, 2016; Lonsdale *et al.*, 2013; Lu *et al.*, 2016).

Mixture models are recently extended to integrate external SNP functional annotations to further improve power of pleiotropic mapping. For example, genetic analysis incorporating pleiotropy and annotation (GPA) (Chung *et al.*, 2014) and its gene-set analysis extension empirical Bayes approach for integrating pleiotropy and tissue-specific information (EPS) (Liu *et al.*, 2016) extend mixture models by jointly modeling SNP association pattern together with the distribution of the SNP functional annotations. SNP functional annotations are developed to characterize the function of genetic variants (Dixon *et al.*, 2015; Kellis, *et al.*, 2014; Lonsdale *et al.*, 2013). For example, we can now classify genetic variants based on their genomic location (e.g. coding, intron and intergenic variants), role in protein structure and function [e.g. sorting tolerant from intolerant (SIFT) score (Kumar *et al.*, 2009) or polymorphism phenotyping (PolyPhen) score (Adzhubei *et al.*, 2013)], ability to regulate gene expression [e.g. expression quantitative trait loci (eQTL) and allele specific expression (ASE) evidence (Pickrell *et al.*, 2010; Tung *et al.*, 2015)], biochemical function [e.g. DNase I hypersensitive sites, metabolomic QTL evidence and chromatin states (Ernst and Kellis, 2012; McVicker *et al.*, 2013; Pique-Regi *et al.*, 2011)], evolutionary significance [e.g. genomic evolutionary rate profiling (GERP) score (Cooper *et al.*, 2005)] and/or a combination of all these annotations [e.g. combined annotation dependent depletion (CADD) score (Kircher *et al.*, 2014) and Eigen score (Ionita-Laza *et al.*, 2016)]. These functional annotations can be important predictors for SNP effects. Studies have shown that SNPs in certain functional categories (e.g. in promoters and enhancers) are more likely to be causal (Pickrell, 2014; Schork *et al.*, 2013), tend to have larger effect sizes and explain more heritability than SNPs in other categories (e.g. introns) (Gusev *et al.*, 2014; Kichaev *et al.*, 2014). Therefore, by integrating functional annotations into pleiotropic association mapping, GPA can identify enriched functional annotations for each SNP association category, with which to further improve association mapping power (Chung *et al.*, 2014).

Despite the promising results from mixture models for pleiotropic association mapping, existing mixture methods have two important limitations. First, perhaps rather surprisingly, existing mixture methods do not account for phenotypic correlation among jointly analyzed traits and effectively assume phenotypic independence. Given that explicit modeling of phenotypic correlation is the key for effective association mapping and phenotype/risk prediction of pleiotropic traits (Hu *et al.*, 2017; Maier *et al.*, 2015; Stephens, 2013), assuming phenotype independence likely reduces the effectiveness of existing mixture models. Second, existing mixture models are only able to integrate binary annotations and can only handle a relatively small number of annotations. Analyzing only a few binary annotations may fail to take advantage of the increasingly large

number of annotations that are being generated today, among which many are continuous (Ionita-Laza *et al.*, 2016; Kircher, *et al.*, 2014; Li and Kellis, 2016; Roadmap Epigenomics Consortium, *et al.*, 2015; The ENCODE Project Consortium, 2012).

Here, we developed a novel mixture modeling method, which we refer to as the integrative **MA**pping of **P**leiotropic association (iMAP), for association mapping of pair-wise traits. iMAP relies on a multinomial logistic regression model to incorporate a large number of binary and continuous SNP annotations, and, with a sparsity-inducing penalty term, is capable of selecting a small, informative set of annotations. In addition, iMAP directly models summary statistics from GWASs and uses a multivariate Gaussian distribution to account for phenotypic correlation between traits. As a result, iMAP is capable of integrating both binary and continuous SNP annotations, selecting informative annotations from a large set of potentially non-informative ones and using GWAS summary statistics while simultaneously accounting for phenotypic correlation between traits. We also adopt a recently developed efficient approximation algorithm for penalized regression (Wang and Leng, 2007; Zeng *et al.*, 2014) to enable scalable inference in iMAP. With simulations, we illustrate the benefits of iMAP in terms of both improving association mapping power and accurate estimation of SNP association patterns. Finally, we apply iMAP to analyze 48 traits from 31 GWAS consortium studies together with 40 tissue-specific SNP annotations from the Roadmap epigenomics project (Roadmap Epigenomics Consortium, *et al.*, 2015). iMAP is freely available at http://www.xzlab.org/software.html.

2 Materials and methods

We provide an overview of iMAP here, details are available in Supplementary Text S1. Our main presentation of iMAP will focus on analyzing pairs of traits that are measured on the same set of individuals from a GWAS. Extensions of our method to situations where the two traits come from different GWASs are trivial and are presented in Supplementary Text S1. Extensions of our method to analyzing more than two traits are not straightforward, with potential statistical and computational complications and will be discussed later in Section 4. For each SNP in turn, we consider the following bivariate model:

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

where n is the number of individuals, m is the number of SNPs, $\mathbf{y}_i$ is a two-vector of phenotypes for 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, $\mathbf{e}_{ij}$ is a two-vector of residual errors that follow a bivariate normal distribution with a two by two covariance matrix $\boldsymbol{\Sigma}_j$ and MVN_2 denotes a two-dimensional multivariate normal distribution (i.e. bivariate normal distribution). iMAP naturally accounts for phenotypic correlation by modeling the covariance matrix $\boldsymbol{\Sigma}_j$. 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 is specified on each SNP j. To infer genome-wide pleiotropic association pattern and improve association mapping power, following (Chung *et al.*, 2014; Pickrell *et al.*, 2016), we model all SNPs jointly by assuming that the joint likelihood for all SNPs is simply a product of the likelihood for each SNP. The joint likelihood approximation by the product of individual likelihoods can be justified from a composite likelihood perspective (Larribe and Fearnhead, 2011; Varin *et al.*, 2011) (Supplementary Text S1). To facilitate information sharing across genome-wide SNPs, we assign a common prior on the effect sizes and assume that each $\boldsymbol{\beta}_j$ *a priori* follows a four-component Gaussian mixture distribution:

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

with prior probabilities π_k ($k = 11$, 10, 01 and 00) that sum to 1. Here π_{11} represents the prior probability that a SNP is associated with both traits; $\mathbf{V}_{11} = ((\sigma^2_{11}, \sigma^2_{21})^T, (\sigma^2_{12}, \sigma^2_{22})^T)$ 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; $\mathbf{V}_{10} = ((\sigma^2_1, 0)^T, (0, 0)^T)$ is a low-rank covariance matrix restricting that only the effect size for the first trait is non-zero. π_{01} represents the prior probability that a SNP is associated with the second trait but not the first; $\mathbf{V}_{01} = ((0, 0)^T, (0, \sigma^2_2)^T)$ is a low-rank covariance matrix restricting non-zero 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 $\mathbf{V}$ and we borrow information across genome-wide SNPs to infer these parameters (Supplementary Text S1). The estimated probability of a SNP having non-zero 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 use PIP to prioritize SNPs in this present study. We also use PIP to provide a conservative estimate of false discovery rate (FDR) (Benjamini and Hochberg, 1995) based on local FDR (Efron *et al.*, 2001) via the direct posterior probability approach (Newton *et al.*, 2004). FDR is a common approach applied for multiple testing settings. Effective control of FDR ensures that the proportion of false discoveries among the identified SNP associations is kept in check regardless of the number of tests performed. However, we also note that an FDR of 0.05 can be less stringent than the usual genome-wide P-value threshold of 5×10^{-8} that aims to control for a family-wise error rate (FWER) of $\sim$0.05.

To increase association mapping power and facilitate association results interpretation, we incorporate functional SNP annotations from external data into the above model. To do so, we assume that all SNPs are characterized by the same set of q annotations. For SNP j, we denote $\mathbf{x}_j = (1, x_{j1}, x_{j2}, \ldots, x_{jq})^T$ as a $(q+1)$-vector of annotation values that include a value of 1 to represent the intercept. These annotations can be either discrete or continuous. For example, one annotation could be a binary indicator on whether a SNP resides in exonic regions, while another annotation could be a continuous value of the CADD (Kircher *et al.*, 2014) or Eigen (Ionita-Laza *et al.*, 2016) score for the given SNP. To simplify presentation, we assemble the annotation vectors across all SNPs into an m by $(q+1)$ annotation matrix $\mathbf{X}$, where each row of $\mathbf{X}$ contains the annotation vector for the corresponding SNP. We then link the annotation matrix $\mathbf{X}$ to the mixture probabilities (π_{11}, π_{10}, π_{01} and π_{00}) through a multinomial logit (mlogit) regression model

$$\log(\pi_{jk}) \propto \sum_{d=0}^{q+1} x_{jd} b_{kd} = \mathbf{x}_j \mathbf{b}_k, \quad (3)$$

where $k = 11$, 10, 01 or 00; and each $\mathbf{b}_k = (b_0, b_1, b_2, \ldots, b_q)$ is a $(q+1)$-vector of annotation coefficients that include an intercept. We choose $k = 00$ as the reference category and set $\mathbf{b}_{00} = \mathbf{0}$ to ensure model identifiability. Note that, in the case where SNPs belong to only two categories (e.g. causal versus non-causal), the mlogit model reduces to a logistic regression model that has been commonly used to link functional annotations to SNP causality (Carbonetto and Stephens, 2013; Wen *et al.*, 2015, 2016). We use the 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).

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.*, 2016b). 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. Subsequently, we impose a penalty term $\lambda \sum_k \sum_{d=1}^{q} |b_{kd}|$ onto the log likelihood of model (3) to select the important annotations via Lasso (Tibshirani, 1996) (Supplementary Text S1; λ is the tuning parameter).

We develop an expectation-maximization (EM) algorithm for parameter inference in iMAP. Briefly, we view the mixture group assignment of each SNP as missing data and impute them in the expectation step. In the maximization step, we adopt the recently developed computational strategy for the mlogit model (Hasan *et al.*, 2016) to take full advantage of the sparse structure of the Hessian matrix, in order to estimate the annotation parameters $\mathbf{b}$ in a computationally efficient fashion. In addition, we employ an efficient approximation algorithm for lasso based on the least square approximation (Wang and Leng, 2007; Zeng *et al.*, 2014) to further improve computational efficiency. We rely on the Bayesian information criterion to select informative annotations. Details of our inference algorithm are provided in Supplementary Text S1.

Finally, we note that, while our main presentation of the model and algorithm rely on individual-level phenotypes and genotypes, iMAP can be fitted using only summary statistics in terms of marginal z scores for each phenotype (Supplementary Text S1). For example, the phenotypic covariance matrix $\mathbf{\Sigma}$ can be estimated using the sample covariance of marginal z scores from null SNPs (Stephens, 2013). The EM algorithm can also be fitted using marginal z scores only. Therefore, in the following simulations, while we use individual-level genotypes to simulate phenotypes, we obtain marginal z scores afterward to fit iMAP. In the real data application, we also use marginal z scores to fit iMAP.

3 Results

3.1 Simulation results

3.1.1 Power to identify associated SNPs

We first performed simulations to compare iMAP with other methods in terms of power to detect true associations. Simulations largely follow previous studies (Chung *et al.*, 2014) with details provided in Supplementary Text S2. Briefly, we obtained 15 495 independent SNPs in 10 000 individuals from the Kaiser Permanente/UCSF Genetic Epidemiology Research Study on Adult Health and Aging (GERA) study (Banda *et al.*, 2015). Among these SNPs, we considered 500 to be causal for each trait, generated causal effects independently from a normal distribution and summed simulated residual errors to form the two traits. Due to the small number of individuals, to ensure sufficient power, our main simulations considered cases where the proportion of phenotypic variance of each phenotype explained (PVE) (Zhou *et al.*, 2013) by all causal SNPs is 0.50. However, in some scenarios, we also performed additional simulations to examine cases where the PVE by all causal SNPs is 0.20. To allow for pleiotropic associations, these 500 causal SNPs were selected in a way that a certain proportion of them (in the range of 0–100%, with 20% increments) have non-zero effects on both traits. To allow for phenotypic correlation, the residual errors were generated from a bivariate normal distribution with a fixed covariance. Besides phenotypes, we also simulated two sets of continuous SNP annotations: an informative set where annotation values are dependent on SNP causality and another non-informative set where annotations do not depend on SNP causality. Overall, the simulation strategy employed here substantially deviates from our own modeling assumption; instead, it largely follows the modeling assumption employed in GPA (Chung *et al.*, 2014). With the genotypes and simulated phenotypes, we obtained marginal z scores and paired them with SNP annotation information to run analysis. We performed 100 simulation replicates for each setting. We considered four different methods for comparisons (Supplementary Text S1): (i) univariate analysis that fits a linear model on each trait separately; (ii) gwas-pw (Pickrell *et al.*, 2016); (iii) GPA (Chung *et al.*, 2014); and (iv) iMAP. We computed the power to detect causal SNPs at the FDR of 0.05 (or 0.10).

We consider three sets of simulations. We performed the first set of simulations to illustrate the benefits of modeling phenotypic correlation between traits. To do so, we varied the residual error covariance between the two simulated traits from −0.8 to 0.8 (= −0.8, −0.5, −0.2, 0, 0.2, 0.5, 0.8) to introduce negative or positive phenotypic correlations. We did not include any annotations in simulations here to exclude the influence of annotation on power comparison. In the simulations, all methods provide slightly conservative estimates of FDR at the given level of 0.05 or 0.10 (Supplementary Fig. S1), suggesting proper control of FDR by these methods. We also show power of different methods at a fixed FDR of 0.05 for positive phenotypic correlations in Figure 1A. The corresponding results for FDR of 0.10 are shown in Supplementary Figure S2, while the corresponding results with negative phenotypic correlations are shown in Supplementary Figures S3 and S4; both sets of results are similar to Figure 1A. Based on Figure 1A, we also contrast iMAP with GPA and display the power gain of iMAP in Figure 1B. In terms of power, both the proportion of pleiotropic effects and phenotypic correlation influence the relative power among methods. First, in the absence of pleiotropic effects, the power of iMAP and GPA is comparable with each other and with that of univariate analysis. However, both iMAP and GPA outperform the univariate analysis in the presence of pleiotropic effects. Second, when phenotypes are independent, the power of iMAP and GPA is comparable with each other. However, with increasing phenotypic correlation, the power of iMAP improves while the power of GPA (and gwas-pw) decays, highlighting the importance of explicit modeling of phenotypic correlation. We notice that gwas-pw is often underpowered when compared to GPA and iMAP, even if we performed simulations according to the gwas-pw study (Supplementary Text S2; Supplementary Fig. S5). We attribute the

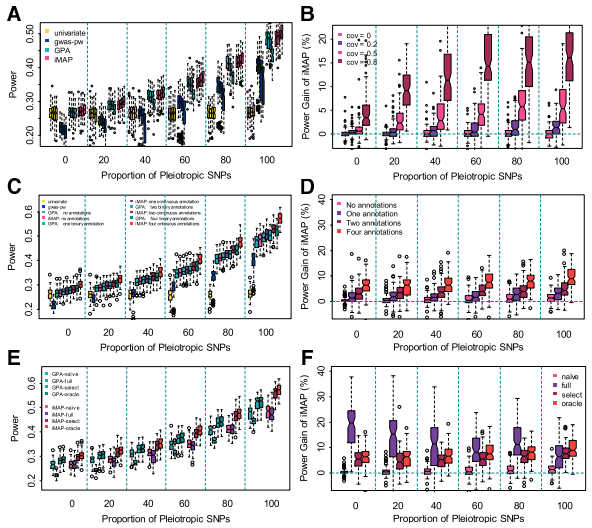

Fig. 1. Comparison of power in detecting associated SNPs by various methods in simulations. In all simulations, the proportion of pleiotropic causal SNPs varies from 0% to 100% (x-axis). Power is measured at a fixed FDR of 0.05. (**A**) Power of four methods (univariate analysis, gwas-pw, GPA and iMAP) in the setting where the two traits are positively correlated. For each method, the four boxplots at each pleiotropic proportion level correspond to four different phenotypic covariance values of 0, 0.2, 0.5 and 0.8, respectively. (**B**) Power gain of iMAP with respect to GPA computed based on panel (A). (**C**) Power of four methods (univariate analysis, gwas-pw, GPA and iMAP) in the presence of informative annotations. Variations of GPA and iMAP that incorporate a different number of annotations (0, 1, 2 and 4) are also presented. (**D**) Power gain of iMAP with respect to GPA computed based on panel (C). (**E**) Power of GPA and iMAP in the presence of four informative annotations and 100 noninformative annotations. Different variations of GPA and iMAP are considered: the naïve version does not incorporate any annotations; the full version includes all the annotations; the select version performs annotation selection; and the oracle version uses the four informative annotations. (**F**) Power gain of iMAP with respect to GPA computed based on panel (E)

relatively low power of gwas-pw to the fact that gwas-pw was not originally designed as an association mapping method.

We performed the second set of simulations to illustrate the benefits of using the original continuous SNP annotations versus dichotomizing them into binary ones. To do so, we simulated four sets of informative SNP annotations that all have continuous values. We used these continuous annotations directly for iMAP but dichotomized them into binary annotations for GPA, as GPA can only take binary annotations. For both GPA and iMAP, we considered settings where 0–4 annotations were incorporated into the model. To minimize the influence of phenotypic correlation on comparison, we set the residual error covariance to be 0 and used independent traits. The power of different methods at FDR = 0.05 is shown in Figure 1C. The corresponding results for FDR = 0.10 are shown in Supplementary Figure S6. We also contrast iMAP with GPA and display the power gain of iMAP in Figure 1D. Again, all methods provide reasonably and slightly conservative estimate of FDR (Supplementary Fig. S7). In terms of power, GPA and iMAP perform similarly as the univariate analysis in the absence of annotations and pleiotropic SNPs. Both GPA and iMAP outperform the univariate analysis in the presence of pleiotropic SNPs or when annotations are available. Importantly, because iMAP models the original continuous annotations directly, iMAP is slightly more powerful than GPA in the presence of annotations and its power gain increases with increasing number of informative annotations. For example, when the proportion of pleiotropic SNPs is 20%, the average power gain of iMAP versus GPA is 0.371, 1.82, 3.43 or 6.12%, when 1-4 annotations are incorporated, respectively (Fig. 1D). The power difference between iMAP and GPA is mainly due to the use of continuous annotations, as iMAP and GPA are comparable to each other when the same set of binary annotations were used (Supplementary Fig. S8). Besides the main setting where the PVE by all causal SNPs was set to be 0.5, we also examined small effect settings where the PVE by all causal SNPs was set to be only 0.2. The power of all the compared methods is low in this setting, although the rank among the compared methods remains the same (Supplementary Fig. S9).

Finally, we performed a third set of simulations to illustrate the benefits of annotation selection when a large number of annotations are present. To do so, we used the same four informative annotations above and simulated a large number (100) of non-informative annotations whose values were independent of SNP causality. Again, we set the residual error covariance to be 0 and used independent traits. iMAP analyzes all annotations jointly through a penalized regression framework to select informative ones among them. To better understand the effectiveness of annotation selection, besides the standard iMAP, we also considered three variations of iMAP in method comparison: (i) an iMAP-naïve method that does not incorporate any annotation; (ii) an iMAP-full method that naively incorporates all annotations without selection; and (iii) an iMAP-oracle method that uses only the four informative annotations. For additional comparison, we also considered different ways of applying GPA: (i) GPA-naïve that does not include any annotation; (ii) GPA-select that examines one annotation at a time, computes for each annotation a likelihood ratio statistic and uses a likelihood ratio statistic cutoff of 12 (which corresponds to an approximate P-value of 5.0×10^{-4}) to select a set of informative annotations that are further included into the GPA model for a final analysis; note that the original GPA method does not provide this option; (iii) GPA-full that includes all annotations to the model; and (iv) GPA-oracle where the correct four informative annotations are supplied. Note that the oracle version of GPA and iMAP represents an upper limit of power for the two methods. Here, we did not consider the univariate analysis and gwas-pw as the two methods cannot handle annotations and were generally less powerful compared to GPA and iMAP. We display the power comparison results among different variations of iMAP and GPA in Figure 1E and F and Supplementary Figure S10. A few patterns are obvious. First, in the presence of a large number of annotations, approaches that naively model all annotations together (i.e. iMAP-full and GPA-full) are often much less powerful compared to approaches that do not consider annotations at all (i.e. iMAP-naïve and GPA-naïve), suggesting that the small degree of freedom in the full model impairs model performance. Second, approaches that select annotations (i.e. iMAP and GPA-select) almost always result in a great power gain compared to naïve approaches (i.e. iMAP-naïve and GPA-naïve), and can often achieve similar power as the corresponding oracle models (i.e. iMAP-oracle and GPA-oracle). The competent performance of selection approaches highlights the importance of performing annotation selection. Importantly, even with annotation selection, iMAP and GPA still provide effective estimate of FDR (Supplementary Fig. S11). Finally, iMAP relies on the formal Lasso penalty to perform annotation selection and is more powerful than GPA-select which relies on a simpler procedure to select annotations (Fig. 1F). The comparative results between iMAP and GPA-select are consistent with early literature on Lasso being more effective compared with the subset selection approach (Tibshirani, 1996; Zeng, *et al.*, 2014; Zou, 2006). Importantly, iMAP is capable of selecting the

Table 1. Accuracy of iMAP in selecting informative annotations and in estimating the annotation coefficients in the setting where four independent annotations with relatively large effects are present

Proportion of pleiotropic causal SNPs (%)	True	False	MSE		
			Oracle	Select	Full
0	3.53	0.38	0.05	0.28	383.52
20	7.81	0.53	0.41	2.38	300.15
40	9.72	1.20	0.29	2.03	68.63
60	5.77	0.81	0.31	2.58	12.53
80	4.98	0.75	0.56	2.70	17.46
100	3.15	0.37	0.12	2.17	18.38

Note: Simulations were carried out in the presence of four informative annotations and 100 non-informative annotations for various proportion of pleiotropic causal SNPs (rows). The 'True' column lists the number of selected correct non-zero annotation parameters inside the mlogit model. Note that a total of 12 ($=3\times4$) non-zero annotation parameters are expected in the presence of four informative annotations. The 'False' column lists the number of selected incorrect non-zero annotation parameters. MSE denotes the median squared error for the estimated annotation parameters across 100 simulation replicates for three different versions of iMAP: the oracle version uses the four informative annotations; the select version performs annotation selection; and the full version includes all annotations without selection.

informative annotations with high precision and can accurately estimate the annotation coefficients (Table 1 and Supplementary Fig. S12). While our main simulation setting uses four independent annotations with relatively large effects, we also examined a setting where we include ten independent annotations with relatively small effects along with another 100 null annotations. As expected, because of small annotation effects, the accuracy to select the true annotations in this setting reduces substantially. However, the relative performance of various methods remains the same, with iMAP greatly outperforming iMAP-full (Supplementary Table S2). We also assessed the performance of iMAP in a four-annotations setting where the first two annotations are correlated with each other and the second two annotations are also correlated with each other, with correlation coefficient varying from 0.2 to 0.8. The 100 null annotations are independent each other and are not correlated to the four true annotations. When correlation is low or moderate (e.g. 0.2 or 0.5), the accuracy of various methods reduces slightly with ranking among them remaining the same. In contrast, when correlation is high (e.g. 0.8), while iMAP outperforms iMAP-full in most cases, it can occasionally perform worse than iMAP-full (Supplementary Table S3). The results with correlated annotations are consistent with those observed previously and can be improved when by replacing the selection method Lasso with, for example, Elastic Net (Zou and Hastie, 2005).

3.1.2 Estimating causal SNP proportions and annotation coefficients

A key feature of mixture models is that they can provide estimates for the proportion of SNPs that have various effects on the two phenotypes. These proportion estimates can help shed light on the genetic architecture of complex traits. In terms of iMAP, we are often interested in estimating: π_{11}, the proportion of pleiotropic SNPs that are associated with both traits; π_{10} (or π_{01}), the proportion of SNPs that are associated with only one trait; and $\pi_{11}/(\pi_{10}+\pi_{11})$ [or $\pi_{11}/(\pi_{01}+\pi_{11})$], the proportion of SNPs associated with one trait that are also associated with the other. The last quantity has been used to evaluate causality of one trait on the other

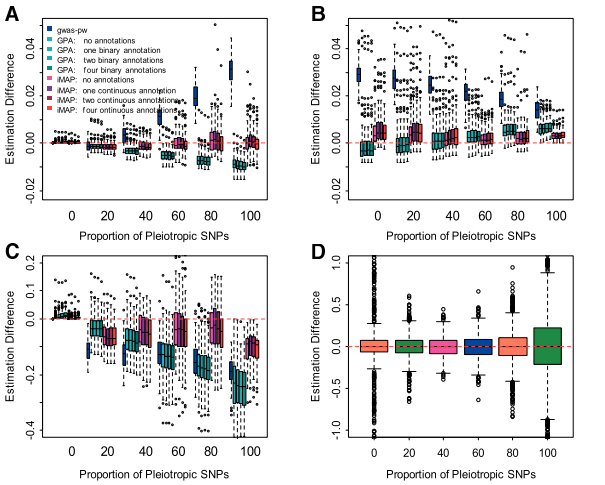

Fig. 2. Estimation accuracy for the proportions of different SNP association categories by different methods in simulations. Methods for comparison include gwas-pw, GPA and iMAP. Because four informative annotations are present, we considered variations of GPA and iMAP that incorporate a different number of annotations (0, 1, 2 and 4). The difference between the estimated values and truth (y-axis) are computed for various settings where the proportion of pleiotropic causal SNPs varies from 0% to 100% (x-axis). Different quantities of interest are considered: (**A**) π_{11}, the proportion of SNPs associated with both traits; (**B**) π_{10}, the proportion of SNPs associated with only the first trait; (**C**) $\pi_{11}/(\pi_{11}+\pi_{10})$, the proportion of SNPs associated with the first trait that are also associated with the second trait; (**D**) estimated parameters for the informative annotations

(Pickrell *et al.*, 2016). Specifically, a large $\pi_{11}/(\pi_{10}+\pi_{11})$ and a small $\pi_{11}/(\pi_{01}+\pi_{11})$ suggest that a large fraction of SNPs associated with the first trait is also associated with the second trait, but not vice versa, indicating that the first trait may causally affect the second trait. A small $\pi_{11}/(\pi_{10}+\pi_{11})$ and a large $\pi_{11}/(\pi_{01}+\pi_{11})$ indicate that the second trait may causally affect the first trait. On the other hand, a large $\pi_{11}/(\pi_{10}+\pi_{11})$ and a large $\pi_{11}/(\pi_{01}+\pi_{11})$ indicate that both traits may share common biological pathways. While we caution against the causal interpretation of $\pi_{11}/(\pi_{10}+\pi_{11})$ [or $\pi_{11}/(\pi_{01}+\pi_{11})$] in association studies, we do consider $\pi_{11}/(\pi_{10}+\pi_{11})$ [or $\pi_{11}/(\pi_{01}+\pi_{11})$] as a useful quantity to be estimated along with the original parameters π_{11}, π_{10} and π_{01}. To compare method performance in estimating the above quantities, we focused on the second simulation setting described in the previous section and applied three different methods (gwas-pw, GPA and iMAP) to examine their performance in the presence of annotations. For better visualization, we contrast the estimated values with the true values and show their differences in Figure 2A–C. Overall, iMAP generates estimates that are slightly closer to the truth than either GPA or gwas-pw across a range of pleiotropic SNP proportions and various numbers of annotations. The slightly accuracy gain in iMAP is likely due to its higher power in identifying SNP associations.

3.2 Results of real data applications

3.2.1 Real data and annotations

We applied our method to analyze 48 traits from 31 GWASs (Supplementary Table S1). These traits span a wide range of phenotypes that include anthropometric traits [e.g. height and body mass index (BMI)], hematopoietic traits [e.g. mean cell haemoglobin concentration (MCHC) and red blood cell count (RBC)], immune diseases [e.g. Crohn's disease (CD) and inflammatory bowel disease (IBD)], and neurological diseases [e.g. Alzheimer's disease and schizophrenia).

We obtained summary statistics for these traits. We also obtained a total of 40 annotations based on genome-wide occupancy information of four histone marks (*H3K27me3*, *H3K36me3*, *H3K4me1* and *H3K4me3*) in 10 tissue groups (blood/immune, adipose, adrenal/pancreas, bone/connective, cardiovascular, CNS, gastrointestinal, liver, muscle and other) from the Roadmap Epigenomics Project (Roadmap Epigenomics Consortium *et al.*, 2015). Data processing details are provided in Supplementary Text S2. We analyzed each of the 1128 trait pairs with different methods. Methods considered include univariate analysis, gwas-pw, GPA and iMAP. For iMAP, we considered two different approaches. The first approach (iMAP-naïve) did not integrate any histone annotation while the second approach (iMAP) analyzed all histone annotations jointly with penalized selection. Because GPA can only examine a small number of histone annotations, to allow for a fair comparison, we applied GPA-select using the same screening procedure described in the simulations section and included all the selected histone annotations into the GPA model for a final analysis. For all methods, we declared association significance based on an estimated FDR of 0.1%.

3.2.2 Integrative analysis by iMAP improves association mapping power

We computed the number of significant associations detected by different methods for each trait pair (Fig. 3). These results are based on an estimated FDR of 0.05, which can be less stringent than the usual 5×10^{-8} *P*-value threshold that aims to control for an FWER of approximately 0.05. Consistent with simulations, iMAP detected more associations than any other methods. Specifically, iMAP identified an average of 658 associated SNPs across trait pairs (Fig. 3A), while iMAP-naïve, GPA-select, gwas-pw and univariate analysis identified 580, 463, 232 and 358 associated SNPs, respectively. In addition, iMAP is ranked as the best method in terms of identifying the largest number of associated SNPs in 703 (62.3%) trait pairs out of the 1128 total (Fig. 3B–E), while iMAP-naïve, GPA-select, gwas-pw and univariate analysis were ranked as the best in 7 (0.6%), 217 (19.2%), 200 (17.7%) and 1 (0.09%) trait pairs, respectively. Among the associated SNPs, an average of 3.8% (25) of them has pleiotropic associations with both traits. The proportion of pleiotropic SNPs is similar whether annotations were accounted for or not; indeed, iMAP-naïve also identified an average of 2.9% (17) SNPs to have pleiotropic associations with both traits. The relatively small proportion of pleiotropic associations detected by iMAP is consistent with previous studies (Sivakumaran *et al.*, 2011). However, among biological related traits, the proportion of pleiotropic associations can be large (Supplementary Fig. S13). For example, among pairs of lipid traits [high-density lipoproteins (HDL), low density lipoproteins (LDL), total cholesterol (TC) and triglycerides (TG)], iMAP identified an average of 24.5% (522) SNPs associated with two traits. Similarly, iMAP-naïve identified an average of 19.9% (364) SNPs associated with two traits.

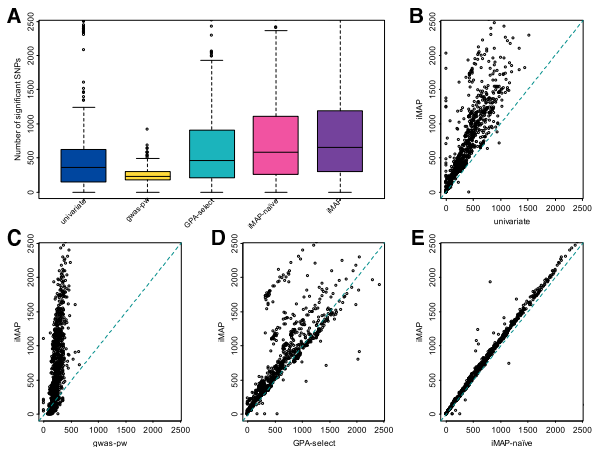

Fig. 3. Power comparison among methods for the 1128 trait pairs analyzed in the real data application. Methods for comparison include: univariate, gwas-pw, GPA-select, iMAP-naïve and iMAP. (**A**) Boxplots show the number of associated SNPs that pass the genome-wide significance threshold identified by various methods across 1128 trait pairs. The number of associated SNPs identified by iMAP for each trait pair is also plotted against that identified by (**B**) univariate, (**C**) gwas-pw, (**D**) GPA-select and (**E**) iMAP-naïve

As an illustrative example, we display a local genetic region that harbors a pleiotropic association for the trait pair of HDL and TG. HDL and TG are blood metabolic phenotypes that are biologically related and have been previously shown to share a common genetic background (Bulik-Sullivan *et al.*, 2015; Pickrell *et al.*, 2016; Teslovich *et al.*, 2010). For the joint analysis of HDL-TG, iMAP identified a total of 2305 associated SNPs, among which 376 (15.9%) of them are pleiotropic associations. The identified associated SNPs span a total of 51 loci (Supplementary Table S4), among which 44 were previously known to be associated with either HDL or TG (MacArthur, *et al.*, 2017). For illustration purpose, we display in Supplementary Figure S14 a genomic region centered at a pleiotropic SNP, *rs1809167*, which resides near the gene *TRIB1* on 8q24.13. *TRIB1* is a protein-coding gene that is associated with multiple lipid traits and cardiovascular disease (MacArthur, *et al.*, 2017; Soubeyrand, *et al.*, 2016; Teslovich, *et al.*, 2010). In the analysis, when we did not include SNP annotations, the PIP of *rs1809167* computed by iMAP-naïve is only 0.753 (Supplementary Fig. S14A), below the genome-wide significance threshold (PIP = 0.936, which corresponds to an FDR of 0.1%). After including SNP annotations, however, the PIP of the same SNP computed by iMAP increases to 0.988 (Supplementary Fig. S14B), which passes the genome-wide significance threshold. Therefore, by integrating SNP functional annotations, iMAP has the potential to identify associations that may otherwise be missed by iMAP-naïve. In this analysis, iMAP also selected four histone annotations from three tissues to be relevant to the two traits, and these four annotations include blood/immune *H3K36me3* and *H3K4me1*, bone/connective *H3K36me3* and liver *H3K4me1*. The estimated annotation coefficients of the four annotations are shown in Supplementary Figure S14B. Importantly, we find that the estimated effects of liver *H3K4me1* and bone/connective H3K36me3 are larger than the other two annotations, suggesting that liver and bone/connective may play an important role in the biology of HDL and TG, consistent with previous findings (Kozlitina *et al.*, 2014; Lories *et al.*, 2013; Rivadeneira and Mäkitie, 2016; Roman *et al.*, 2015).

3.2.3 iMAP identifies important histone annotations relevant to complex traits

Next, we examine the important histone annotations selected by iMAP across all trait pairs. iMAP selected at least one histone annotation for most pairs of traits examined (99.6%). On average, it selected 2.8 histone annotations (median = 2.0) that influence the probability of a SNP to be associated with one trait and selected 0.20 histone annotations (median = 0) that influence the probability of a SNP to be associated with both traits. The small number of histone annotations selected in the later case is likely due to the relatively small number of pleiotropic SNPs and the subsequent low statistical power there. To examine individual histone annotations,

for each histone annotation in turn, we counted the proportion of times that the histone annotation of interest was selected by iMAP across all analyzed trait pairs. In addition, we obtained the coefficient estimates for these selected histone annotations. Intuitively, if a histone annotation is an important predictor of SNP association, then it would be selected more often than the other non-important ones. In addition, it would have a higher estimated annotation coefficient than the others. We show both the coefficient estimates and the probability of being selected for the four histone annotations in Figure 4A. Specifically, across all tissues, *H3K36me3* and *H3K4me1* are selected more often than *H3K27me3* and *H3K4me3* (47.3% and 33.1% of the times versus 9.97% and 9.59% of the times). In addition, the estimated annotation coefficients of *H3K36me3* and *H3K4me1* are also larger than those of *H3K27me3* and *H3K4me3* (an average of 0.158 and 0.168 versus an average of 0.036 and 0.077). Both results are consistent with the important role of *H3K36me3* and *H3K4me1* in marking promoter or enhancer regions that are known to be predictive of SNP causality from previous univariate analyses (Roadmap Epigenomics Consortium, *et al.*, 2015; The ENCODE Project Consortium, 2012).

To further investigate trait-relevance of these histone annotations, for each trait/histone annotation pair in turn, we counted the proportion of times that the particular histone annotation was selected to have non-zero effects in all trait pairs that involve the trait of interest. The resulting value, which we simply refer to as the trait *relevance score*, is in the range of 0 and 1 and represents the relevance of the specific histone annotation to the given trait. Because histone annotations are tissue-specific, the resulting relevance scores also quantify trait-tissue relevance for trait tissue pairs. Results show that SNP associations in some traits are primarily predicted by histone annotations in a single tissue, while associations in other traits are predicted by histone annotations in a range of tissues (Fig. 4B). For example, SNP associations in many immune diseases (e.g. UC, CD, IBD, T1D, Lupus, PBC and RA) are primarily related to two histone annotations, *H3K36me3* and *H3K4me1*, in the blood/immune tissue. Associations for most anthropometric traits (e.g. Height, BMI2, FNBMD, LSBMD, BMI1, BW2, G10, GPC and Obesity) can be predicted by marks in both the bone/connective and CNS tissues. Psychiatric disorders (e.g. SCZ, BPSCZ and BIP) are related to both blood/immune and CNS tissues. Lipids-related traits and diseases (e.g. HDL, LDL, TC, TG, CAD and X2hrGlucose) are related to blood/immune, bone/connective, gastrointestinal and liver tissues. While hematopoietic traits (e.g. MCHC, MCH, HB, MCV, MPV, PCV, PLT and RBC) are related to blood/immune, gastrointestinal and bone/connective tissues. Overall, the results are largely consistent with previous univariate analyses and reveal the complexity of trait-tissue relevance (Bradfield *et al.*, 2011; Liu *et al.*, 2015; Teslovich *et al.*, 2010).

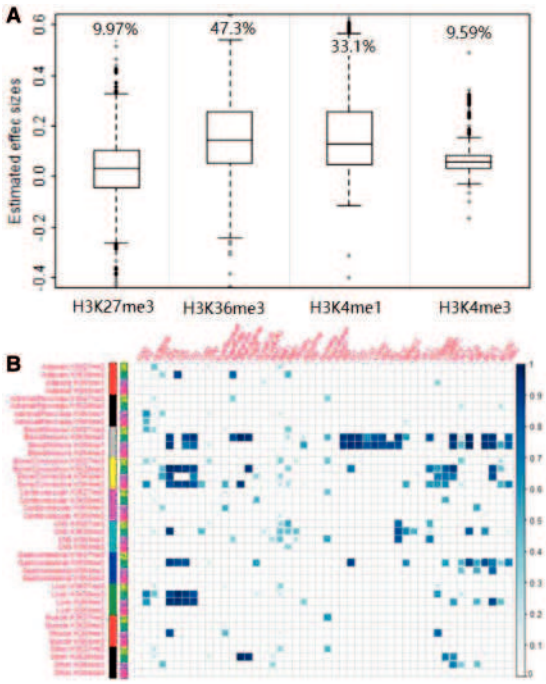

Fig. 4. Annotation selection in the real data application. (**A**) Estimated annotation effect sizes for the selected annotations across all analyzed trait pairs. On top of the boxplots lists the percentage of times a histone annotation is selected across these trait pairs. (**B**) Relevance score for each annotation (rows) across all traits (columns). The relevance score quantifies the importance of an annotation for a particular trait of interest and is computed based on analysis of all trait pairs

3.2.4 iMAP characterizes genetic relationship between pair of traits

Finally, we examine the estimated $\pi_{11}/(\pi_{10}+\pi_{11})$ [or $\pi_{11}/(\pi_{01}+\pi_{11})$] values across all trait pairs. The proportion $\pi_{11}/(\pi_{10}+\pi_{11})$ [or $\pi_{11}/(\pi_{01}+\pi_{11})$] can be asymmetric between any two traits and have been used to infer causality between the two traits (Pickrell *et al.*, 2016). The proportions estimated by iMAP are shown in Figure 5 and are largely consistent with the estimates by iMAP-naïve (Supplementary Figs S15A–C also show the proportions estimated by gwas-pw and GPA-select). In either case, the estimated proportions are largely symmetric, though with substantial asymmetric patterns. In particular, among all trait-pairs, 331 (29.3%) of them have an asymmetric pattern with the difference between the two proportions estimated to be larger than 10%. For example, the estimated proportions of SNPs associated with FG and HDL that are also associated with T2D are 0.983 and 0.507, respectively. On the other hand, the estimated proportions of SNPs associated with T2D that are also associated with FG and HDL are 0.568 and 0.058, respectively. The large proportion estimates between T2D and FG suggest that T2D and FG may share common biological pathways (Solovieff *et al.*, 2013).

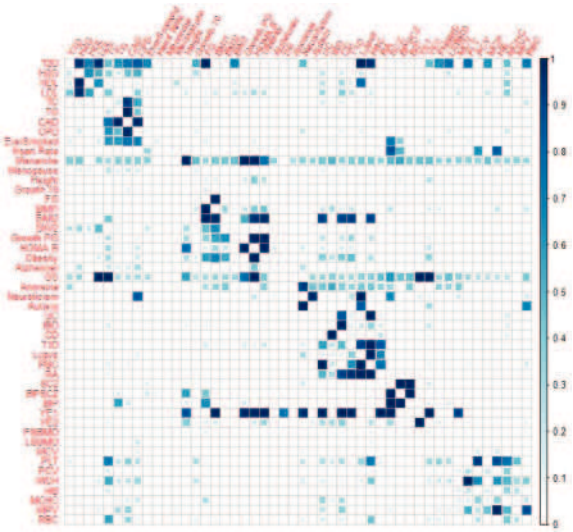

Fig. 5. Estimated probability that a SNP associated with one trait (y-axis) is also associated with the other trait (x-axis), for 48 trait pairs in the real data application. Results are based on iMAP that performs annotation selection among 40 annotations

In contrast, the fact that SNPs associated with HDL often are also associated with T1D, but not vice versa, suggests that HDL may mediate SNPs effects onto T1D (Solovieff *et al.*, 2013). As another example, the estimated proportions of SNPs associated with the four blood lipid traits (HDL, LDL, TC and TG) that are also associated with CAD are high (0.669, 0.681, 0.815 and 0.724, respectively). In contrast, the proportions of SNPs associated with CAD that are also associated with the four blood lipid traits are small (0.068, 0.066, 0.064 and 0.096, respectively). The asymmetric relationship between lipid traits and CAD again suggests that lipid traits may mediate SNP effects onto CAD, consistent with previous findings (Bjornsson *et al.*, 2017; Pickrell *et al.*, 2016; Willer, *et al.*, 2008).

4 Discussion

We have presented a penalized Gaussian mixture model method, iMAP, which incorporates functional annotations for association mapping of pair-wise traits. iMAP properly accounts for phenotypic correlation, models association pattern across genome-wide SNPs, can accommodate both continuous and binary annotations, while capable of selecting informative annotations among a large set. iMAP is particularly useful for integrative analyses of GWAS summary statistics with a potentially large number of SNP annotations. While we have mainly focused on the use of Lasso (Tibshirani, 1996) for annotation selection, we have implemented the Elastic Net (Zou and Hastie, 2005) in the software to improve the performance of iMAP in the presence of correlated annotations. In addition, we have implemented the standard iMAP model without annotation input, so that the users can access the contribution of likelihood versus prior from annotation in the final association results. iMAP is also reasonably fast in the real data: it takes an average of 48 min (median = 45 min) to analyze a pair of traits in the presence of one annotation and takes an average of 7.2 h (median = 4.8 h) when 40 annotations are used. With extensive simulations and real data application to 48 GWAS traits, we have illustrated the benefits of iMAP in terms of improving association mapping power, identifying important annotations, revealing genetic architecture underlying complex traits, as well as investigating relationship among phenotypes.

Throughout the text, we have used FDR thresholds to identify significant associations. While using FDR allows us to fairly compare the power of different methods based on the same threshold, using FDR also requires care in results interpretation for two reasons. First, compared to the commonly used *P*-value threshold of 5×10^{-8} that controls for an FWER of $\sim$0.05, depending on the number of identified SNPs and the number of total SNPs, an FDR of 0.05 can be less stringent and can lead to the identification of unexpected many associated variants (Brzyski *et al.*, 2017). Consequently, the identified associations based on an FDR of 0.05 may suffer from lower replication rate as compared to those identified based on a *P*-value threshold of 5×10^{-8}. Second, compared to FWER control, FDR control does not have the sub-setting property (Goeman and Solari, 2014). Specifically, because FWER controls for the probability of making a type I error, a FWER control at any level on the entire set of the identified SNPs guarantees the same (or more stringent) FWER level on any subset of the identified SNPs, regardless of SNP association evidence measured by p-value within the subset (Goeman and Solari, 2014). In contrast, FDR controls for the expected proportion of false discoveries, and an FDR control at any level on the entire set of the identified SNPs does not guarantee the same FDR level on any subset of the identified SNPs. In particular, SNPs with low PIP will have high local FDR, and consequently, the subset of identified SNPs with low PIP will have higher FDR than what is controlled for on the entire set. Therefore, for FDR control, we recommend examining PIP as additional association evidence within the identified SNP set. We highlight these two important differences between FWER and FDR to help practitioners better interpret the results from iMAP and other Bayesian methods that rely on FDR for error control.

In this study, we have mainly focused on the relatively simple task of analyzing pairs of traits. Extensions of iMAP to analyzing more than two traits may seem trivial conceptually, but present important statistical and computational challenges that need to be properly addressed. Specifically, with increasing number of traits (d), the number of possible association patterns increases exponentially (2^d): a SNP can be associated with none of the traits, with any one trait, with any two traits, ..., or with all traits. Unfortunately, direct and naïve modeling of the large number of possible association patterns would require an exponentially large number of π parameters and an exponentially large number of annotation coefficients inside the mlogit model. Employing a large number of parameters not only would reduce the degrees of freedom and subsequently power but also imposes computational hurdles (e.g. the computational complexity of iMAP also scales exponentially to the number of traits). Therefore, additional modeling assumptions on the possible association patterns are necessary to enable efficient and powerful analysis. Previous methods developed in eQTL mapping setting for identifying eQTLs in multiple tissues (Flutre *et al.*, 2013) have suggested that using either sparse (that a SNP is only associated with a small set of phenotypes) or group-structured (that there exist a number of phenotype groups, where a SNP is either associated with all phenotypes within the same group or associated with none of them) effect size assumptions are effective in modeling association patterns across tissues. Adapting these eQTL mapping methods to association mapping of pleiotropic effects is an interesting future avenue to explore.

Like previous other mixture methods (Chung *et al.*, 2014; Pickrell *et al.*, 2016), iMAP makes a key assumption that the joint likelihood for all SNPs is a simple product of the likelihoods from every SNP. However, because SNPs are in linkage disequilibrium (LD) and their genotypes are correlated with each other (Wall and Pritchard, 2003), assuming a simplified form of the joint likelihood is unrealistic. To examine whether iMAP remains effective under LD, we have performed a series of simulations using correlated SNPs instead of independent ones (Supplementary Text S2). Briefly, we show that the power comparison results among methods are relatively stable with respect to LD and that iMAP still outperforms the other methods in identifying associated SNPs in the presence of LD. However, LD strongly influences the estimation of the proportions of causal SNPs in different association categories, and none of the methods examined here are accurate in estimating the causal proportions in the presence of LD (Text S2). We have attempted to improve the estimation of causal SNPs proportions by incorporating LD information as weights for the SNP likelihoods as suggested in (Liley *et al.*, 2017). However, we found that using either LD scores (Finucane, *et al.*, 2015) or linkage disequilibrium adjusted kinships (LDAK) weights (Speed *et al.*, 2012) does not improve the accuracy in causal SNP proportion estimation. Besides the weighted likelihood, we have also attempted to prune SNPs before performing analysis for estimating causal SNP proportions (Nishino *et al.*, 2018). However, we found that the results from pruning can be unstable depending on how causal SNPs were simulated. Therefore, we view it as an important direction in the future to incorporate LD

pattern into iMAP or other mixture model methods in a principled way. Such extension, together with efficient algorithmic innovations, will facilitate the accurate estimation of causal SNP proportion in various association categories and will likely further improve the power in fine mapping of pleiotropic traits (Kichaev and Pasaniuc, 2015; Kichaev *et al.*, 2014; Spain and Barrett, 2015).

Acknowledgements

We are grateful to the associate editor and the reviewers for their valuable suggestions and comments which have improved our manuscript. We thank all the GWAS consortium studies for making the summary data publicly available. The GWAS summary data source to these data consortium are given in the Supplementary Material.

Funding

This work was supported by the National Institutes of Health [NIH R01HG009124, R01GM126553] and the National Science Foundation [NSF DMS1712933].

Conflict of Interest: none declared.

References

Adzhubei,I. *et al.* (2013) Predicting functional effect of human missense mutations using PolyPhen-2. In: Jonathan,L. (eds) *Current Protocols in Human Genetics*. John Wiley & Sons, Inc., New York.

Andreassen,O.A. *et al.* (2014) Identifying common genetic variants in blood pressure due to polygenic pleiotropy with associated phenotypes. *Hypertension*, **63**, 819–826.

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.

Bjornsson,E. *et al.* (2017) A rare splice donor mutation in the haptoglobin gene associates with blood lipid levels and coronary artery disease. *Hum. Mol. Genet.*, **26**, 2364–2376.

Bradfield,J.P. *et al.* (2011) A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci. *PLoS Genet.*, **7**, e1002293.

Brzyski,D. *et al.* (2017) Controlling the rate of GWAS false discoveries. *Genetics*, **205**, 61–75.

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,L. *et al.* (2016a) DIVAN: accurate identification of non-coding disease-specific risk variants using multi-omics profiles. *Genome Biol.*, **17**, 252.

Chen,W. *et al.* (2016a) 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.

Cooper,G.M. *et al.* (2005) Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.*, **15**, 901–913.

Cotsapas,C. *et al.* (2011) Pervasive sharing of genetic effects in autoimmune disease. *PLoS Genet.*, **7**, e1002254.

Cross-Disorder Group of the Psychiatric Genomics Consortium. (2013a) Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.*, **45**, 984–994.

Cross-Disorder Group of the Psychiatric Genomics Consortium. (2013b) Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet*, **381**, 1371–1379.

Dixon,J.R. *et al.* (2015) Chromatin architecture reorganization during stem cell differentiation. *Nature*, **518**, 331–336.

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

Ernst,J. *et al.* (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, **473**, 43–49.

Ernst,J. and Kellis,M. (2010) Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat. Biotechnol.*, **28**, 817–825.

Ernst,J. and Kellis,M. (2012) ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods*, **9**, 215–216.

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

Fletcher,O. and Houlston,R.S. (2010) Architecture of inherited susceptibility to common cancer. *Nat. Rev. Cancer*, **10**, 353–361.

Flutre,T. *et al.* (2013) A statistical framework for joint eQTL analysis in multiple tissues. *PLoS Genet.*, **9**, e1003486.

Fu,Y. *et al.* (2014) FunSeq2: a framework for prioritizing noncoding regulatory variants in cancer. *Genome Biol.*, **15**, 480.

Giambartolomei,C. *et al.* (2014) Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.*, **10**, e1004383.

Goeman,J.J. and Solari,A. (2014) Multiple hypothesis testing in genomics. *Stat. Med.*, **33**, 1946–1978.

Gusev,A. *et al.* (2014) Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *Am. J. Hum. Genet.*, **95**, 535–552.

Hasan,A. *et al.* (2016) Fast estimation of multinomial logit models: r package mnlogit. *J. Stat. Softw.*, **75**, 1–24.

He,Q. *et al.* (2013) A general framework for association tests with multivariate traits in large-scale genomics studies. *Genet. Epidemiol.*, **37**, 759–767.

Hu,Y. *et al.* (2017) Joint modeling of genetically correlated diseases and functional annotations increases accuracy of polygenic risk prediction. *PLoS Genet.*, **13**, e1006836.

Ionita-Laza,I. *et al.* (2016) A spectral approach integrating functional genomic annotations for coding and noncoding variants. *Nat. Genet.*, **48**, 214–220.

Ji,S.G. *et al.* (2017) Genome-wide association study of primary sclerosing cholangitis identifies new risk loci and quantifies the genetic relationship with inflammatory bowel disease. *Nat. Genet.*, **49**, 269–273.

Jostins,L. *et al.* (2012) Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*, **491**, 119–124.

Kellis,M. *et al.* (2014) Defining functional DNA elements in the human genome. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 6131–6138.

Kichaev,G. *et al.* (2014) Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Genet.*, **10**, e1004722.

Kichaev,G. and Pasaniuc,B. (2015) Leveraging functional-annotation data in trans-ethnic fine-mapping studies. *Am. J. Hum. Genet.*, **97**, 260–271.

Kim,Y.J. *et al.* (2011) Large-scale genome-wide association studies in East Asians identify new genetic loci influencing metabolic traits. *Nat. Genet.*, **43**, 990–995.

Kircher,M. *et al.* (2014) A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.*, **46**, 310–315.

Kozlitina,J. *et al.* (2014) Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. *Nat. Genet.*, **46**, 352–356.

Kumar,P. *et al.* (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.*, **4**, 1073–1081.

Lane,J.M. *et al.* (2017) Genome-wide association analyses of sleep disturbance traits identify new loci and highlight shared genetics with neuropsychiatric and metabolic traits. *Nat. Genet.*, **49**, 274–281.

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

Lee,P.H. *et al.* (2016) Partitioning heritability analysis reveals a shared genetic basis of brain anatomy and schizophrenia. *Mol. Psychiatr.*, **21**, 1680–1689.

Li,Y. and Kellis,M. (2016) Joint Bayesian inference of risk variants and tissue-specific epigenomic enrichments across multiple complex human diseases. *Nucleic Acids Res.*, **44**, e144.

Liley,J. *et al.* (2017) A method for identifying genetic heterogeneity within phenotypically defined disease subgroups. *Nat. Genet.*, **49**, 310–316.

Liu,J. *et al.* (2016) EPS: an empirical Bayes approach to integrating pleiotropy and tissue-specific information for prioritizing risk genes. *Bioinformatics*, **32**, 1856–1864.

Liu,J.Z. *et al.* (2015) Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.*, **47**, 979–986.

Lonsdale,J. *et al.* (2013) The genotype-tissue expression (GTEx) project. *Nat. Genet.*, **45**, 580–585.

Lories,R.J. *et al.* (2013) To Wnt or not to Wnt: the bone and joint health dilemma. *Nat. Rev. Rheumatol.*, **9**, 328–339.

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.

MacArthur,J. *et al.* (2017) The new NHGRI-EBI catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.*, **45**, D896–D901.

Maier,R. *et al.* (2015) Joint analysis of psychiatric disorders increases accuracy of risk prediction for schizophrenia, bipolar disorder, and major depressive disorder. *Am. J. Hum. Genet.*, **96**, 283–294.

McLaughlin,R.L. *et al.* (2017) Genetic correlation between amyotrophic lateral sclerosis and schizophrenia. *Nat. Commun.*, **8**, 14774.

McVicker,G. *et al.* (2013) Identification of genetic variants that affect histone modifications in human cells. *Science*, **342**, 747–749.

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.

Nishino,J. *et al.* (2018) Empirical Bayes estimation of semi-parametric hierarchical mixture models for unbiased characterization of polygenic disease architectures, *Front. Genet.*, **9**, 115.

Pickrell,J.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.

Pickrell,J.K. *et al.* (2010) Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature*, **464**, 768–772.

Pique-Regi,R. *et al.* (2011) Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res.*, **21**, 447–455.

Rivadeneira,F. and Mäkitie,O. (2016) Osteoporosis and bone mass disorders: from gene pathways to treatments. *Trends Endocrinol. Metabol.*, **27**, 262–281.

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

Roman,T.S. *et al.* (2015) Multiple hepatic regulatory variants at the GALNT2 GWAS locus associated with high-density lipoprotein cholesterol. *Am. J. Hum. Genet.*, **97**, 801–815.

Schork,A.J. *et al.* (2013) All SNPs are not created equal: genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs. *PLoS Genet.*, **9**, e1003449.

Sivakumaran,S. *et al.* (2011) Abundant pleiotropy in human complex diseases and traits. *Am. J. Hum. Genet.*, **89**, 607–618.

Smith,E.N. *et al.* (2010) Longitudinal genome-wide association of cardiovascular disease risk factors in the Bogalusa heart study. *PLoS Genet.*, **6**, e1001094.

Solovieff,N. *et al.* (2013) Pleiotropy in complex traits: challenges and strategies. *Nat. Rev. Genet.*, **14**, 483–495.

Soubeyrand,S. *et al.* (2016) TRIB1 is regulated post-transcriptionally by proteasomal and non-proteasomal pathways. *PLoS ONE*, **11**, e0152346.

Spain,S.L. and Barrett,J.C. (2015) Strategies for fine-mapping complex traits. *Hum. Mol. Genet.*, **24**, R111–R119.

Speed,D. *et al.* (2012) Improved heritability estimation from genome-wide SNPs. *Am. J. Hum. Genet.*, **91**, 1011–1021.

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.

Teslovich,T.M. *et al.* (2010) Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*, **466**, 707–713.

The ENCODE Project Consortium. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.

The Wellcome Trust Case Control Consortium. (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, **447**, 661–678.

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

Tung,J. *et al.* (2015) The genetic architecture of gene expression levels in wild baboons. *Elife*, **4**,

van der Sluis,S. *et al.* (2013) TATES: efficient multivariate genotype-phenotype analysis for genome-wide association studies. *PLoS Genet*, **9**, e1003235.

Van der Sluis,S. *et al.* (2015) MGAS: a powerful tool for multivariate gene-based genome-wide association analysis. *Bioinformatics*, **31**, 1007–1015.

Varin,C. *et al.* (2011) An overview of composite likelihood methods. *Stat. Sin.*, **21**, 5–42.

Visscher,P.M. *et al.* (2017) 10 Years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.*, **101**, 5–22.

Wall,J.D. and Pritchard,J.K. (2003) Haplotype blocks and linkage disequilibrium in the human genome. *Nat. Rev. Genet.*, **4**, 587–597.

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

Warren,H.R. *et al.* (2017) Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nat. Genet.*, **49**, 403–415.

Weissbrod,O. *et al.* (2016) Multikernel: linear mixed models for complex phenotype prediction. *Genome Res.*, **26**, 969–979.

Wen,X. *et al.* (2015) Cross-population joint analysis of eQTLs: fine mapping and functional annotation. *PLoS Genet.*, **11**, e1005176.

Wen,X. *et al.* (2016) Efficient integrative multi-SNP association analysis via deterministic approximation of posteriors. *Am. J. Hum. Genet.*, **98**, 1114–1129.

Willer,C.J. *et al.* (2008) Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat. Genet.*, **40**, 161–169.

Zeng,P. *et al.* (2014) Variable selection approach for zero-inflated count data via adaptive lasso. *J. Appl. Stat.*, **41**, 879–894.

Zhernakova,A. *et al.* (2009) Detecting shared pathogenesis from the shared genetics of immune-related diseases. *Nat. Rev. Genet.*, **10**, 43–55.

Zhou,X. *et al.* (2013) Polygenic modeling with Bayesian sparse linear mixed models. *PLoS Genet.*, **9**, e1003264.

Zhou,X. and Stephens,M. (2014) Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat. Methods*, **11**, 407–409.

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.

Zou,H. (2006) The adaptive Lasso and its oracle properties. *J. Am. Stat. Assoc.*, **101**, 1418–1429.

Zou,H. and Hastie,T. (2005) Regularization and variable selection via the Elastic Net. *J. R. Stat. Soc. Ser. B*, **67**, 301–320.