ARTICLE

PLEIO: a method to map and interpret pleiotropic loci with GWAS summary statistics

Cue Hyunkyu Lee,[1,2] Huwenbo Shi,[3] Bogdan Pasaniuc,[4,5,6] Eleazar Eskin,[4,6,7] and Buhm Han[1,8,*]

Summary

Identifying and interpreting pleiotropic loci is essential to understanding the shared etiology among diseases and complex traits. A common approach to mapping pleiotropic loci is to meta-analyze GWAS summary statistics across multiple traits. However, this strategy does not account for the complex genetic architectures of traits, such as genetic correlations and heritabilities. Furthermore, the interpretation is challenging because phenotypes often have different characteristics and units. We propose PLEIO (Pleiotropic Locus Exploration and Interpretation using Optimal test), a summary-statistic-based framework to map and interpret pleiotropic loci in a joint analysis of multiple diseases and complex traits. Our method maximizes power by systematically accounting for genetic correlations and heritabilities of the traits in the association test. Any set of related phenotypes, binary or quantitative traits with different units, can be combined seamlessly. In addition, our framework offers interpretation and visualization tools to help downstream analyses. Using our method, we combined 18 traits related to cardiovascular disease and identified 13 pleiotropic loci, which showed four different patterns of associations.

Introduction

Genome-wide association studies (GWASs) have identified genetic variants associated with multiple traits, a phenomenon called pleiotropy.[1,2] The identification of pleiotropic loci is important to understanding the shared etiology among diseases and complex traits. Since GWAS summary statistics results are publicly available for many traits, these results can be used to find pleiotropic loci. Methods to identify pleiotropic loci are based on meta-analysis,[3–5] trait-specific effect size estimation,[6] or Bayesian approaches.[7] Methods based on meta-analysis give one p value per locus and are therefore convenient if the primary goal is identifying new risk loci. However, the meta-analysis results (the pooled statistic and the p value) alone are insufficient to determine the degree of association for each trait at a locus, making downstream interpretation (i.e., which trait is significant and which one is not) difficult. Trait-specific methods give an updated effect size and p value per trait per locus and thus have an advantage in the interpretation and risk prediction. However, an additional multiple testing correction may be required if one wants to obtain a single p value per locus. Here, we developed methods for identifying pleiotropic loci based on meta-analysis approaches.

Applying an existing meta-analysis method to multi-trait analyses is not optimal for several reasons. First, many existing meta-analysis methods do not adequately model the genetic architectures of complex traits. However, explicitly modeling genetic correlations across pairs of traits and their heritability can provide information on the direction and magnitude of effect sizes across different traits. Second, the meta-analysis methods depend on the scales and units of the phenotypes; the units often differ among quantitative traits, and the effect size definitions differ between binary and continuous traits. Most meta-analysis methods ignore the unit difference in effect size and use the observed effect size estimates as input. Therefore, they may not provide optimal results. For the same reason, interpretation tools such as the forest plot[5] or m-value[4] are less useful. Third, environmental correlations may exist among traits collected from the same individuals. Without systematically estimating and correcting for environmental correlations, a naïve application of meta-analysis methods can inflate false positives.

Here, we propose a multi-trait method to map and interpret pleiotropic loci called PLEIO (Pleiotropic Locus Exploration and Interpretation using Optimal test). As with meta-analysis methods, our method uses only GWAS summary statistics. Our method starts by estimating the genetic correlations, environmental correlations, and heritability for each trait from the whole-genome GWAS summary statistics. We then standardize the effect sizes of all traits and convert the effect sizes of binary traits to the liability scale. The standardization allows us to jointly analyze diseases and complex traits with different units and compare the magnitude of effect sizes. We assume that genetic effect is random and develop a test of

[1]Department of Biomedical Sciences, BK21 Plus Biomedical Science Project, Seoul National University College of Medicine, Seoul 03080, Republic of Korea; [2]Department of Convergence Medicine, University of Ulsan College of Medicine, Asan Medical Center, Seoul 05505, Republic of Korea; [3]Bioinformatics Interdepartmental Program, University of California, Los Angeles, Los Angeles, CA 90095, USA; [4]Department of Human genetics, University of California, Los Angeles, Los Angeles, CA 90095, USA; [5]Department of Pathology and Laboratory Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA; [6]Department of Computational Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA; [7]Department of Computer Science, University of California, Los Angeles, Los Angeles, CA 90095, USA; [8]Interdisciplinary Program in Bioengineering, Seoul National University, Seoul 03080, Republic of Korea
*Correspondence: buhm.han@snu.ac.kr
https://doi.org/10.1016/j.ajhg.2020.11.017.

non-zero genetic variance components where the covariance matrix is the cross-trait genetic covariance matrix. This test can take into account both the genetic correlations and heritabilities to maximize power and control false positive rate by accounting for environmental correlations. To increase computational efficiency in maximum likelihood estimation, we developed an optimization technique by using the spectral decomposition on the covariance matrix of the linearly transformed effect sizes. Even with this technique, obtaining the p value is computationally challenging because the small number of traits induces the small sample problem. We overcome this challenge by implementing an importance sampling method that provides accurate p value estimates.

We demonstrate the power of PLEIO in identifying pleiotropic loci by using both simulations and analysis of real traits. In simulations, PLEIO was consistently more powerful than other methods in almost all simulated genetic architectures because it could flexibly adapt to each genetic architecture, whereas other methods only performed well under certain genetic architectures. We applied PLEIO to combine 18 traits related to cardiovascular disease and identified 13 "novel" pleiotropic loci, i.e., loci not present in the GWAS catalog and not identified ($p_{GWAS} > 5\times10^{-8}$) by the original GWAS of the individual traits. These loci were categorized into four groups on the basis of their association patterns, which may represent distinct pathways. In addition to the powerful association test, PLEIO offers a visualization tool for the interpretation of the pleiotropic loci. PLEIO is publicly available to the research community.

Material and methods

PLEIO analysis in five steps

Here we describe our framework, PLEIO. PLEIO aggregates GWAS summary statistics of multiple traits to identify pleiotropic loci shared across traits. Suppose we have Q traits that we expect to share genetic components. We can collect T sets of genome-wide summary statistics for these traits. T can be greater than Q because more than one study can be included per trait. These traits can be a mixture of binary and quantitative traits whereby the quantitative traits can have differing phenotypic units. Suppose we have M SNPs that are shared by all studies we collected. Let $\widehat{\beta}_{it}$ denote the effect size estimate of the i^{th} SNP for the t^{th} study, $\text{SE}[\widehat{\beta}_{it}]$ denote the standard error estimate, and N_t denote the number of samples in the t^{th} study. Given this input, PLEIO performs a multi-trait joint analysis in the following five steps.

Step 1: decomposition of correlation

Correlation of GWAS marginal effect sizes can be attributable to correlation of causal genetic effect sizes and correlation of environmental effects. We decompose this correlation into genetic correlation $\mathbf{C_g}$ and environmental correlation $\mathbf{C_e}$ by applying cross-trait linkage disequilibrium score regression (ct-LDSC)[8] to each pair of studies. It is straightforward to estimate $\mathbf{C_g}$ and the heritabilities $\mathbf{h}^2$ from ct-LDSC. We combine $\mathbf{C_g}$ and $\mathbf{h}^2$ to get the genetic covariance matrix Ω.

We also use LDSC to estimate $\mathbf{C_e}$, which reflects the correlated errors of the effect size estimates driven by sample overlap.[8] We first correct the confounding factors of each trait by dividing the Z scores by the square root of the LDSC intercept. Then, the intercept of the ctLDSC (after running LDSC with –rg flag to compute genetic correlation) becomes the estimate of the correlation of environmental effects between the two traits. This approach was recently suggested by the MTAG.[6] We also describe another method for estimating $\mathbf{C_e}$. We combine a pair of traits with fixed effects meta-analysis based on the inverse variance of the effect size. Then, we apply this pooled summary statistic to get the LDSC intercept (by running LDSC with –h2 flag to compute heritabililty). Given the LDSC intercept α_{meta}, the environmental correlation is

$$\rho_e \approx \frac{N_j + N_k}{2\sqrt{N_j N_k}}(\alpha_{meta} - 1)$$

where N_j and N_k are the sample sizes of the two studies. We found that the two approaches give similar estimates. For details, see Supplemental methods.

Step 2: standardization of effect sizes

In the input data, the scales of effect sizes can be heterogeneous across the studies. We calculate the standardized effect sizes of SNP i for the trait t as

$$\widehat{\eta}_{it} = \frac{\sqrt{\delta_t}\,\frac{\widehat{\beta}_{it}}{SE[\widehat{\beta}_{it}]}}{\sqrt{N_t + \delta_t\theta_t\left[\frac{\widehat{\beta}_{it}}{SE[\widehat{\beta}_{it}]}\right]^2}},\ \text{and}\ \text{SE}[\widehat{\eta}_{it}] = \frac{SE[\widehat{\beta}_{it}]}{\widehat{\beta}_{it}}\widehat{\eta}_{it}. \quad \text{(Equation 1)}$$

δ_t is a scaling factor that is 1 for quantitative traits and $(K_t^2(1-K_t)^2/P_t(1-P_t))\cdot(1/[\psi(\phi^{-1}(1-K_t))]^2)$ for binary traits, where K_t refers to the disease prevalence, $P_t = (N_t|y=1)/N_t$ refers to the sample prevalence, ψ refers to the probability density function of the standard normal distribution, and ϕ^{-1} refers to the inverse of the cumulative density function of the standard normal distribution. θ_t is an additional scaling factor that is 0 for quantitative traits and $(i_t\times(P_t - K_t/1-K_t))(i_t\times(P_t - K_t/1-K_t) - t)$ for binary traits, where $i_t = (\psi(\phi^{-1}(1-K_t))/K_t)$ refers to the mean liability of cases, and $t = \phi^{-1}(1-K_t)$ refers to the liability threshold for cases. For quantitative traits, $\widehat{\eta}_{it}$ can be simplified to $\widehat{\eta}_{it} = (\widehat{\beta}_{it}/SE[\widehat{\beta}_{it}])\cdot(1/\sqrt{N_t})$, which corresponds to the effect size based on the standardized phenotypes and the standardized genotypes. For binary traits, $\widehat{\eta}_{it}$ is the effect size for the liability, assuming that the Z score $(\widehat{\beta}_{it}/SE[\widehat{\beta}_{it}])$ was obtained from a linear model with an observed scale (by setting the phenotypes 0 and 1). The use of the two scaling factors (δ_t and θ_t) in a non-randomly ascertained case-control study using a linear model was suggested by Lee et. al.[9] Typically, the Z scores come from the logistic regression model rather than the observed scale linear model. However, it is a common practice to use these Z scores as if they came from the linear model.[10] $\widehat{\eta}_{it}$ can be used conveniently to interpret the pleiotropic effects of a variant because, in contrast to the original effect size, $\widehat{\beta}_{it}$, it is independent of the units of phenotypes. We verified the accuracy of the proposed scaling for different combinations of population prevalence and sample prevalence (Figure S1) in a simulation setting similar to one used by Choi et al.[11]

Step 3: mapping pleiotropic loci with a variance component test

We build a statistical model optimized for the identification of pleiotropic loci. We assume that an individual phenotype is influenced by K causal SNPs whose individual contribution is very small. For simplicity, we assume that K causal SNPs are shared by

T traits. Let $\boldsymbol{\eta}_i$ denote a $T\times1$ vector of the true effect sizes of the causal SNP i under the standardized scale. Following the common model widely used in previous studies,[6,10,12] we assume that K SNPs have equal contributions. Then, $\boldsymbol{\eta}_i \sim \mathrm{MVN}\left(0, \frac{\Omega}{K}\right)$, where Ω denotes the genetic covariance matrix, of which diagonal elements are the narrow sense heritabilities. We assume $\boldsymbol{\eta}_i = 0$ for non-casual SNPs.

Let $\widehat{\boldsymbol{\eta}}_i$ denote the observed effect sizes and $SE(\widehat{\boldsymbol{\eta}}_i)$ denote the standard errors. We can model $\widehat{\boldsymbol{\eta}}_{\boldsymbol{i}}$ as the sum of the true genetic effect and the error:

$$\widehat{\boldsymbol{\eta}}_{\boldsymbol{i}} = \boldsymbol{\eta}_i + \boldsymbol{\epsilon}_{\boldsymbol{i}},$$

where $\boldsymbol{\epsilon}_{\boldsymbol{i}}$ is a random variable denoting the error, which follows $\boldsymbol{\epsilon}_{\boldsymbol{i}} \sim \mathrm{MVN}(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = diag(SE[\widehat{\boldsymbol{\eta}}_{\boldsymbol{i}}])\cdot\mathbf{C}_{\boldsymbol{e}}\cdot diag(SE[\widehat{\boldsymbol{\eta}}_{\boldsymbol{i}}])$. Thus, $\mathrm{Var}(\widehat{\boldsymbol{\eta}}_{\boldsymbol{i}}) = \frac{\Omega}{\boldsymbol{K}} + \boldsymbol{\Sigma}$ for causal SNPs and $\mathrm{Var}(\widehat{\boldsymbol{\eta}}_{\boldsymbol{i}}) = \boldsymbol{\Sigma}$ for non-causal SNPs. As described earlier, applying LDSC to $\widehat{\boldsymbol{\eta}}_i$ and $SE(\widehat{\boldsymbol{\eta}}_i)$ of all M SNPs can produce an estimate of the genetic covariance matrix ($\widehat{\Omega}$) as well as the error correlation ($\widehat{\mathbf{C}_{\boldsymbol{e}}}$).

We then relax the assumption that K SNPs have equal contributions. Then, the true effect $\boldsymbol{\eta}_i$ needs not have the fixed variance $\frac{\Omega}{K}$. We now model $\widehat{\boldsymbol{\eta}}_{\boldsymbol{i}}$ as

$$\widehat{\boldsymbol{\eta}}_{\boldsymbol{i}} = \boldsymbol{\gamma}_{\boldsymbol{i}} + \boldsymbol{\epsilon}_{\boldsymbol{i}},$$

where $\boldsymbol{\gamma}_{\boldsymbol{i}}$ is a new random variable denoting the genetic effect that follows $\boldsymbol{\gamma}_{\boldsymbol{i}} \sim \mathrm{MVN}(0, \tau_i^2\Omega)$, where $\tau_i^2 > 0$ for causal SNPs and $\tau_i^2 = 0$ for non-causal SNPs. That is, the scaling factor τ_i^2 of the variance can model SNP-by-SNP differences in genetic contributions. As a special case, if we set $\tau_i^2 = \frac{1}{K}$ for K causal SNPs and $\tau_i^2 = 0$ for non-casual SNPs, this model is reduced to the previous model assuming equal contributions of causal SNPs. Note that although we relaxed the assumption of the equal contribution, the variance of $\boldsymbol{\gamma}_{\boldsymbol{i}}$ is still proportional to Ω, which models the relative heritability differences of the traits and the genetic correlations among the traits. Under this model, testing whether a SNP is causal or not corresponds to testing the null hypothesis $\tau_i^2 = 0$ versus the alternative hypothesis $\tau_i^2 > 0$.

The underlying intuitions of our model are as follows. Our key assumption is that the genetic component $\boldsymbol{\gamma}_{\boldsymbol{i}}$ in the effect size is a random variable whose variance is proportional to the genetic covariance matrix Ω. This implies that (1) phenotypes with larger heritability show larger genetic effects and (2) phenotypes show genetic effects concordant to their genetic correlations. Because $\widehat{\Omega}$ and $\widehat{\boldsymbol{\Sigma}}$ are summarized information from the whole genome, this approach can maximize the overall power. In that sense, our model resembles empirical Bayes approaches.[7]

To test the hypothesis $\tau_i^2 > 0$, we can fit a variance component model to get the maximum likelihood estimate (MLE) $\widehat{\tau}_i^2$ that maximizes $\mathcal{L}(\tau_i^2|\widehat{\eta}_i; \widehat{\Omega}, \widehat{\boldsymbol{\Sigma}})$. A numerical optimization algorithm such as the pseudo Newton-Raphson method can be used to find $\widehat{\tau}_i^2$. However, updating the value of the likelihood function at each iteration requires a matrix inversion. With a large T, this can significantly increase the overall analysis time. To solve this challenge, we developed an optimization technique that considerably reduces the computational burden for finding the MLE (see Supplemental methods). The proposed optimization method carries out a linear transformation on $\widehat{\eta}_i$ via $\widehat{\Omega}^{-\frac{1}{2}}$. The transformed observed effect sizes follow

$$\widehat{\Omega}^{-\frac{1}{2}}\widehat{\boldsymbol{\eta}}_{\boldsymbol{i}} \sim \mathbf{MVN}\left(0, \tau_i^2\boldsymbol{I} + \widehat{\Omega}^{-\frac{1}{2}}\widehat{\boldsymbol{\Sigma}}\widehat{\Omega}^{-\frac{1}{2}}\right),$$

where the corresponding $\widehat{\tau}_i^2$ maximizes $\mathcal{L}\left(\tau_i^2\middle|\widehat{\Omega}^{-\frac{1}{2}}\widehat{\boldsymbol{\eta}}_{\boldsymbol{i}}; \widehat{\Omega}^{-\frac{1}{2}}\widehat{\boldsymbol{\Sigma}}\widehat{\Omega}^{-\frac{1}{2}}\right)$ under the constraint of $\tau_i^2 > 0$. We apply a spectral decomposition $\boldsymbol{D} = \widehat{\Omega}^{-\frac{1}{2}}\widehat{\boldsymbol{\Sigma}}\widehat{\Omega}^{-\frac{1}{2}} = \boldsymbol{P_D}(\boldsymbol{\Lambda_D})\boldsymbol{P_D^T}$, where $\boldsymbol{\Lambda_D}$ is a diagonal matrix of the eigenvalues, the diagonal elements of which are arranged in ascending order, and $\boldsymbol{P_D}$ is an eigenvector matrix, the i^{th} column of which corresponds to the i^{th} eigenvalue. Then, we only need to calculate $\boldsymbol{P_D}(\boldsymbol{\Lambda_D} + \tau_i^2\boldsymbol{I})^{-1}\boldsymbol{P_D^T}$ per each iteration, which is much easier to calculate than $(\tau_i^2\widehat{\Omega} + \widehat{\boldsymbol{\Sigma}})^{-1}$. Note that the values of the matrices $\boldsymbol{P_D}$ and $\boldsymbol{\Lambda_D}$ remain unchanged with iterations. The log-likelihood function obtained through the linear transformation (ℓ_1') can be shown as follows:

$$\begin{aligned}\ell_1' &= -\frac{1}{2}\left[T\ln(2\pi) + \sum_{t=1}^{p}\ln(\xi_t + \tau_{\boldsymbol{i}}^2)\right.\\ &\quad\left. + \left(\boldsymbol{P_D}\boldsymbol{E}\left[\widehat{\Omega}^g\right]^{\frac{1}{2}}\widehat{\boldsymbol{\eta}}_{\boldsymbol{i}}\right)^{\boldsymbol{T}}\left[\boldsymbol{\Lambda}_{\boldsymbol{D}}^{+}\right]^{-1}\left(\boldsymbol{P_D}\boldsymbol{E}\left[\widehat{\Omega}^g\right]^{\frac{1}{2}}\widehat{\boldsymbol{\eta}}_{\boldsymbol{i}}\right)\right]\\ &= -\frac{1}{2}\left[T\ln(2\pi) + \sum_{t=1}^{p}\ln(\xi_t + \tau_{\boldsymbol{i}}^2) + \sum_{t=1}^{p}\frac{\delta_t^2}{\xi_t + \tau_i^2}\right],\end{aligned}$$

where p is the number of non-zero eigenvalues, ξ_t is the t^{th} diagonal element of $\boldsymbol{\Lambda_D}$, δ_t^2 is the t^{th} element of the vector $\boldsymbol{P_D}\boldsymbol{E}\widehat{\Omega}^{-\frac{1}{2}}\widehat{\boldsymbol{\eta}}_{\boldsymbol{i}}$, and $\boldsymbol{E}$ is a diagonal matrix of which the first p elements are 1 and the rest are 0.

The first and second derivatives of ℓ_1' with respect to τ_i^2 are as follows:

$$\frac{d\ell_1'}{d\tau_{\boldsymbol{i}}^2} = -\frac{1}{2}\left[\sum_{t=1}^{p}\frac{1}{\xi_t + \tau_{\boldsymbol{i}}^2} - \sum_{t=1}^{p}\frac{\delta_t^2}{(\xi_t + \tau_i^2)^2}\right]$$

$$\frac{d^2\ell_1'}{d(\tau_{\boldsymbol{i}}^2)^2} = \frac{1}{2}\left[\sum_{t=1}^{p}\frac{1}{(\xi_t + \tau_{\boldsymbol{i}}^2)^2} - 2\sum_{t=1}^{p}\frac{\delta_t^2}{(\xi_t + \tau_i^2)^3}\right].$$

The optimal $\widehat{\tau}_i^2$ can be obtained with the Newton Raphson method. As a result, we get the log-likelihood ratio test (LRT) statistic

$$S_{PLEIO} = \left[\sum_{t=1}^{p}\ln\left(\frac{\xi_t}{\xi_t + \widehat{\tau}_i^2}\right)\right] + \left[\sum_{t=1}^{p}\frac{\delta_t^2}{\xi_t} - \sum_{t=1}^{p}\frac{\delta_t^2}{\xi_t + \widehat{\tau}_i^2}\right].$$

This technique can substantially shorten the time to complete our test, and the time reduction increases with increasing number of traits (Figures S2 and S3). We note that our technique was inspired by the technique used in efficient mixed-model association (EMMA).[13] Although the exact model and formulation are different, the general scheme using eigen decomposition to simplify the problem to one-dimensional search is the same.

Step 4: assessing statistical significance via importance sampling

Here, we describe how to assess an accurate p value of S_{PLEIO} that asymptotically follows a 50 : 50 mixture of χ_0^2 and χ_1^2.[14] However, the asymptotic approximation is not accurate if the number of traits (T) is small. We found that even when T is as large as 100, the null p values calculated from asymptotic distribution deviate from uniform distribution (Figure S4). Moreover, it turns out that the null distribution depends on the genetic covariance matrix $\widehat{\Omega}$ and error correlation matrix $\widehat{\boldsymbol{\Sigma}}$. Thus, an alternative approach would be simulating null distributions on the basis of these study-specific factors ($\widehat{\Omega}$ and $\widehat{\boldsymbol{\Sigma}}$). But the standard sampling is overly inefficient for assessing very small p values (e.g., 5×10^{-8}).

Instead, we use an importance sampling approach to assess the p value of S_{PLEIO}. Let $\boldsymbol{x}$ be a random variable denoting the standardized effect sizes. Let $q(\boldsymbol{x})$ denote the probability density function (PDF) of $\boldsymbol{x}$ under the null. By definition, $\int_B q(\boldsymbol{x})d\boldsymbol{x} = 1$ where $B = \mathbb{R}^T$. We can consider S_{PLEIO} as a function of $\boldsymbol{x}$ given $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\Omega}$. Given an observed S_{PLEIO} statistic from data, which we call θ, we want to calculate the p value of it. To this end, let $f(\boldsymbol{x}, \theta)$ denote an indicator function as follows:

$$f(\boldsymbol{x}, \theta|\widehat{\boldsymbol{\Sigma}}, \widehat{\Omega}) = \begin{cases} 1 \text{ if } S_{PLEIO}(x|\widehat{\boldsymbol{\Sigma}}, \widehat{\Omega}) \geq \theta \\ 0 \text{ if } S_{PLEIO}(\boldsymbol{x}|\widehat{\boldsymbol{\Sigma}}, \widehat{\Omega}) < \theta \end{cases}.$$

For simplicity, we replace $f(\boldsymbol{x}, \theta|\widehat{\boldsymbol{\Sigma}}, \widehat{\Omega})$ with a simpler expression, $f(\boldsymbol{x})$. The p value of θ can be expressed as

$$I = \int_B f(\boldsymbol{x})q(\boldsymbol{x})d\boldsymbol{x},$$

To estimate I, we can exploit the importance sampling algorithm. In importance sampling, we use a sampling distribution $p(\boldsymbol{x})$ that differs from $q(\boldsymbol{x})$. Let $\boldsymbol{X^p} \sim p(\boldsymbol{x})$ denote a $M \times T$ matrix of the sampled effect sizes generated from $p(\boldsymbol{x})$, where M is the number of sampling. Then, we can estimate I by using $\boldsymbol{X^p}$ as follows:

$$\widehat{I} = \mathrm{E}^p\left[\frac{f(\boldsymbol{x})q(\boldsymbol{x})}{p(\boldsymbol{x})}\right] = \frac{1}{M}\left[\sum_{i=1}^{M}\frac{f(\boldsymbol{X_i^p})q(\boldsymbol{X_i^p})}{p(\boldsymbol{X_i^p})}\right],$$

where $\mathrm{E}^p[\cdot]$ denotes the expectation over $\boldsymbol{X^p}$, and $\boldsymbol{X_i^p}$ is the i^{th} row vector of $\boldsymbol{X^p}$.

The challenge in importance sampling is choosing an appropriate $p(\boldsymbol{x})$. It is particularly challenging in GWASs because the range of p values is very wide, from 1.0 to 5×10^{-8}. Thus, it is difficult to select a single distribution that can minimize variance for all range of p values. To solve this challenge, we applied the importance sampling method developed by Owen and Zhou.[15] The method generates samples from a mixture distribution. Let $p_j(\boldsymbol{x})$ denote the j^{th} sampling distribution where $j = \{1, 2, \ldots, F\}$, and let $p_\alpha(\boldsymbol{x})$ denote the mixture distribution of F sampling distributions. We select F distributions so that the variance can be reduced for a wide range of p values. We assume that each sampling distribution has the equal chance to generate a sample such that $p_\alpha(\boldsymbol{x}) = \frac{1}{F}\sum_{j=1}^{F} p_j(\boldsymbol{x})$. Detailed information on the selection of $p_j(\boldsymbol{x})$ can be found in Supplemental methods. Here, we use $\frac{p_j(\boldsymbol{x})}{p_\alpha(\boldsymbol{x})}$ as a control variate of $m(\boldsymbol{x}) = \frac{f(\boldsymbol{x})q(\boldsymbol{x})}{p_\alpha(\boldsymbol{x})}$. Then, we can define

$$m^*(\boldsymbol{x}, \boldsymbol{\beta}) = \frac{f(\boldsymbol{x})q(\boldsymbol{x})}{p_\alpha(\boldsymbol{x})} - \sum_{j=1}^{K}\beta_j\left(\frac{p_j(\boldsymbol{x})}{p_\alpha(\boldsymbol{x})} - \mu_{p_j}\right),$$

where $\mathrm{E}[m^*] = \mathrm{E}[m] = I$ and $\mu_{p_j} = E^{p_\alpha}\left[\frac{p_j(\boldsymbol{x})}{p_\alpha(\boldsymbol{x})}\right] = \int_B p_j(\boldsymbol{x})d\boldsymbol{x} = 1$. The control variate method maximizes the variance reduction of $\mathrm{Var}(m^*)$ by using the optimal control variate coefficient ($\boldsymbol{\beta}^*$). Then, the variance $\mathrm{Var}(m^*)$ becomes equal to or smaller than $\mathrm{Var}(m)$. Owen and Zhou[15] showed that the p value estimate of θ can be shown as follows:

$$\widehat{I} = \mathrm{E}^{p_\alpha}[m^*] = \frac{1}{M}\left(\sum_{i=1}^{M}\frac{f(\boldsymbol{X_i^p})q(\boldsymbol{X_i^p}) - \sum_{j=1}^{K}\beta_j p_j(\boldsymbol{X_i^p})}{p_\alpha(\boldsymbol{X_i^p})}\right) + \sum_{k=1}^{K}\beta_k.$$

Given $\boldsymbol{X^p}$ from $p_\alpha(\boldsymbol{x})$, we calculate p values of 40 different θ that are in the range $(0, 40)$, which roughly correspond to p values from 1.0 to 3×10^{-11}. For each θ, we calculate the optimal $\boldsymbol{\beta}$ for the control variate method to maximize the variance reduction of the p value estimate. See Supplemental methods for how we obtained the optimal control variate coefficients ($\boldsymbol{\beta}^*$). Using these 40 points, we interpolate p values for $\theta < 40$ by using B-spline fit and extrapolate p values for $\theta > 40$ by using the linear fit on the logarithmic p value scale.

In our method, we generate null samples once and use them for all SNPs. One challenge with this procedure is that, by definition, $\widehat{\boldsymbol{\Sigma}}$ is dependent on SNP i, as shown in Equation 1, if the trait is binary. Although the proposed scaling scheme can accurately convert $\widehat{\beta}_{it}$ into $\widehat{\eta}_{it}$ (see Figure S1), the drawback is that it imposes a dependency between $\widehat{\boldsymbol{\Sigma}}$ and i. To overcome this challenge, in the null sample generation, we assume that all traits are quantitative. That is, we use approximations $\widehat{\eta}_{it} = \frac{\widehat{\beta}_{it}}{SE[\widehat{\beta}_{it}]} \times \frac{1}{\sqrt{N_t}}$ and $\mathrm{SE}[\widehat{\eta}_{it}] = \frac{1}{\sqrt{N_t}}$ for all traits so that $\widehat{\boldsymbol{\Sigma}} = diag\left(\frac{1}{\sqrt{\boldsymbol{N}}}\right)\cdot\mathbf{C_e}\cdot diag\left(\frac{1}{\sqrt{\boldsymbol{N}}}\right)$ when $\boldsymbol{N} = \{N_1, N_2, \ldots, N_T\}$. Under this assumption, $\widehat{\boldsymbol{\Sigma}}$ becomes independent of SNP i, and therefore, the null samples generated once can be used for all SNPs. We empirically confirmed that the use of this approximation does not much affect the robustness of the false positive rate control (data not shown).

Step 5: pleiotropy plot

PLEIO offers a tool to visualize the pleiotropic effects of a SNP, which we named "pleiotropy plot" (Figure S5). This circular plot provides information about the standardized effect sizes, the local heritabilities, and the local Manhattan plots of a SNP. The outer part is partitioned by the traits, each of which contains (1) the effect size of each trait on the original scale as text and on the standardized scale as a horizontal bar and (2) the local Manhattan plot within a 1 Mb window. The inner part is a ribbon plot linking multiple traits. The ribbon color indicates genetic correlations. The ribbon width at the end indicates the relative locus-heritability per trait (squared standardized effect size), where the width of the largest locus-heritability is adjusted to 100%.

Data analysis

Collection of GWAS summary statistics

We collected public GWAS summary statistics of 18 diseases and complex traits related to cardiovascular disease from large-scale genetic consortia, as described in Table S1. When a consortium database contained more than one GWAS for the same phenotype, we selected the most recent study. We obtained the summary statistics of four quantitative traits from the Global Lipids Genetics consortium.[16] The data consisted of the results of GWASs from 94,595 individuals from 23 studies genotyped with GWAS arrays and 93,982 individuals from 37 studies genotyped with the Metabochip array. We obtained the summary statistics of the twelve binary traits in the UK biobank data from the Neale lab website (Table S2). The data consisted of the results of GWASs from 361,193 individuals in the UK biobank cohort. We obtained the summary data on coronary artery disease (CAD) from the CARDIo+C4D consortium.[17] The data consisted of the results of GWAS meta-analysis from 60,801 CAD-affected individuals and 123,504 control individuals from 48 studies. We obtained the summary data on fasting glucose (FG) from MAGIC (Meta-Analysis of Glucose and Insulin-related traits Consortium).[18] The data consisted of the analysis results from 46,186 non-diabetic patients from 21 GWASs. All samples were from individuals of European descent, except for those from the participants included in the CAD data

from CARDIo+C4D consortium. For CAD data, participants were a mixture of European ancestry (77%), South Asian ancestry (India and Pakistan; 13%), East Asian ancestry (China and Korea; 6%), and others (Hispanic and African American; 4%).[17]

Summary statistics data quality control

For each summary statistics dataset, we removed SNPs that were not included in 1000 Genomes.[19] We checked the consistency of allele pair of each SNP with the corresponding allele pair of the SNP in 1000 Genomes. To eliminate potential strand mismatches, we pruned SNPs with the allele pair GC and AT. The genetic covariance and error correlation were estimated from summary statistics of the remaining SNPs. A total of 1,777,411 SNPs was included in the joint analysis of 18 traits.

Identification of novel pleiotropic loci

In the joint analysis of 18 traits, we identified 7,932 SNPs that were genome-wide significant ($p_{PLEIO} < 5\times10^{-8}$). We clumped these SNPs with threshold ($r^2 < 0.1$) and found 625 approximately independent hits (Table S3). To estimate LD between SNPs, we used the European samples in the 1000 Genomes data. To determine whether the remaining variants were novel loci, we excluded variants that met any of the following two conditions: (1) the variant had a moderate LD ($r^2 > 0.1$) with a variant that is listed in the GWAS catalog as associated with the CVD-related traits or (2) the variant already reached the genome-wide significance threshold of 5×10^{-8} in the original summary statistics of a single trait. From this, we identified 13 "novel" pleiotropic variants (Table S4). For GWAS catalog summary data, we used the file named "All associations v1.0," downloaded on September 3, 2020.

Results

Overview of method

PLEIO is a multi-trait framework to map and interpret pleiotropic loci. PLEIO estimates the genetic covariance and environmental covariance from GWAS summary statistics data and uses this information to increase the power of association test (Figure S6). Consider a toy example that involves three traits (A, B, and C) (Figure S7). At SNP X_1, we observed the effect sizes of $(2.2, 2.8, -1.2)$, and at another SNP X_2, we observed the effect sizes of $(-1.5, 0.4, -2.7)$. For simplicity, we assume that the variances of all estimates were one. Then, if we apply the fixed effects meta-analysis (inverse-variance method), we get the same p value for both SNPs ($p = 0.03$) because the average effect size is the same. However, suppose we know that traits A and B have a positive genetic correlation and trait C has a negative genetic correlation with the rest. Then, SNP X_1 is more likely to be a true signal than SNP X_2 because the effect directions conform to the genetic correlations. Moreover, suppose we know that trait B has the largest heritability and trait C has the smallest heritability. Then, the association at SNP X_1 is even more likely because the relative strengths of the effect size conform to the heritabilities. Our method accounts for both the genetic correlations and heritabilities and gives a more significant p value at SNP X_1 ($p = 0.0006$) than SNP X_2 ($p = 0.1$).

PLEIO consists of five steps. First, we apply the LDSC[10] to the genome-wide summary data of traits to obtain the genetic correlations $\mathbf{C_g}$, the environmental correlation $\mathbf{C_e}$, and the heritabilities $\mathbf{h}^2$. We summarize $\mathbf{C_g}$ and $\mathbf{h}^2$ into the genetic covariance Ω. Second, we transform the effect sizes $\widehat{\boldsymbol{\beta}}$ into the standardized effect sizes $\widehat{\boldsymbol{\eta}}$, converting the effect sizes of binary traits to the effect sizes for liabilities. Third, we apply our variance component test to map pleiotropic loci. We assume $\widehat{\boldsymbol{\eta}} = \mathbf{g} + \mathbf{e}$, where $\mathbf{g}$ is the genetic effect and $\mathbf{e}$ is the error (Figure S6). Our main assumption is that the genetic effects follow the genetic covariance, $\mathrm{Var}(\mathbf{g}) = \tau^2\Omega$. We then test the hypothesis $\tau^2 > 0$ versus $\tau^2 = 0$. To find the MLE $\widehat{\tau}^2 \geq 0$ efficiently, we utilize an optimization technique by using spectral decomposition of the variance. Fourth, we apply an importance sampling method to assess the one-tailed p value. Fifth, we report and visualize the results to help interpretation.

Evaluation of false positive rates in null simulations

We evaluated the false positive rate (FPR) of PLEIO by using simulations. We assumed the null hypothesis of no genetic effect at a SNP for all T traits. Overall, we varied four factors: (1) the number of traits (T), (2) the environmental correlation matrix ($\mathbf{C_e}$), (3) the heritability parameter for PLEIO ($\mathbf{h}^2$), and (4) the genetic correlation parameter for PLEIO ($\mathbf{C_g}$). Note that $\mathbf{h}^2$ and $\mathbf{C_g}$ are input parameters for PLEIO describing what PLEIO thinks to be the true $\mathbf{h}^2$ and $\mathbf{C_g}$ but are not the actual $\mathbf{h}^2$ and $\mathbf{C_g}$ because the true $\mathbf{h}^2$ is zero in this null simulation. In a real analysis of PLEIO, $\mathbf{h}^2$ and $\mathbf{C_g}$ are estimated from GWAS statistics and given to the test method. The test method combines them to genetic covariance (Ω) and performs a variance component test. Because PLEIO's test method depends on the input parameters $\mathbf{h}^2$ and $\mathbf{C_g}$, we wanted to evaluate FPR when different $\mathbf{h}^2$ and $\mathbf{C_g}$ were given.

Specifically, we simulated three different numbers of traits ($T = 5, 10, 20$). We set the off-diagonal elements of $\mathbf{C_e}$ to 0.0 and 0.5 to simulate uncorrelated and correlated environmental effects, respectively. We simulated two different patterns of $\mathbf{h}^2$. In the "equal $\mathbf{h}^2$" scenario, we set genome-wide heritability to be the same ($\boldsymbol{h}^2 = 0.5$) for all traits. In the "different $\mathbf{h}^2$" scenario, we simulated heritabilities ranging from 0.1 to 0.5. We simulated two different patterns of $\mathbf{C_g}$. In the "uniform $\mathbf{C_g}$" scenario, the off-diagonal elements of $\mathbf{C_g}$ were all set to 0.3. In the "partitioned $\mathbf{C_g}$" scenario, we set two subgroups and set off-diagonal elements to 0.3 within a group and 0 between groups. Thus, we tested 24 different scenarios ($3\times2\times2\times2$). We generated one million null datasets for each situation and calculated the FPR at $\alpha = 0.05$. Table S5 shows that the FPR of PLEIO is well calibrated in all situations.

Next, we examined the FPR at a lower threshold. We increased the number of null datasets to a billion to measure the FPR at the conventional GWAS threshold (5×10^{-8}). We tested three numbers of traits ($T = 5, 10, 20$) while assuming the equal $\mathbf{h}^2$, partitioned

$\mathbf{C_g}$, and no sample overlap. Table S6 shows that PLEIO's FPR is well calibrated for α down to 5×10^{-8}.

So far, we directly simulated effect sizes without generating genotypes. See Supplemental methods for a detailed explanation for the simulation. We confirmed that when we actually generated genotypes under the null, the results were similar and the FPR was controlled regardless of the minor allele frequency (Table S7).

Evaluation of power in alternate simulations

We compared the power of PLEIO against two meta-analysis approaches: the fixed effects meta-analysis method and ASSET.[3] For the fixed effects method, we used the inverse-variance method of METAL.[20] We used our own R code implementation because the original METAL code cannot account for the environmental correlation due to sample overlap. We implemented the strategy suggested by Lin and Sullivan,[21] which can be thought of as a general extension of METAL. ASSET is a subset-based method assuming that the true effects could only exist in a subset of the studies. We confirmed that FPRs were well calibrated with both meta-analysis methods (Table S8).

Additionally, we compared the power with a trait-specific approach, MTAG.[6] Unlike other meta-analysis approaches, MTAG gives T p values given T traits. Because we measured the power as the proportion of simulations whose p value exceeds a threshold, we needed to combine T p values into one p value. A straightforward approach was to choose the minimum p value. However, additional multiple testing burden was required with this approach. When we measured the FPR, indeed, the FPR was inflated by choosing the minimum p value (Table S8). To correct for multiple testing, we applied the Bonferroni correction by multiplying the minimum p value by T. This approach controlled the FPR but was conservative because the T effect size estimates were correlated (Table S8). In our simulation, we measured the power of MTAG both before the Bonferroni correction (MTAG-U; uncorrected) and after the Bonferroni correction (MTAG-C; corrected). Because MTAG-U is anti-conservative and MTAG-C is conservative, they can give upper and lower bounds of the power of MTAG. We used our own Python code that implements the MTAG method because the MTAG software thought that the input was flawed if the median of the Z scores was far from zero, which was the case in the power simulations.

We assessed the power of the methods in various simulation settings. Each setting defined a specific genetic correlation structure $\mathbf{C_g}$, heritabilities $\mathbf{h}^2$, phenotypic units ($\boldsymbol{U}$), and the types of traits (quantitative [Q] or binary [B]). In each setting, we assumed $T = 7$ traits and repeated simulations 10,000 times. The power was estimated as the proportion of the simulations in which the p value was $< 5\times10^{-8}$. We assumed that the true $\mathbf{C_g}$ and $\mathbf{h}^2$ were provided to PLEIO and MTAG. In power simulations, we generated actual genotypes instead of directly sampling effect sizes from a distribution. See Supplemental methods for a detailed explanation for the simulation.

First, we assumed a fixed heritability and perfect correlations ($r^2 = 1.0$) among the seven traits. This represents the scenario in which the same traits were collected in multiple studies. In this situation, all methods performed similarly well except ASSET (Figure 1A). With a sample size of $N = 50,000$, the powers of PLEIO, METAL, MTAG-U, MTAG-C, and ASSET were 63.79%, 63.81%, 63.81%, 61.67%, and 30.66%, respectively. As expected, METAL performed well because METAL is optimal for the fixed effect scenario. PLEIO attained similar power, within the 95% confidence interval with METAL, because it can account for the genetic correlations. In this situation, MTAG was analytically equivalent to METAL.[6] Because the T p values of MTAG are identical in this scenario, the multiple testing correction is not needed. Thus, MTAG-U represents the correct power of MTAG, while MTAG-C is overly conservative.

Second, we simulated different heritabilities for seven traits, varying from 0.005 to 0.7. We simulated a uniform genetic correlation $r = 0.5$ between all trait pairs. In this scenario, PLEIO outperformed the other methods (Figure 1B). With a sample size of $N = 50,000$, PLEIO achieved a power of 77.6%, while the second-best method (MTAG-U) achieved 67.2% and the third-best method (MTAG-C) achieved 62.7%. PLEIO achieved higher power than METAL because PLEIO accounts for different heritabilities of the traits.

Third, we simulated a complex correlation pattern with both negative and positive correlations. We divided seven traits into two groups (three traits and four traits). We set the correlations in the first group to 0.95 and the correlations in the second group to 0.90. We set the correlations between the groups to a negative value of -0.9. We assumed a uniform heritability of 0.3 for all traits. PLEIO showed the highest power among all methods (Figure 1C). With a sample size of $N = 50,000$, PLEIO achieved a power of 78.6%, while the second-best method (MTAG-U) achieved 66.3% and the third-best method (MTAG-C) achieved 62.6%. PLEIO achieved higher power than METAL because PLEIO can take into account the genetic correlation structure of the traits.

Fourth, we simulated a mixture of quantitative and binary traits. We simulated four quantitative traits and three binary traits. For quantitative traits, we simulated different phenotypic units ranging from $0.1U$ to $10U$, where U was the standard unit we assumed. We simulated a fixed heritability of 0.5 and a uniform genetic correlation of 0.5 for all traits. Again, PLEIO achieved the highest power (Figure 1D). With a sample size of $N = 50,000$, PLEIO achieved a power of 83.6%, while the second-best method (MTAG-U) achieved 67.0% and the third-best method (MTAG-C) achieved 60.8%. PLEIO achieved higher power than METAL because PLEIO systematically combines heterogeneous traits by standardizing the effect sizes.

So far, we varied only one factor in each simulation: different heritabilities, a complex pattern of genetic correlations, and different phenotypic units. In reality, all three

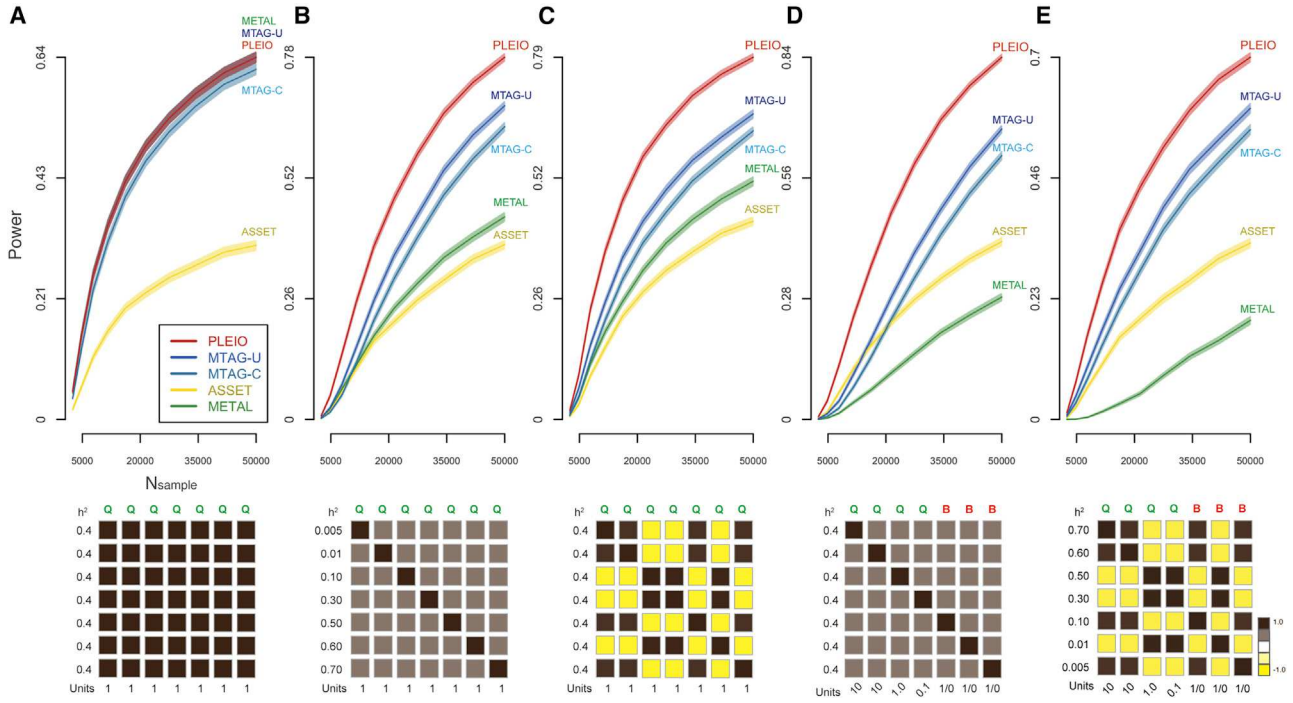

Figure 1. The simulation result comparing the performance of PLEIO and other methods
(A–E) We performed a total five power tests and labeled the results with (A)–(E). Each line shows the statistical power of a model gained from an association test using seven summary statistics. We compared PLEIO (red), MTAG-U (blue), MTAG-C (light blue), METAL (green), and ASSET (yellow). At the bottom of the figure, we visualized the simulation setting of each test. The boxplot shows the genetic correlation. "Q" and "B" indicate whether the phenotype is quantitative or binary. The heritability values of the traits are shown on the left side of the boxplot. The trait phenotype units are shown at the bottom of the boxplot. The line thickness indicates the 95% confidence interval.

can occur together. We simulated such a combined situation. With a sample size of 50,000, PLEIO achieved a power of 69.7%, while the power of the second-best method (MTAG-U) was 59.9% (Figure 1E).

Next, we wanted to simulate with real-data-based parameters. To this end, we assumed that there is one focal trait of interest and we want to borrow information from multiple non-focal traits. Non-focal traits are selected so that they are closely correlated with focal trait, but the correlation between non-focal traits may not necessarily be strong. Since MTAG gives trait-specific p values, we can use MTAG to only look at the p value of the focal trait, which we call MTAG-F.

Here, we used low density lipoprotein (LDL) as the focal trait and selected the following six non-focal traits that have a strong association with LDL ($0.35 \geq |r_g| \geq 0.17$) on the basis of the genetic correlations reported in LDHub:[22] triglyceride (TG), coronary artery disease (CAD), age at Smoking (Age_Smo) childhood IQ (cIQ), hemoglobin A1c (HbA1C), and waist-hip ratio (WHR). For simplicity, we assumed that 1,000 causal variants were shared by all seven traits. When we used the heritability estimates reported in LDHub[22] for our simulation, there was a phenomenon that the overall p value was driven by the trait with the largest h^2 if the sample sizes were set the same. For this reason, we adjusted sample sizes so that Nh^2 is constant for all traits. Then for the focal trait, we doubled the sample size.

Figure S8 shows the results of the power simulation. Again, PLEIO achieved the highest power. With sample sizes satisfying $Nh^2 = 10,000$, PLEIO achieved a power of 72.6%, while the second-best method (MTAG-U) achieved 52.8% and the third-best method (ASSET) achieved 37.3%. We note that the interpretation is different for MTAG-F than other methods because other methods are not trait specific. That is, in other methods, a careful interpretation is required before concluding that the association is driven by the focal trait.

Computation time and memory usage comparison

We compared the computation time and maximum memory usage of the methods. We assumed the simulation settings in the focal-trait power simulation ($T = 7$). As previously noted, we used our own implementations of MTAG and METAL. For importance sampling, we used $N_{sample} = 100K$. We generated a simulation input for performing $10K$ and $1M$ associative tests, and tested each method using one CPU. Table S9 shows that PLEIO, MTAG and METAL can perform $1M$ associative tests in an hour with less than $4GB$ of free memory, in this setting.

Joint analysis of multiple traits related to cardiovascular disease

We applied PLEIO to identify pleiotropic loci associated with traits related to cardiovascular disease (CVD). We

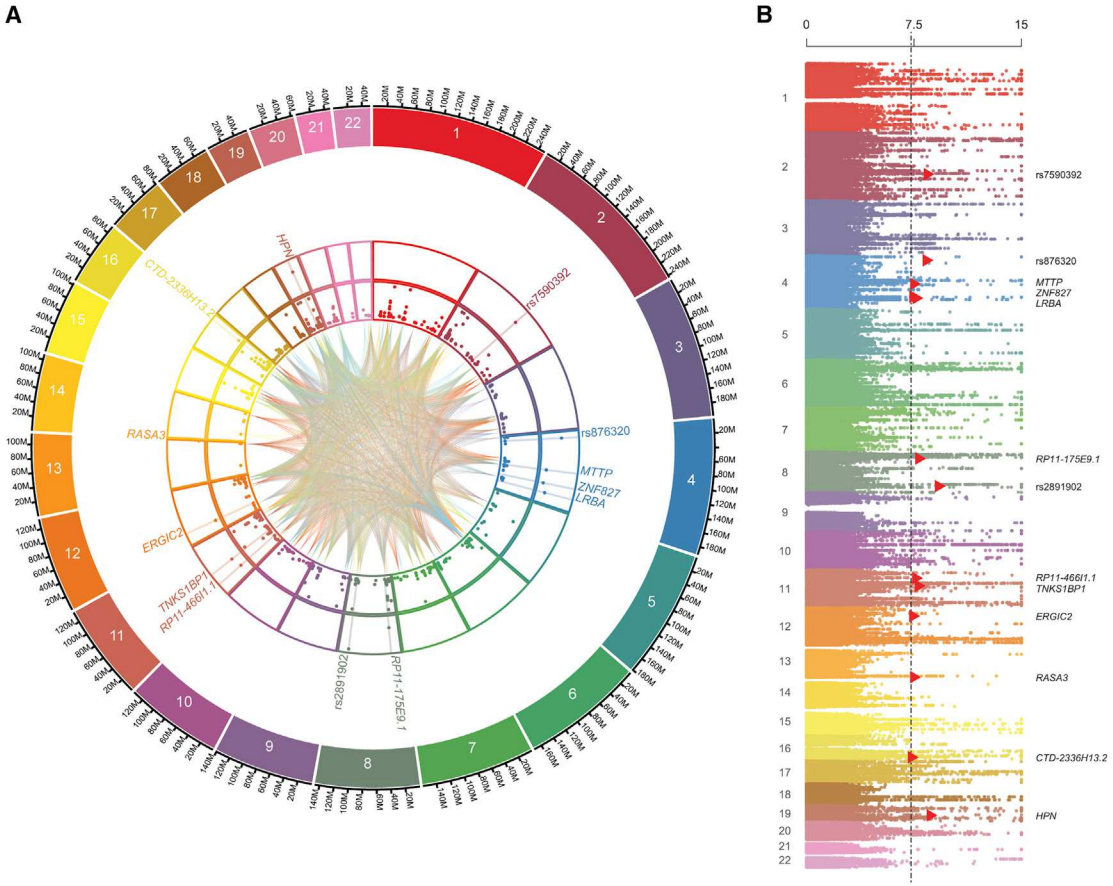

Figure 2. The summary of the PLEIO analysis result with GWAS summary statistics of the 18 CVD-related traits
(A) The circular plot shows the locations and the statistical significances of the 13 novel variants (outer edge) and the 625 GWAS top SNPs (inner edge). The inner ribbons connect the variants in the same functional category found by the DAVID analysis.
(B) The Manhattan plot of the PLEIO association results. Red triangles indicate the 13 novel loci.

collected summary statistics of 18 diseases and complex traits from multiple consortia (Table S1). We selected 12 binary traits from the Neale lab's UK Biobank GWAS results (Table S2) by using the following search terms: heart, hypertension, obesity, lipoproteins, cholesterol, and diabetes. We collected four lipid traits from the Global Lipid consortium,[16] one binary trait (CAD) from the CARDIoGRAM+C4D consortium,[17] and one trait (FG) from MAGIC.[18] In total, we collected 13 binary and five quantitative traits. See Material and methods for details of the trait selection. Quantitative traits had differing units. Lipid traits had the unit of mg/dL, whereas the FG had the unit of mmol/L.[16,18] We used the intersection of $1,777,411$ imputed SNPs across all datasets. These traits showed differing heritabilities and non-zero genetic and environmental correlations (Figure S9).

PLEIO identified 625 independent GWAS top hits that exceeded the threshold $p = 5\times10^{-8}$ (Figure 2 and Table S3). Among those, we found 13 independent novel variants, which have no known associations to CVD traits and were not significant in each single study (Table S4). The local Manhattan plots of these loci are shown in Figure S10. Figure 2A shows a circular plot whose radial position indicates the genomic position, and the heights of the points are the statistical significances of the variants. The genome-wide Manhattan plot is shown in Figure 2B. We compared the results of PLEIO to the original summary statistics by using a mirrored Manhattan plot in Figure S11.

We used LDSC to investigate whether our statistics had systematic inflation. To apply LDSC, we should assume that the chi-square statistic for a SNP in LD decreases by r^2. Although it is unclear whether this assumption is correct in the PLEIO statistics, we have accepted this assumption and applied LDSC. The LDSC intercept was close to one ($\alpha = 1.11$), which showed that our results did not have much inflation.

We used the Variant Effect Predictor (VEP v.97.2) in ENSEMBL GRCh37 and obtained the annotations of the identified variants. The 13 novel variants included six intronic variants, three non-coding transcript variants, three intergenic variants, and one upstream gene variant. The 625 top hits included 374 intronic variants, 112 intergenic variants, 41 upstream gene variants, 25 downstream variants, 23 missense variants, 21 3′ UTR variants, 12 non-coding transcript exon variants, 12 synonymous variants, and five 5′ UTR variants. The detailed annotations are in Tables S10 and S11.

Using the 625 top hits, we performed an additional analysis with DAVID v.6.8. Given the list of genes obtained by

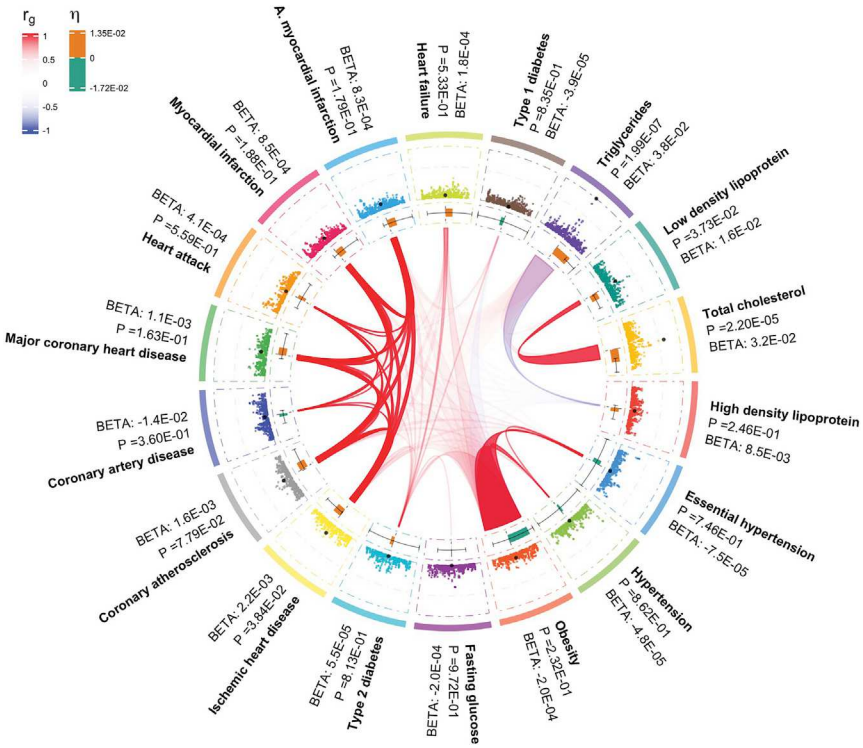

Figure 3. Pleiotropy plot of rs1688030, an intronic variant of HPN

The radial axis of the circular plot is divided by the 18 traits included in the real data analysis. The outermost layer shows the p values and the effect size estimates of the variant obtained from the original GWAS summary statistics. The next layer shows the local Manhattan plots of the variant within 1 Mb window. The horizontal bar plot shows the direction and maginitude of the standardized effect size (η) with the 95% confidence interval for each trait. The inner ribbons show the genetic correlations (as the color: positive r_g as red and negative r_g as blue) and the relative SNP heritability per trait (as the width of the ribbon end). The upper left corner shows the color scale used in the inner ribbon plots (left) and the range of observed standardized effect sizes (right)

VEP, we used DAVID to search for the presence of known trait-gene associations based on the Genetic Association Database (GAD, Table S12). We curated the reported trait-gene associations into eight categories: CAD, FG, hypertension, diabetes, high density lipoprotein (HDL), LDL, total cholesterol, and total glycerides. That is, we categorized the variants into eight groups on the basis of the trait category of the known association. We visualized the results in the inner circle of Figure 3A, where each ribbon indicates a pair of genes in the same phenotypic category.

For comparison, we applied MTAG to the same dataset. Because MTAG gave 18 p values per SNP, we first considered looking at all $18 \times 1,777,411$ p values. For the purpose of discovering the associated locus, this is equivalent to looking at the minimum p value per each SNP without considering multiple testing (MTAG-U). MTAG-U identified 622 independent GWAS hits (Figure S12). Thus, MTAG-U found a slightly fewer number of associations than PLEIO (625 hits) in this analysis. Although MTAG-U found a comparable number of hits, we note that the number of p values MTAG-U examined was much larger than PLEIO. When we applied LDSC, MTAG-U showed an inflated intercept ($\alpha = 3.89$) as expected because MTAG-U is a minimum p value approach (Table S13). Next, we considered a scenario that we want to correct for multiple testing. After applying the Bonferroni correction, MTAG-C identified 493 GWAS hits. Another possible approach to correct for multiple testing would be to adjust the χ^2 statistic so that the LDSC intercept would be similar to PLEIO ($\alpha = 1.10$). However, this approach further reduced the number of GWAS hits to 102 (Table S14), suggesting that the inflation caused by multiple testing is not well corrected by the intercept adjustment. For a detailed description of this analysis, see Material and methods.

We measured the computation time needed for this real data analysis by using a single CPU core. Running LDSC for 18 traits took 0.2 h and running pairwise LDSC for $\binom{18}{2}$ pairs took 1.5 h. Building the importance sampling distribution (with $N_{sample} = 1M$) for PLEIO took 1.89 h. Then, testing $1,777,411$ SNPs with PLEIO took 1.83 h. In total, PLEIO spent 3.72 h excluding LDSC preprocessing and required $2.1GB$ memory at peak.

Interpretation of the joint analysis results

To further interpret the multi-trait associations at each locus we identified, we visualized the result of each locus by using a circular plot, which we call "pleiotropy plot." The pleiotropy plot includes the local Manhattan plot and the bar plot of the standardized effect sizes. The inner ribbons show the genetic correlations as colors and the explained heritabilities by the locus as widths. We drew pleiotropy plots of the 13 novel variants we identified (Figure 3 and Figure S5). On the basis of the patterns observed in these plots, we manually categorized the 13 variants into four non-overlapping groups, which may imply distinct underlying pathways (Figure 4).

The first group of variants had associations with seven binary traits: six traits from the UK Biobank (acute myocardial infarction, myocardial infarction, heart attack, major coronary heart disease, coronary atherosclerosis, and ischemic heart disease) and one trait (CAD) from CARDIoGRAM+C4D. These seven traits showed high genetic correlations (Figure 4). We categorized variants into this group if the variant had the strongest association with one of the seven traits and associations ($p < 0.001$) with at least three traits out of the seven traits. The variants showing

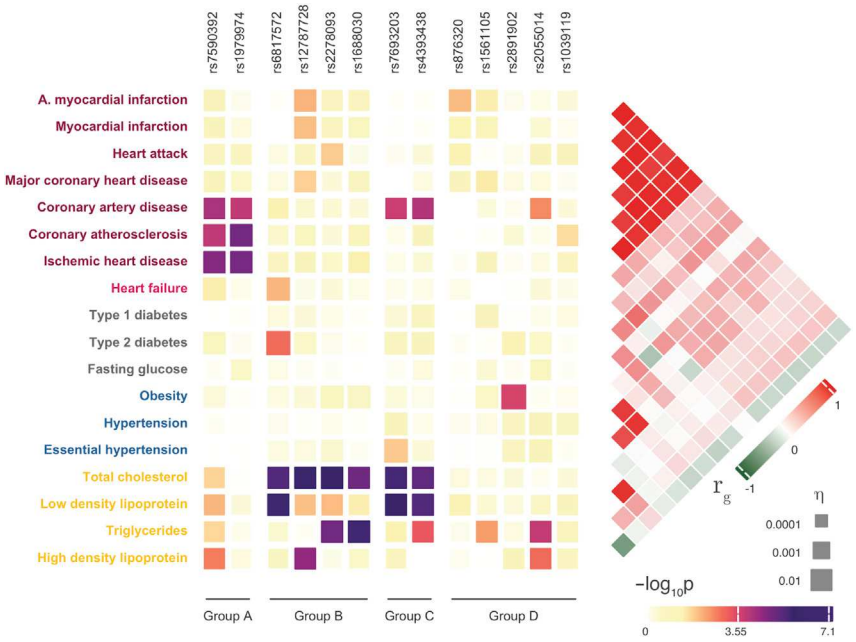

Figure 4. Distinct association patterns of 13 novel variants identified by PLEIO
Each box represents the association of a variant with a trait; the size of the box indicates the magnitude of the standardized effect size (η), and the color of the box indicates the statistical significance. The right-side heatmap shows the genetic correlations. We divided the variants into four groups on the basis of their association patterns. In the lower right corner, we provide the color scale of the genetic correlations, the size scale of the effect sizes, and the color scale of the associations.

this pattern were rs7590392 near ACVR2A (2q22.3) and rs1979974 in ZNF827 (4q31.22).

The second group of variants had associations with lipid phenotypes (triglycerides, LDL, HDL, and total cholesterol). We categorized variants into this group if the variant had the strongest association with one of the lipid traits and associations ($p < 0.001$) with at least two lipid traits. The variants showing this pattern were rs6817572 in LRBA (6p22.3), rs12787728 in TNKS1BP1 (11q12.1), rs2278093 in ERGIC2 (12p11.22), and rs1688030 in HPN (19q13.12).

These variants showed differing associations to the lipid phenotypes. rs6817572 showed the strongest associations to the total cholesterol and LDL. rs12787728 showed the strongest associations to the total cholesterol and HDL. rs2278093 and rs1688030 showed the strongest associations to the total cholesterol and triglycerides.

The third group of variants had associations with both the CAD and the lipid phenotypes. We categorized variants into this group if the variant had associations ($p < 0.001$) with both CAD and one of the lipid traits at the same time. Although these variants satisfied both the condition for group 1 and the condition for group 2, we categorized them separately as the third group. The variants showing this pattern were rs7693203 in MTTP (4q23) and rs4393438 in RASA3 (13q34). The variants in this group showed strong associations ($p < 0.0001$) to the total cholesterol and LDL.

The fourth group of variants was the variants that were not categorized into the three aforementioned groups. The variants in this group were rs876320 near FGFBP1 (4p15.32), rs1561105 in RP11-175E9.1 (8p21.2), rs2891902 near RPL35AP19 (8q24.12), rs2055014 in RP11-466I1.1 (8q24.12), and rs1039119 in AC106729.1 (16q23.1). rs2891902 showed the strongest association to obesity ($p < 0.001$) and weak associations to type 2 diabetes and hypertensions. rs876320, rs1561105, and rs1039119 were interesting because their associations to all traits were weak ($p > 0.01$). The strongest associations of rs1039119 were to coronary atherosclerosis ($p = 0.02$) and triglycerides ($p = 0.08$). However, this SNP's effect size directions to the seven binary traits in the first group were all concordant to the genetic correlations of these traits. The strongest associations of rs1561105 were to triglycerides ($p = 0.005$) and major coronary heart disease ($p = 0.03$), acute myocardial infarction ($p = 0.04$), and myocardial infarction ($p = 0.05$). This SNP's effect size directions to these three traits were all concordant to the genetic correlations. The strongest associations of rs876320 were to acute myocardial infarction ($p = 0.01$), myocardial infarction ($p = 0.04$), and heart attack ($p = 0.04$). This SNP's effect size directions to these three traits were all concordant to the genetic correlations. Thus, PLEIO seems to have captured the aggregate information in multiple weak associations by considering the fact that the effect size directions were concordant to the genetic correlations. Further follow-ups would be needed to determine whether these loci with weak associations to multiple traits present true associations or false positives.

Discussion

We have presented PLEIO, a framework to identify and interpret pleiotropic loci with GWAS summary statistics of multiple traits. PLEIO increased statistical power by using a test of variance components in a random effect model that models genetic correlations and heritabilities and by using standardized units of effect sizes across traits. Our method offers interpretation and visualization tools to help understand shared association patterns of pleiotropic loci.

PLEIO is a general method that includes other previous meta-analysis methods as special cases. If we set the genetic covariance matrix to a matrix of ones and the environmental correlations to zero, the test is approximately equivalent to the fixed effects meta-analysis method. If we assume environmental correlations, the test is

approximately equivalent to the Lin-Sullivan method.[21] If we set the genetic covariance matrix to an identity matrix and the environmental correlations to zero, the resulting test is similar to the heterogeneity test in the Han-Eskin random effects model.[23] If we set the genetic covariance matrix to an identity matrix and assume environmental correlations, the resulting test is similar to the heterogeneity test in the RE2C framework.[24] A difference of PLEIO is that, unlike other methods optimized for specific scenarios, it estimates the genetic covariance and the environmental correlations from data and adjusts itself to each scenario. For example, if we have a collection of the studies for the same trait, PLEIO will learn this information and act as though it were a fixed effects meta-analysis method.

PLEIO can combine the traits from different populations. When we combine the same traits of the same population, the genetic correlations will be one. However, when we combine the same traits from multiple ethnicities, the genetic correlation is usually positive but imperfect ($0 < r_g < 1$). Recent methods can estimate genetic correlations across different populations by accounting for population-specific LD.[25,26] One can use these methods to estimate r_g for the PLEIO analysis if the traits come from multiple populations.

In a multi-trait analysis, one must decide which traits should be included. Selection of traits can be performed on the basis of the literature describing comorbidity, shared candidate genes, or observed genetic correlations. If one includes a trait with no pleiotropy to other traits, the power to detect pleiotropic loci shared across all traits will decrease. In real data analysis, our trait selection was based on literature search, and the observed r_g between selected traits was greater than 0.15. One approach to choose traits can be based on an estimated r_g with the whole genome. However, a possible pitfall is that the region-specific pleiotropic effects can be ignored because there can be specific regions whose local co-heritabilities are greater than other regions.[11]

There exist two types of multi-trait analyses. The first is a joint meta-analysis in which the statistics of several traits are combined into one. The goal of this type of analysis is to find pleiotropic loci that are associated to multiple traits. These analyses have the same strengths and weaknesses as a typical meta-analysis. Aggregating more traits can provide additional power, but modeling heterogeneity between traits and interpreting results can often be challenging. The second type is a trait-specific analysis in which related traits help the association test of a specific trait.[6,7,27,28] The goal of this type of analysis is to maximize power for the analysis of each trait. In this study, we focused on the meta-analysis methods. Because our framework provides tools to facilitate interpretations, our method can minimize the weaknesses of the joint meta-analysis.

PLEIO has similarities and differences to a popular approach, MTAG.[6] Both methods model the genetic correlations, heritabilities, and environmental correlations. Both methods can deal with binary traits and quantitative traits with different units. The main difference is that PLEIO is a meta-analysis approach, while MTAG is a trait-specific method. Given T traits, PLEIO produces one p value per locus, while MTAG produces T trait-specific p values. Therefore, if one wants to calculate a single association p value per locus, PLEIO can be the method of choice. One advantage of MTAG is that the polygenic risk prediction can be made more accurate with the updated trait-specific effect sizes. In contrast, PLEIO is an aggregate meta-analysis method that does not update trait-specific effect sizes. Thus, for risk prediction, MTAG can be the method of choice. In the future, it will be interesting to expand the PLEIO framework to update effect sizes via techniques such as the best linear unbiased predictor (BLUP).

The pleiotropic loci identified by PLEIO can be attributed to biological or mediated pleiotropy.[29] In the former case, the variant has an independent association for each trait tested. In the latter case, however, the variant may have cross-trait associations resulting from causal relationship of two or more traits' being tested. PLEIO does not have the ability to distinguish these two types of pleiotropy and will identify loci with any of them. It will be an interesting research direction to examine the effect of the type of pleiotropy to PLEIO's power and to develop methods to distinguish between the two via incorporation of Mendelian randomization into the PLEIO framework.

We developed PLEIO under the assumption that only GWAS summary statistics are available. If individual-level genotype data are available, multivariate regression approaches can be used to combine information from multiple traits.[12,30,31] These methods can utilize individual-level information and control for confounding factors consistently across traits. However, to run these methods, sample data of all traits must be available at one location. Considering that the transfers of genotype data are becoming increasingly difficult because of privacy issues,[32,33] collecting all samples would be challenging. Moreover, models using individual genotypes commonly require large computing resources. As for the statistical power, Lin and Zeng[34] have shown that the use of individual-level data did not much improve statistical power over the use of summary statistics in the context of traditional meta-analysis. In multi-trait analysis, it would be interesting to compare power between the two types of methods in the future studies.

In summary, we proposed a general and flexible meta-analysis framework for the identification and interpretation of pleiotropic loci. We expect that our framework can help discover core genes that contribute to multiple phenotypes, which can lead us to a better understanding of the common etiology of traits and the development of shared drug targets.

Data and code availability

PLEIO is publicly available at https://github.com/cuelee/pleio. The summary statistics data used for the multi-trait association analysis are available from UK biobank GWAS results, the Global Lipids Genetics consortium, the CARDIo+C4D consortium, and MAGIC. The multi-trait association results are available upon request.

Supplemental Data

Supplemental Data can be found online at https://doi.org/10.1016/j.ajhg.2020.11.017.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF, 2019R1A2C2002608) funded by the Korean government, Ministry of Science, and ICT. This work was supported by the Creative-Pioneering Researchers Program funded by Seoul National University (SNU). C.L. was supported by the Graduate Student Scholarship by Asan Foundation and the Lecture and Research Scholarship by Seoul National University College of Medicine.

Declaration of interests

B.H. is the CTO of Genealogy Inc.

Received: July 22, 2020
Accepted: November 23, 2020
Published: December 21, 2020

Web resources

CARDIo+C4D consortium, http://www.cardiogramplusc4d.org
Global Lipids Genetics consortium, http://lipidgenetics.org
GWAS catalog, https://www.ebi.ac.uk/gwas/home
LDSC, https://github.com/bulik/ldsc
MAGIC, https://www.magicinvestigators.org
MTAG, https://github.com/JonJala/mtag
PLEIO, https://github.com/cuelee/pleio
UK biobank GWAS results, http://www.nealelab.is/uk-biobank

References

1. Gratten, J., and Visscher, P.M. (2016). Genetic pleiotropy in complex traits and diseases: implications for genomic medicine. Genome Med. *8*, 78.
2. Watanabe, K., Stringer, S., Frei, O., Umićević Mirkov, M., de Leeuw, C., Polderman, T.J.C., van der Sluis, S., Andreassen, O.A., Neale, B.M., and Posthuma, D. (2019). A global overview of pleiotropy and genetic architecture in complex traits. Nat. Genet. *51*, 1339–1348.
3. Bhattacharjee, S., Rajaraman, P., Jacobs, K.B., Wheeler, W.A., Melin, B.S., Hartge, P., Yeager, M., Chung, C.C., Chanock, S.J., Chatterjee, N.; and GliomaScan Consortium (2012). A subset-based approach improves power and interpretation for the combined analysis of genetic association studies of heterogeneous traits. Am. J. Hum. Genet. *90*, 821–835.
4. Han, B., and Eskin, E. (2012). Interpreting meta-analyses of genome-wide association studies. PLoS Genet. *8*, e1002555.
5. Kang, E.Y., Park, Y., Li, X., Segrè, A.V., Han, B., and Eskin, E. (2016). ForestPMPlot: A Flexible Tool for Visualizing Heterogeneity Between Studies in Meta-analysis. G3 (Bethesda) *6*, 1793–1798.
6. Turley, P., Walters, R.K., Maghzian, O., Okbay, A., Lee, J.J., Fontana, M.A., Nguyen-Viet, T.A., Wedow, R., Zacher, M., Furlotte, N.A., et al.; 23andMe Research Team; and Social Science Genetic Association Consortium (2018). Multi-trait analysis of genome-wide association summary statistics using MTAG. Nat. Genet. *50*, 229–237.
7. Liley, J., and Wallace, C. (2015). A pleiotropy-informed Bayesian false discovery rate adapted to a shared control design finds new disease associations from GWAS summary statistics. PLoS Genet. *11*, e1004926.
8. Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.R., Duncan, L., Perry, J.R., Patterson, N., Robinson, E.B., et al.; ReproGen Consortium; Psychiatric Genomics Consortium; and Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3 (2015). An atlas of genetic correlations across human diseases and traits. Nat. Genet. *47*, 1236–1241.
9. Lee, S.H., Goddard, M.E., Wray, N.R., and Visscher, P.M. (2012). A better coefficient of determination for genetic profile analysis. Genet. Epidemiol. *36*, 214–224.
10. Bulik-Sullivan, B.K., Loh, P.R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., Neale, B.M.; and Schizophrenia Working Group of the Psychiatric Genomics Consortium (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. *47*, 291–295.
11. Ni, G., Moser, G., Wray, N.R., Lee, S.H.; and Schizophrenia Working Group of the Psychiatric Genomics Consortium (2018). Estimation of Genetic Correlation via Linkage Disequilibrium Score Regression and Genomic Restricted Maximum Likelihood. Am. J. Hum. Genet. *102*, 1185–1194.
12. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. *88*, 76–82.
13. Kang, H.M., Sul, J.H., Service, S.K., Zaitlen, N.A., Kong, S.Y., Freimer, N.B., Sabatti, C., and Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. Nat. Genet. *42*, 348–354.
14. Self, S.G., and Liang, K.-Y. (1987). Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions. J. Am. Stat. Assoc. *82*, 605–610.
15. Owen, A., and Zhou, Y. (2000). Safe and Effective Importance Sampling. J. Am. Stat. Assoc. *95*, 135–143.
16. Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M.L., Mora, S., et al.; Global Lipids Genetics Consortium (2013). Discovery and refinement of loci associated with lipid levels. Nat. Genet. *45*, 1274–1283.
17. Nikpay, M., Goel, A., Won, H.H., Hall, L.M., Willenborg, C., Kanoni, S., Saleheen, D., Kyriakou, T., Nelson, C.P., Hopewell, J.C., et al. (2015). A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat. Genet. *47*, 1121–1130.
18. Dupuis, J., Langenberg, C., Prokopenko, I., Saxena, R., Soranzo, N., Jackson, A.U., Wheeler, E., Glazer, N.L., Bouatia-

Naji, N., Gloyn, A.L., et al.; DIAGRAM Consortium; GIANT Consortium; Global BPgen Consortium; Anders Hamsten on behalf of Procardis Consortium; and MAGIC investigators (2010). New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat. Genet. *42*, 105–116.

19. Sudmant, P.H., Rausch, T., Gardner, E.J., Handsaker, R.E., Abyzov, A., Huddleston, J., Zhang, Y., Ye, K., Jun, G., Fritz, M.H., et al.; 1000 Genomes Project Consortium (2015). An integrated map of structural variation in 2,504 human genomes. Nature *526*, 75–81.
20. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics *26*, 2190–2191.
21. Lin, D.Y., and Sullivan, P.F. (2009). Meta-analysis of genome-wide association studies with overlapping subjects. Am. J. Hum. Genet. *85*, 862–872.
22. Zheng, J., Erzurumluoglu, A.M., Elsworth, B.L., Kemp, J.P., Howe, L., Haycock, P.C., Hemani, G., Tansey, K., Laurin, C., Pourcain, B.S., et al.; Early Genetics and Lifecourse Epidemiology (EAGLE) Eczema Consortium (2017). LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics *33*, 272–279.
23. Han, B., and Eskin, E. (2011). Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. Am. J. Hum. Genet. *88*, 586–598.
24. Lee, C.H., Eskin, E., and Han, B. (2017). Increasing the power of meta-analysis of genome-wide association studies to detect heterogeneous effects. Bioinformatics *33*, i379–i388.
25. Brown, B.C., Ye, C.J., Price, A.L., Zaitlen, N.; and Asian Genetic Epidemiology Network Type 2 Diabetes Consortium (2016). Transethnic Genetic-Correlation Estimates from Summary Statistics. Am. J. Hum. Genet. *99*, 76–88.
26. Galinsky, K.J., Reshef, Y.A., Finucane, H.K., Loh, P.R., Zaitlen, N., Patterson, N.J., Brown, B.C., and Price, A.L. (2019). Estimating cross-population genetic correlations of causal effect sizes. Genet. Epidemiol. *43*, 180–188.
27. Andreassen, O.A., Thompson, W.K., Schork, A.J., Ripke, S., Mattingsdal, M., Kelsoe, J.R., Kendler, K.S., O'Donovan, M.C., Rujescu, D., Werge, T., et al.; Psychiatric Genomics Consortium (PGC); and Bipolar Disorder and Schizophrenia Working Groups (2013). Improved detection of common variants associated with schizophrenia and bipolar disorder using pleiotropy-informed conditional false discovery rate. PLoS Genet. *9*, e1003455.
28. Chung, D., Yang, C., Li, C., Gelernter, J., and Zhao, H. (2014). GPA: a statistical approach to prioritizing GWAS results by integrating pleiotropy and annotation. PLoS Genet. *10*, e1004787.
29. Solovieff, N., Cotsapas, C., Lee, P.H., Purcell, S.M., and Smoller, J.W. (2013). Pleiotropy in complex traits: challenges and strategies. Nat. Rev. Genet. *14*, 483–495.
30. Korte, A., Vilhjálmsson, B.J., Segura, V., Platt, A., Long, Q., and Nordborg, M. (2012). A mixed-model approach for genome-wide association studies of correlated traits in structured populations. Nat. Genet. *44*, 1066–1071.
31. Zhou, X., and Stephens, M. (2014). Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nat. Methods *11*, 407–409.
32. Erlich, Y., and Narayanan, A. (2014). Routes for breaching and protecting genetic privacy. Nat. Rev. Genet. *15*, 409–421.
33. Kim, K., Baik, H., Jang, C.S., Roh, J.K., Eskin, E., and Han, B. (2019). Genomic GPS: using genetic distance from individuals to public data for genomic analysis without disclosing personal genomes. Genome Biol. *20*, 175.
34. Lin, D.Y., and Zeng, D. (2010). On the relative efficiency of using summary statistics versus individual-level data in meta-analysis. Biometrika *97*, 321–332.