Supplementary Information for "Multi-trait analysis of genome-wide association summary statistics using MTAG"

Table of Contents

Supplementary Note

1 Theory

1.1 Background

The derivation of MTAG assumes that each SNP j's effect is drawn from a distribution with a common variance-covariance matrix $\mathbf{\Omega}$. We refer to this assumption as the "homogeneous-$\mathbf{\Omega}$ assumption."

MTAG is derived in two steps. In step 1, we show that the MTAG estimate of SNP j on trait t depends on (i) the vector of single-trait GWAS estimates of the effect of SNP j on each of the traits included in the analysis, (ii) the variance-covariance matrix of estimation errors, $\mathbf{\Sigma}_j$, and (iii) the variance-covariance matrix of SNP effects, $\mathbf{\Omega}$. In practice, the matrices $\mathbf{\Sigma}_j$ and $\mathbf{\Omega}$ are usually not known. In step 2, we therefore show how they can be estimated to yield an operational (feasible) version of the MTAG estimator.

Next, we derive analytic formulas that characterize several of MTAG's theoretical properties. We establish theoretically that for *any* joint genetic architecture across the traits—regardless of whether or not the homogeneous-$\mathbf{\Omega}$ assumption holds—MTAG estimates have lower genome-wide mean-squared error than estimates from the corresponding single-trait GWAS. Then, for individual-SNP associations identified by MTAG, we analytically characterize MTAG's power and false discovery rate, and we calculate them in some illustrative examples. Finally, we describe a procedure that in many practical applications can be used to calculate an informative upper bound on the false discovery rate under violations of the homogeneous-$\mathbf{\Omega}$ assumption.

1.2 Framework

1.2.1 Statistical Model

MTAG was developed for settings where an investigator has genome-wide association study (GWAS) summary statistics for T traits, possibly obtained from samples with unknown overlap. Denote the length-T vector of the true, unconditional effects of SNP j on each of the traits by $\boldsymbol{\beta}_j$. Note that, in contrast to much of the literature, we use the vector notation $\boldsymbol{\beta}_j$ to refer to the effect of the same SNP on multiple traits rather than the effect of multiple SNPs on the same trait. By "true," we mean that it is the effect that would be estimated in a sample of infinite size that is free from bias, and by "unconditional" we mean that it is the marginal effect of SNP j, not controlling for other SNPs. We normalize $\boldsymbol{\beta}_j$ so that it is measured in standardized units; that is, it is the vector of effects of SNP j after the genotype of the SNP and all the traits have been standardized to have mean zero and variance one.

Denote the corresponding vector of GWAS-estimated effects by $\hat{\boldsymbol{\beta}}_j$ and the vector of z-statistics by $\boldsymbol{Z}_j$. We assume that the $\hat{\boldsymbol{\beta}}_j$'s were calculated using some estimator that is consistent and asymptotically normal (as is true for the standard GWAS estimators, such as ordinary least squares and logistic regression). Because the genotypes are standardized, $\hat{\boldsymbol{\beta}}_j$ is the estimated effect of a one-standard-deviation change in genotype. We can therefore express $\boldsymbol{Z}_j$ as

$$\boldsymbol{Z}_j = \mathbf{W}_j \hat{\boldsymbol{\beta}}_j, \tag{1}$$

where $\mathbf{W}_j$ is a diagonal matrix whose k^{th} diagonal entry is a "weight" equal to the square root of sample size used to estimate the k^{th} element of $\hat{\boldsymbol{\beta}}_j$. By allowing $\mathbf{W}_j$ to vary across j, we account for the possibility that the available sample size differs across SNPs. Note that since $\boldsymbol{W}_j$ is symmetric, $\boldsymbol{W}_j = \boldsymbol{W}_j'$. Therefore, throughout this derivation, we omit the transpose on $\boldsymbol{W}_j$ when it would otherwise be necessary.

Denote the vector of GWAS estimation errors by $\boldsymbol{\varepsilon}_j \equiv \hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j$. It represents factors that cause $\hat{\boldsymbol{\beta}}_j$ to differ from $\boldsymbol{\beta}_j$, the combination of sampling variation and bias from sources such as stratification, cryptic relatedness, or technical artifacts. We assume that this estimation error has a distribution with

$$\begin{aligned} \mathbb{E}\left(\boldsymbol{\varepsilon}_j\right) &= 0 \\ \text{Var}\left(\boldsymbol{\varepsilon}_j\right) &= \boldsymbol{\Sigma}_j, \end{aligned}$$

for some variance-covariance matrix $\boldsymbol{\Sigma}_j$. The matrix $\boldsymbol{\Sigma}_j$ may vary across SNPs due to differences in sample size and sample overlap. In our empirical implementation of MTAG, we allow $\boldsymbol{\Sigma}_j$ to vary depending on SNP j's sample size, but we make the simplifying assumption that the amount of sample overlap across any two traits is the same for all SNPs. To be more precise, letting $N_{t,j}$ and $N_{s,j}$ denote the sample size used to estimate the effect of SNP j on trait t and trait s, respectively, and letting $N_{ts,j}$ denote the overlapping sample size, we assume that $N_{ts,j}/\sqrt{N_{t,j}N_{s,j}}$ is constant for all SNPs j.

The off-diagonal elements of $\boldsymbol{\Sigma}_j$ will be nonzero if (a) the phenotypic correlations are nonzero and there is sample overlap across traits, or if (b) biases in the SNP effect estimates (e.g., population stratification or cryptic relatedness) have correlated effects across traits. Accounting for the possibility of non-zero off-diagonal elements allows us to apply this method to GWAS summary statistics that are estimated using overlapping or related samples.

We use a random effects framework, i.e., we assume that the effect sizes, $\boldsymbol{\beta}_j$, are independently and identically distributed across j. Because the genotypes are standardized (and therefore have equal variance across SNPs), the assumption of identically distributed random effects implies that the expected amount of phenotypic variance explained is equal for each SNP, regardless of SNP characteristics such as allele frequency.[1] We assume that that each trait has non-zero heritability and that each SNP's effect is mean zero. The zero-mean assumption is justified because the choice of reference allele, and therefore the sign of the effect, is arbitrary. Our key substantive assumption is that $\boldsymbol{\beta}_j$ has a variance-covariance matrix across traits, denoted $\boldsymbol{\Omega}$, that is the same for all j. We do *not* make any assumptions about the shape of the distribution of true effect sizes $\boldsymbol{\beta}_j$ (such as normal, t-distribution, etc.). Nonetheless, the homogeneous-$\boldsymbol{\Omega}$ assumption is strong, and an extended discussion of its implications is found in section 1.2.5.

[1] While that is a strong implicit assumption about each trait's genetic architecture, it is an assumption that is common to many methods in statistical genetics [1, 2, 3]. If we instead assumed that the effect sizes of the *unstandardized* genotypes were identically distributed across SNPs j, then the expected contribution to heritability of rare variants would be less than the expected contribution of common variants.

1.2.2 Generalized Method of Moments

MTAG is a generalized method of moments estimator [4, 5] defined by a set of *moment conditions* and a weight matrix. For each observation $i = 1, 2, ..., N$, denote the vector of data we have by $\boldsymbol{D}_i$. Each $\boldsymbol{D}_i$ is treated as a random variable generated by an assumed statistical model that has an unknown vector-valued parameter $\boldsymbol{\theta}_0$. Suppose there is a vector-valued function $\boldsymbol{g}(\boldsymbol{\theta}; \boldsymbol{D}_i)$ such that

$$\mathbb{E}\left[\boldsymbol{g}(\boldsymbol{\theta}_0; \boldsymbol{D}_i)\right] = \mathbf{0}.$$

This set of equations is referred to as the set of *population moment conditions*. The GMM estimator is the value of $\boldsymbol{\theta}$ that minimizes the objective function

$$\sum_i \boldsymbol{g}(\boldsymbol{\theta}_0; \boldsymbol{D}_i)' \boldsymbol{\Psi} \boldsymbol{g}(\boldsymbol{\theta}_0; \boldsymbol{D}_i)$$

for some positive semi-definite weight matrix $\boldsymbol{\Psi}$. Under some regularity conditions, including the assumption that

$$\boldsymbol{\Psi}\mathbb{E}\left[\boldsymbol{g}(\boldsymbol{\theta}; \boldsymbol{D}_i)\right] = \mathbf{0}$$

if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, GMM is consistent and asymptotically normal as $N \to \infty$. Additionally, if we choose to use the weight matrix

$$\boldsymbol{\Psi} = \left(\mathrm{Var}\left[\boldsymbol{g}(\boldsymbol{\theta}_0; \boldsymbol{D}_i)\right]\right)^{-1},$$

then GMM will be the most efficient estimator among all consistent and asymptotically normal estimators [4]. Below, we derive the moment conditions used by the MTAG estimator.

1.2.3 Derivation of Moment Conditions

The MTAG moment conditions are based on the projection of the GWAS-estimated effect of SNP j on some trait s onto the space spanned by a constant and the true (but unknown) effect of SNP j on trait t. More precisely, the projection is defined by the coefficients

$$\boldsymbol{\gamma} \equiv \begin{bmatrix} \gamma_0 \\ \gamma_1 \end{bmatrix}$$

that minimize the expected squared difference between the GWAS estimate $\hat{\beta}_{j,s}$ and a linear function of the true marginal effect coefficient $\beta_{j,t}$,

$$\mathbb{E}\left[\left(\hat{\beta}_{j,s} - \begin{bmatrix} 1 & \beta_{j,t} \end{bmatrix} \boldsymbol{\gamma}\right)^2\right].$$

To find $\boldsymbol{\gamma}$, we take the derivative of this expression with respect to $\boldsymbol{\gamma}$ and set it equal to zero. This gives us

$$\begin{aligned}
0 &= \frac{\partial}{\partial \boldsymbol{\gamma}} \mathbb{E}\left[\left(\hat{\beta}_{j,s} - \begin{bmatrix} 1 & \beta_{j,t} \end{bmatrix} \boldsymbol{\gamma}\right)^2\right] \\
&= \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\gamma}}\left(\hat{\beta}_{j,s} - \begin{bmatrix} 1 & \beta_{j,t} \end{bmatrix} \boldsymbol{\gamma}\right)^2\right] \\
&= 2\mathbb{E}\left[\begin{bmatrix} 1 \\ \beta_{j,t} \end{bmatrix}\left(\hat{\beta}_{j,s} - \begin{bmatrix} 1 & \beta_{j,t} \end{bmatrix} \boldsymbol{\gamma}\right)\right] \\
&= 2\mathbb{E}\left[\left(\begin{bmatrix} \hat{\beta}_{j,s} \\ \beta_{j,t}\hat{\beta}_{j,s} \end{bmatrix} - \begin{bmatrix} 1 & \beta_{j,t} \\ \beta_{j,t} & \beta_{j,t}^2 \end{bmatrix} \boldsymbol{\gamma}\right)\right] \\
0 &= \mathbb{E}\left[\left(\begin{bmatrix} \hat{\beta}_{j,s} \\ \beta_{j,t}\hat{\beta}_{j,s} \end{bmatrix} - \begin{bmatrix} 1 & \beta_{j,t} \\ \beta_{j,t} & \beta_{j,t}^2 \end{bmatrix} \boldsymbol{\gamma}\right)\right] \qquad (2) \\
&= \mathbb{E}\left(\begin{bmatrix} \beta_{j,s} \\ \beta_{j,t}\beta_{j,s} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_{j,s} \\ \beta_{j,t}\boldsymbol{\varepsilon}_{j,s} \end{bmatrix} - \begin{bmatrix} 1 & \beta_{j,t} \\ \beta_{j,t} & \beta_{j,t}^2 \end{bmatrix} \boldsymbol{\gamma}\right) \\
&= \begin{bmatrix} 0 \\ \omega_{ts} \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & \omega_{tt} \end{bmatrix} \boldsymbol{\gamma} \\
\begin{bmatrix} 1 & 0 \\ 0 & \omega_{tt} \end{bmatrix} \boldsymbol{\gamma} &= \begin{bmatrix} 0 \\ \omega_{ts} \end{bmatrix} \\
\boldsymbol{\gamma} &= \begin{bmatrix} 1 & 0 \\ 0 & \omega_{tt} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \omega_{ts} \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{\omega_{tt}} \end{bmatrix} \begin{bmatrix} 0 \\ \omega_{ts} \end{bmatrix} \\
&= \begin{bmatrix} 0 \\ \frac{\omega_{ts}}{\omega_{tt}} \end{bmatrix}. \qquad (3)
\end{aligned}$$

Therefore, we see that the projection of $\hat{\beta}_{j,s}$ onto the space spanned by a constant and $\beta_{j,t}$ is $\frac{\omega_{ts}}{\omega_{tt}}\beta_{j,t}$.

1.2.4 MTAG Estimator

Substituting (3) into the first element of (2) above, we have

$$\mathbb{E}\left(\hat{\beta}_{j,s} - \frac{\omega_{ts}}{\omega_{tt}}\beta_{j,t}\right) = 0. \qquad (4)$$

Since this is a function of the unknown parameter we wish to estimate ($\beta_{j,t}$) and the data ($\hat{\beta}_{j,s}$) that is zero in expectation, we will use it as our moment condition. In fact, there are T moment conditions of the form (4), one for each trait s. Defining $\boldsymbol{\omega}_t$ as the t^{th} column of $\boldsymbol{\Omega}$, our vector of moment conditions is

$$\mathbb{E}\left(\hat{\boldsymbol{\beta}}_j - \frac{\boldsymbol{\omega}_t}{\omega_{tt}}\beta_{j,t}\right) = \mathbf{0}. \qquad (5)$$

We would like to estimate $\beta_{j,t}$ given $\hat{\boldsymbol{\beta}}_j$, $\boldsymbol{\Omega}$, and $\boldsymbol{\Sigma}_j$. The GMM estimator of $\beta_{j,t}$ corresponding the moment conditions above is the value of $\beta_{j,t}$ that minimizes

$$Q\left(\beta_{j,t}\right) = \left(\hat{\boldsymbol{\beta}}_j - \frac{\boldsymbol{\omega}_t}{\omega_{tt}}\beta_{j,t}\right)' \boldsymbol{W}_Q \left(\hat{\boldsymbol{\beta}}_j - \frac{\boldsymbol{\omega}_t}{\omega_{tt}}\beta_{j,t}\right)$$

for some positive semi-definite weight matrix $\boldsymbol{W}_Q$. The efficient GMM estimator uses the weight matrix:

$$\begin{aligned}
\boldsymbol{W}_Q &= \left[\mathrm{Var}\left(\hat{\boldsymbol{\beta}}_j - \frac{\boldsymbol{\omega}_t}{\omega_{tt}}\beta_{j,t}\right)\right]^{-1} \\
&= \left[\mathrm{Var}\left(\hat{\boldsymbol{\beta}}_j\right) + \mathrm{Var}\left(\frac{\boldsymbol{\omega}_t}{\omega_{tt}}\beta_{j,t}\right) - 2\mathrm{Var}\left(\hat{\boldsymbol{\beta}}_j, \frac{\boldsymbol{\omega}_t}{\omega_{tt}}\beta_{j,t}\right)\right]^{-1} \\
&= \left[\mathrm{Var}\left(\hat{\boldsymbol{\beta}}_j\right) + \frac{\boldsymbol{\omega}_t}{\omega_{tt}}\mathrm{Var}\left(\beta_{j,t}\right)\frac{\boldsymbol{\omega}_t'}{\omega_{tt}} - 2\mathrm{Var}\left(\hat{\boldsymbol{\beta}}_j, \beta_{j,t}\right)\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\right]^{-1} \\
&= \left[\boldsymbol{\Omega} + \boldsymbol{\Sigma}_j + \frac{\boldsymbol{\omega}_t}{\omega_{tt}}\omega_{tt}\frac{\boldsymbol{\omega}_t'}{\omega_{tt}} - 2\boldsymbol{\omega}_t\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\right]^{-1} \\
&= \left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1}. \qquad (6)
\end{aligned}$$

So our efficient GMM objective function is

$$Q_{\text{eff}}\left(\beta_{j,t}\right) = \left(\hat{\boldsymbol{\beta}}_j - \frac{\boldsymbol{\omega}_t}{\omega_{tt}}\beta_{j,t}\right)' \left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1} \left(\hat{\boldsymbol{\beta}}_j - \frac{\boldsymbol{\omega}_t}{\omega_{tt}}\beta_{j,t}\right). \qquad (7)$$

To obtain the MTAG estimator of $\beta_{j,t}$, we minimize (7) with respect to $\beta_{j,t}$. The first-order condition of this minimization problem is

$$\begin{aligned}
0 &= \frac{\partial}{\partial\beta_{j,t}} Q_{\text{eff}}\left(\beta_{j,t}\right) \\
&= \frac{\partial}{\partial\beta_{j,t}}\left[\left(\hat{\boldsymbol{\beta}}_j - \frac{\boldsymbol{\omega}_t}{\omega_{tt}}\beta_{j,t}\right)' \left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1} \left(\hat{\boldsymbol{\beta}}_j - \frac{\boldsymbol{\omega}_t}{\omega_{tt}}\beta_{j,t}\right)\right] \\
&= \frac{\partial}{\partial\beta_{j,t}}\left[\hat{\boldsymbol{\beta}}_j' \left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1} \hat{\boldsymbol{\beta}}_j - 2\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1} \hat{\boldsymbol{\beta}}_j\beta_{j,t} + \frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1} \frac{\boldsymbol{\omega}_t}{\omega_{tt}}\beta_{j,t}^2\right] \\
&= -2\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1} \hat{\boldsymbol{\beta}}_j + 2\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1} \frac{\boldsymbol{\omega}_t}{\omega_{tt}}\beta_{j,t}.
\end{aligned}$$

Solving for $\beta_{j,t}$ gives

$$\hat{\beta}_{\text{MTAG},j,t} = \frac{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1} \hat{\boldsymbol{\beta}}_j}{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1} \frac{\boldsymbol{\omega}_t}{\omega_{tt}}}. \qquad (8)$$

MTAG estimates of the effect of each SNP on each trait by using this equation, substituting estimates of the variance-covariance matrices of the effect sizes, $\boldsymbol{\Omega}$, and of the estimation error, $\boldsymbol{\Sigma}_j$, in place of the matrices. In some special cases, we may be able to increase the precision and computational speed of MTAG by making particular assumptions about these matrices. In the **Online Methods**, we discussed the four special cases that are implemented in the MTAG software. Here we discuss two of the cases in more detail.

No sample overlap across traits. To be more precise, the assumption in this special case is that the estimation error is uncorrelated across phenotypes, in which case $\mathbf{\Sigma}_j$ is a diagonal matrix. This is equivalent to assuming not only no sample overlap across estimates but also that the biases in the GWAS estimates are uncorrelated. Under this assumption, it is only necessary to conduct univariate LD score regressions for each trait, which will speed estimation. We highlight, however, that it is difficult to assess the assumption that the biases in the GWAS estimates are uncorrelated. For this reason, when this special case of MTAG is applied, we recommend running LD score regression on the MTAG results, and then inflating the MTAG standard errors by the square root of the estimated intercept. Doing so will correct the MTAG estimates for any new bias introduced by possible correlation of biases in the GWAS estimates.

Perfect genetic correlation but different heritabilities. This special case applies when the "traits" are different measures of the same trait, with different amounts of measurement error. In this case, we set the correlation of the effect sizes to one, which implies

$$\omega_{uv} = \sqrt{\omega_{uu}\omega_{vv}},$$

where ω_{uv} is the $(u,v)^{\text{th}}$ element of $\mathbf{\Omega}$. In this case, the $(u,v)^{\text{th}}$ element of $\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}$ is

$$\begin{aligned}\left[\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}\right]_{uv} &= \frac{\omega_{tu}\omega_{tv}}{\omega_{tt}} \\ &= \frac{\sqrt{\omega_{tt}\omega_{uu}}\sqrt{\omega_{tt}\omega_{vv}}}{\omega_{tt}} \\ &= \sqrt{\omega_{uu}\omega_{vv}} \\ &= \omega_{uv}.\end{aligned}$$

It follows that

$$\mathbf{\Omega} = \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}, \tag{9}$$

and therefore

$$\hat{\beta}_{\text{MTAG},j,t} = \frac{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\mathbf{\Sigma}_j^{-1}\hat{\boldsymbol{\beta}}_j}{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\mathbf{\Sigma}_j^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}}.$$

1.2.5 MTAG Standard Errors

The estimator of the MTAG standard error for SNP j's effect on trait t is:

$$\text{SE}\left(\hat{\beta}_{\text{MTAG},j,t}\right) = \frac{1}{\sqrt{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\mathbf{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \mathbf{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}}} \tag{10}$$

We will show that the standard error calculated using equation (10) has three properties: (i) it is the asymptotic standard error of the MTAG estimator, (ii) under the homogeneous-$\mathbf{\Omega}$ assumption, it is also the exactly correct small-sample standard error of the estimator, and (iii) when the homogeneous-$\mathbf{\Omega}$ assumption does not hold, using (10) will lead to standard errors that are conservative (i.e., too large) on average. We establish each of these properties in turn.

Asymptotic Standard Error. Applying a standard result about GMM estimators, the efficient MTAG esti-

mator has asymptotic distribution

$$\hat{\beta}_{\mathrm{MTAG},j,t} \,|\, \beta_{j,t} \overset{a}{\sim} N\left(\beta_{j,t},\, \left[\frac{\partial g'(b)}{\partial b} W_Q^{-1} \frac{\partial g(b)}{\partial b}\right]^{-1}\right) \tag{11}$$

where

$$g(b) = \left(\hat{\boldsymbol{\beta}}_j - \frac{\boldsymbol{\omega}_t}{\omega_{tt}} b\right) \tag{12}$$

is the moment condition used in MTAG. Substituting (6) and (12) into (11), the asymptotic sampling variance is

$$\left[\frac{\partial g'(b)}{\partial b} W_Q^{-1} \frac{\partial g(b)}{\partial b}\right]^{-1} = \frac{1}{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}} \left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t \boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1} \frac{\boldsymbol{\omega}_t}{\omega_{tt}}},$$

which yields the standard error in (10). In MTAG, however, the unit of observation is a trait, so these asymptotics may not be a good approximation if only a small number of traits is used.

Finite-Sample Standard Error. We will show that if the homogeneous-$\boldsymbol{\Omega}$ assumption holds, then equation (10) holds exactly. We will use the fact that under the homogeneous-$\boldsymbol{\Omega}$ assumption,

$$\mathrm{Var}\left(\boldsymbol{\beta}_j \,|\, \beta_{j,t}\right) = \boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t \boldsymbol{\omega}_t'}{\omega_{tt}}.$$

Given this, we calculate

$$\begin{aligned}
\mathrm{Var}\left(\hat{\beta}_{\mathrm{MTAG},j,t} \,|\, \beta_{j,t}\right) &= \mathrm{Var}\left(\frac{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}} \left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t \boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1} \hat{\boldsymbol{\beta}}_j}{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}} \left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t \boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1} \frac{\boldsymbol{\omega}_t}{\omega_{tt}}} \,|\, \beta_{j,t}\right) \\
&= \frac{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}} \left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t \boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1} \mathrm{Var}\left(\hat{\boldsymbol{\beta}}_j \,|\, \beta_{j,t}\right) \left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t \boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1} \frac{\boldsymbol{\omega}_t}{\omega_{tt}}}{\left[\frac{\boldsymbol{\omega}_t'}{\omega_{tt}} \left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t \boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1} \frac{\boldsymbol{\omega}_t}{\omega_{tt}}\right]^2} \\
&= \frac{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}} \left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t \boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1} \left[\mathrm{Var}\left(\boldsymbol{\beta}_j \,|\, \beta_{j,t}\right) + \mathrm{Var}\left(\boldsymbol{\epsilon}_j \,|\, \beta_{j,t}\right)\right] \left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t \boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1} \frac{\boldsymbol{\omega}_t}{\omega_{tt}}}{\left[\frac{\boldsymbol{\omega}_t'}{\omega_{tt}} \left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t \boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1} \frac{\boldsymbol{\omega}_t}{\omega_{tt}}\right]^2} \\
&= \frac{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}} \left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t \boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1} \left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t \boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right) \left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t \boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1} \frac{\boldsymbol{\omega}_t}{\omega_{tt}}}{\left[\frac{\boldsymbol{\omega}_t'}{\omega_{tt}} \left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t \boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1} \frac{\boldsymbol{\omega}_t}{\omega_{tt}}\right]^2} \\
&= \frac{1}{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}} \left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t \boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1} \frac{\boldsymbol{\omega}_t}{\omega_{tt}}}.
\end{aligned} \tag{13}$$

Equation (13) follows. We note that even if the homogeneous-$\boldsymbol{\Omega}$ assumption does not hold, the standard error formula is likely to provide an accurate approximation if all traits are highly genetically correlated, since in such cases $\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t \boldsymbol{\omega}_t'}{\omega_{tt}} \approx \mathbf{0}$.

Conservative (on average) when the homogeneous-$\boldsymbol{\Omega}$ assumption does not hold. In a slight abuse of notation, we let $\boldsymbol{\Omega}$ denote the genome-wide (i.e., across-SNP) variance-covariance matrix of effect sizes, and we calculate

the expectation of the sampling variance of the MTAG estimator:

$$
\begin{aligned}
\mathbb{E}\left[\mathrm{Var}\left(\hat{\beta}_{\mathrm{MTAG},j,t}-\beta_{j,t}\,|\,\beta_{j,t}\right)\right] &= \mathbb{E}\left[\mathbb{E}\left[\left(\hat{\beta}_{\mathrm{MTAG},j,t}-\beta_{j,t}\right)^2\,|\,\beta_{j,t}\right]-\mathbb{E}\left(\hat{\beta}_{\mathrm{MTAG},j,t}-\beta_{j,t}\,|\,\beta_{j,t}\right)^2\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\left(\hat{\beta}_{\mathrm{MTAG},j,t}-\beta_{j,t}\right)^2\,|\,\beta_{j,t}\right]\right]-\mathbb{E}\left[\mathbb{E}\left(\hat{\beta}_{\mathrm{MTAG},j,t}-\beta_{j,t}\,|\,\beta_{j,t}\right)^2\right] \\
&= \mathbb{E}\left[\left(\hat{\beta}_{\mathrm{MTAG},j,t}-\beta_{j,t}\right)^2\right]-\mathbb{E}\left[\mathbb{E}\left(\hat{\beta}_{\mathrm{MTAG},j,t}-\beta_{j,t}\,|\,\beta_{j,t}\right)^2\right] \\
&\leq \mathbb{E}\left[\left(\hat{\beta}_{\mathrm{MTAG},j,t}-\beta_{j,t}\right)^2\right] \\
&= \mathbb{E}\left(\hat{\beta}^2_{\mathrm{MTAG},j,t}\right)+\mathbb{E}\left(\beta^2_{j,t}\right)-2\mathbb{E}\left(\hat{\beta}_{\mathrm{MTAG},j,t}\beta_{j,t}\right) \\
&= \frac{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}\mathbb{E}\left(\hat{\boldsymbol{\beta}}_j^2\right)\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}}{\left[\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}\right]^2} \\
&\quad +\mathbb{E}\left(\beta^2_{j,t}\right)-2\frac{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}\mathbb{E}\left(\hat{\boldsymbol{\beta}}_j\beta_{j,t}\right)}{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}} \\
&= \frac{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}\left(\boldsymbol{\Omega}+\boldsymbol{\Sigma}_j\right)\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}}{\left[\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}\right]^2} \\
&\quad +\omega_{tt}-2\frac{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}\boldsymbol{\omega}_t}{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}\boldsymbol{\omega}_t}\omega_{tt} \\
&= \frac{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}\left(\boldsymbol{\Omega}+\boldsymbol{\Sigma}_j\right)\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}}{\left[\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}\right]^2}-\omega_{tt} \\
&= \frac{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}\left(\boldsymbol{\Omega}+\boldsymbol{\Sigma}_j\right)\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}}{\left[\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}\right]^2} \\
&\quad -\frac{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}\omega_{tt}\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}}{\left[\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}\right]^2} \\
&= \frac{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}\left(\boldsymbol{\Omega}+\boldsymbol{\Sigma}_j\right)\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}}{\left[\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}\right]^2} \\
&\quad -\frac{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}}{\left[\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}\right]^2}
\end{aligned}
$$

$$= \frac{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1}\left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)\left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}}{\left[\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}\right]^2}$$

$$= \frac{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}}{\left[\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}\right]^2}$$

$$= \frac{1}{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}}.$$

Since this is the same sampling variance formula as above, we see that in expectation (10) will be larger than the true standard errors when the homogeneous-$\boldsymbol{\Omega}$ assumption does not hold. This reduces MTAG's power to detect true SNP associations.

1.3 Estimation of Model Parameters

The derivation of MTAG above assumed the matrices $\boldsymbol{\Sigma}_j$ and $\boldsymbol{\Omega}$ were known. In practice, they are not known, and it is necessary to first estimate the matrices and plug these estimates into equation 8 to obtain feasible versions of equations (8) and (10). In brief, we use LD score regression [1, 6] to estimate the elements of $\boldsymbol{\Sigma}_j$, and we estimate $\boldsymbol{\Omega}$ using the method of moments. We now describe these procedures.

1.3.1 Estimating $\boldsymbol{\Sigma}_j$

The matrix $\boldsymbol{\Sigma}_j$ is the variance-covariance matrix of the estimation error of the SNPs' effects, $\boldsymbol{\varepsilon}_j \equiv \hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j$, which includes both sampling variation and biases (e.g., stratification, cryptic relatedness, or technical artifacts). To construct an estimate for $\boldsymbol{\Sigma}_j$, we exploit the fact that the intercept from LD score regression estimates how much of the variance (or covariance) in the z-statistics is due to such biases [1].

More precisely, for some SNP j and for phenotypes t and s (where t may be equal to s), LD score regression partitions the product of z-statistics as

$$\begin{aligned}\mathbb{E}\left(Z_{j,t}Z_{j,s}\right) &= W_{j,t,t}W_{j,s,s}\mathbb{E}\left(\beta_{j,t}\beta_{j,s}\right) + W_{j,t,t}W_{j,s,s}\mathbb{E}\left(\varepsilon_{j,t}\varepsilon_{j,s}\right)\\ &= W_{j,t,t}W_{j,s,s}\mathbb{E}\left(\beta_{j,t}\beta_{j,s}\right) + W_{j,t,t}W_{j,s,s}\Sigma_{j,t,s}.\end{aligned}$$

Note that $\mathbb{E}\left(\beta_{j,t}\beta_{j,s}\right)$ is proportional to the LD score of SNP j, which can be estimated in a reference sample. Thus, the intercept from a regression of the product of SNP j's z-statistics for traits t and s on SNP j's LD score is an estimator of $W_{j,t,t}W_{j,s,s}\Sigma_{j,t,s}$.

For MTAG, we run univariate LD score regression for each trait (i.e., we estimate the above equation with $t = s$ for each trait) and bivariate LD score regression for each pair of traits. We then construct a matrix, $\boldsymbol{\Sigma}_{LD}$, whose $(t,s)^{\text{th}}$ element is the intercept from the LD score regression for traits t and s. The LD score regression equation above implies that

$$\boldsymbol{\Sigma}_{LD} = \boldsymbol{W}_j\boldsymbol{\Sigma}_j\boldsymbol{W}_j.$$

Since $W_{j,t,t}$ and $W_{j,s,s}$ are known, we back out an estimate of each $\Sigma_{j,t,s}$ by calculating

$$\mathbf{\Sigma}_j = \mathbf{W}_j^{-1}\mathbf{\Sigma}_{LD}\mathbf{W}_j^{-1}. \tag{14}$$

We use this formula to calculate an estimate, $\hat{\mathbf{\Sigma}}_j$, for each SNP. Under our procedure, the estimates $\hat{\mathbf{\Sigma}}_j$ can only differ across SNPs because of differences in sample size; that is, the relative contribution of sampling variation and bias is assumed to be the same across SNPs.

Because we are using the intercepts from LD score regressions to estimate $\mathbf{\Sigma}_j$, MTAG relies on many of the assumptions underlying LD score regression. For example, if the reference sample and estimation sample are not drawn from the same population, then the LD score regression intercept will be biased. Note, however, that some assumptions mainly matter for the slope of LD score regression rather than the intercept, making MTAG robust to violations of those assumptions. For instance, if the distribution of a SNP's effect size is related to the LD score of that SNP, the heritability estimate from LD score regression may be biased, but the intercept will be largely unaffected [7].

1.3.2 Estimating $\mathbf{\Omega}$

Using method of moments, we will estimate each of the elements of $\mathbf{\Omega}$ independently. Suppose we would like to estimate the $(t,s)^{\text{th}}$ element of $\mathbf{\Omega}$, which we denote by $\omega_{t,s}$. Note that for some SNP j,

$$\begin{aligned}
\mathbb{E}\left(\hat{\beta}_{j,t}\hat{\beta}_{j,s}\right) &= \mathbb{E}\left[\left(\beta_{j,t}+e_{j,t}\right)\left(\beta_{j,s}+e_{j,s}\right)\right] \\
&= \mathbb{E}\left(\beta_{j,t}\beta_{j,s}+\beta_{j,s}e_{j,t}+\beta_{j,t}e_{j,s}+e_{j,t}e_{j,s}\right) \\
&= \mathbb{E}\left(\beta_{j,t}\beta_{j,s}\right)+\mathbb{E}\left(\beta_{j,s}e_{j,t}\right)+\mathbb{E}\left(\beta_{j,t}e_{j,s}\right)+\mathbb{E}\left(e_{j,t}e_{j,s}\right) \\
&= \omega_{t,s}+\Sigma_{j,t,s},
\end{aligned}$$

where $\Sigma_{j,t,s}$ is the $(t,s)^{\text{th}}$ element of $\mathbf{\Sigma}_j$. We can therefore use the moment condition

$$\mathbb{E}\left(\hat{\beta}_{j,t}\hat{\beta}_{j,s}-\omega_{t,s}-\Sigma_{j,t,s}\right)=0.$$

This implies that the objective function for the method of moments estimator for $\omega_{t,s}$ is

$$Q_{t,s}=\sum_{j=1}^{M}\left(\hat{\beta}_{j,t}\hat{\beta}_{j,s}-\omega_{t,s}-\Sigma_{j,t,s}\right)^2,$$

where M is the number of SNPs in our data.[2] Minimizing this function with respect to $\omega_{t,s}$ gives us the first-order condition

$$\begin{aligned}
0 &= \frac{\partial Q_{t,s}}{\partial \omega_{t,s}} \\
&= \frac{\partial}{\partial \omega_{t,s}} \sum_{j=1}^{M} \left(\hat{\beta}_{j,t}\hat{\beta}_{j,s} - \omega_{t,s} - \Sigma_{j,t,s} \right)^2 \\
&= \sum_{j=1}^{M} \frac{\partial}{\partial \omega_{t,s}} \left(\hat{\beta}_{j,t}\hat{\beta}_{j,s} - \omega_{t,s} - \Sigma_{j,t,s} \right)^2 \\
&= -2 \sum_{j=1}^{M} \left(\hat{\beta}_{j,t}\hat{\beta}_{j,s} - \omega_{t,s} - \Sigma_{j,t,s} \right), \\
0 &= \sum_{j=1}^{M} \left(\hat{\beta}_{j,t}\hat{\beta}_{j,s} - \Sigma_{j,t,s} \right) - \sum_{j=1}^{M} \omega_{t,s} \\
&= \sum_{j=1}^{M} \left(\hat{\beta}_{j,t}\hat{\beta}_{j,s} - \Sigma_{j,t,s} \right) - M\omega_{t,s}.
\end{aligned}$$

Solving for $\omega_{t,s}$ gives us the method of moments estimator:

$$\hat{\omega}_{t,s} = \frac{1}{M} \sum_{j=1}^{M} \left(\hat{\beta}_{j,t}\hat{\beta}_{j,s} - \Sigma_{j,t,s} \right),$$

which is consistent (and asymptotically normal). The estimator for each component can simply be extended to the whole matrix $\mathbf{\Omega}$ as

$$\hat{\mathbf{\Omega}} = \frac{1}{M} \sum_{j=1}^{M} \left(\hat{\boldsymbol{\beta}}_j \hat{\boldsymbol{\beta}}_j' - \mathbf{\Sigma}_j \right).$$

Note that, while $\mathbf{\Omega} - \frac{\boldsymbol{\omega}_t \boldsymbol{\omega}_t'}{\omega_{tt}} + \mathbf{\Sigma}_j$ is a variance-covariance matrix (which implies it is positive definite and therefore invertible), since each of the elements of $\hat{\mathbf{\Omega}}$ and $\hat{\mathbf{\Sigma}}_j$ is estimated separately, it is possible that the estimate of $\mathbf{\Omega} - \frac{\boldsymbol{\omega}_t \boldsymbol{\omega}_t'}{\omega_{tt}} + \mathbf{\Sigma}_j$ will be non-invertible and that the MTAG estimates will be undefined. Such an event is in general unlikely, however, for two reasons. First, it will be non-invertible if and only if at least one of its eigenvalues is exactly one (which is unlikely to occur unless there is user error, e.g., including the same set of GWAS results twice).[3] Second, $\hat{\mathbf{\Sigma}}_j$ will be positive definite except in very extreme circumstances. To understand why, first note that elements of $\mathbf{\Sigma}_{LD}$ tend to be fairly precisely estimated when the GWAS summary statistics are from moderately well-powered studies. For example, in our data, the standard errors

[2] In principle, it is possible to improve upon the efficiency of the estimator by using the efficient GMM estimator with M moment conditions defined by $\mathbb{E}\left(\hat{\beta}_{j,t}\hat{\beta}_{j,s} - \omega_{t,s} - \Sigma_{j,t,s} \right) = 0$ for each SNP j. Our method of moments estimator corresponds to GMM where the weight matrix is the identity matrix. Since $\boldsymbol{\beta}_j$ corresponds to the unconditional effect of SNP j on the vector of traits, our "observations"—which are SNPs in this case—will be correlated. Therefore the efficient GMM weight matrix is not the identity matrix. However, the efficient weight matrix is computationally intensive to calculate, whereas our method of moments estimator has a closed-form solution. For this reason—and because $\mathbf{\Omega}$ will typically be estimated precisely anyway—we simply use the method of moments estimator here.

[3] In the limit of an infinite sample, this can happen if one set of summary statistics is a linear combination of some other set of summary statistics included in the analysis. In that case, the term can be made invertible without a loss of power by dropping the redundant trait. Because we are using estimates of $\mathbf{\Omega}$ and $\mathbf{\Sigma}_j$, non-invertibility may also be observed even if the matrix would be invertible with the true values of $\mathbf{\Omega}$ and $\mathbf{\Sigma}_j$. In practice, this is extremely unlikely unless the same set of summary statistics have been passed into MTAG more than once. Because such a condition is so unlikely absent any user error, we have programmed the MTAG software to return an error under this condition.

on our estimates of the elements of $\mathbf{\Sigma}_{LD}$ are on the order of 0.01. This implies that the estimate of $\mathbf{\Sigma}_{LD}$ is unlikely to be non-positive definite unless $\mathbf{\Sigma}_{LD}$ is nearly non-positive definite to begin with. Next, note that if the GWAS included reasonably effective controls for bias (e.g., principal components), the diagonal elements of $\mathbf{\Sigma}_{LD}$ primarily capture the sampling variances of the z-statistics (which are equal to one), while the off-diagonal elements primarily capture the degree of sample overlap jointly with the phenotypic correlation (see [1]). These off-diagonal elements approach one as sample overlap and phenotypic correlations approach 100%, but as long as they are substantially smaller than 100%, the off-diagonal elements will be substantially smaller than one, and $\mathbf{\Sigma}_{LD}$ will be far from non-positive definite.

1.3.3 Consistency of MTAG

Aside from variation due to bias in the GWAS estimates that may persist even in a large sample, MTAG is a consistent estimator. Recall from section 1.2.2 that in general, a GMM estimator is consistent irrespective of the GMM weight matrix that is used. In MTAG, however, the unit of observation is a trait, so this standard consistency property only applies as $T \to \infty$.[4] Here, we show that the estimate $\hat{\beta}_{\text{MTAG},j,t}$ converges to the true parameter $\beta_{j,t}$ as the GWAS sample size for trait t goes to infinity. For this sample-size consistency, it is necessary that MTAG uses the efficient GMM weight matrix.

We first note that for all s, the $(t,s)^{\text{th}}$ element of $\frac{\boldsymbol{\omega}_t \boldsymbol{\omega}_t'}{\omega_{tt}}$ is

$$\begin{aligned}\left[\frac{\boldsymbol{\omega}_t \boldsymbol{\omega}_t'}{\omega_{tt}}\right]_{t,s} &= \frac{\omega_{ts}\omega_{tt}}{\omega_{tt}} \\ &= \omega_{ts}.\end{aligned}$$

Therefore the t^{th} row and t^{th} column of $\mathbf{\Omega} - \frac{\boldsymbol{\omega}_t \boldsymbol{\omega}_t'}{\omega_{tt}}$ contain only zeros. Without loss of generality, we reorder the traits such that the t^{th} trait is the first trait. Then we can can write $\left(\mathbf{\Omega} - \frac{\boldsymbol{\omega}_t \boldsymbol{\omega}_t'}{\omega_{tt}} + \mathbf{\Sigma}_j\right)$ as a block matrix

$$\begin{bmatrix} A & \boldsymbol{B}' \\ \boldsymbol{B} & \boldsymbol{C} \end{bmatrix},$$

where A is the $(t,t)^{\text{th}}$ element of $\mathbf{\Sigma}_j$ (a scalar), $\boldsymbol{B}$ is the t^{th} column of $\mathbf{\Sigma}_j$ omitting the t^{th} element, and $\boldsymbol{C}$ is some positive definite matrix. Similarly, $\frac{\boldsymbol{\omega}_t}{\omega_{tt}}$ can be partitioned as

$$\begin{bmatrix} 1 \\ \boldsymbol{Y} \end{bmatrix}$$

where $\boldsymbol{Y}$ is some length-$(T-1)$ vector.

Our assumption that bias in the GWAS estimate vanishes in a large sample implies that A and all elements of $\boldsymbol{B}$ go to zero. Since $\hat{\mathbf{\Omega}}$, $\hat{\mathbf{\Sigma}}_j$, and $\hat{\boldsymbol{\beta}}_j$ are consistent estimators of $\mathbf{\Omega}$, $\mathbf{\Sigma}_j$, and $\boldsymbol{\beta}_j$, respectively, the probability

[4] This result requires an additional assumption: the sequence of GWAS estimates is a weakly stationary, ergodic process. Intuitively, this means that the information added by each additional trait is large enough that as the number of traits goes to infinity, the total information content from the set of traits goes to infinity as well.

limit of the MTAG estimator is

$$\begin{aligned}
\operatorname*{plim}_{N_t\to\infty}\hat{\beta}_{\text{MTAG},\,j,t} &= \operatorname*{plim}_{N_t\to\infty}\frac{\frac{\hat{\boldsymbol{\omega}}_t'}{\hat{\omega}_{tt}}\left(\hat{\boldsymbol{\Omega}}-\frac{\hat{\boldsymbol{\omega}}_t\hat{\boldsymbol{\omega}}_t'}{\hat{\omega}_{tt}}+\hat{\boldsymbol{\Sigma}}_j\right)^{-1}\hat{\boldsymbol{\beta}}_j}{\frac{\hat{\boldsymbol{\omega}}_t'}{\hat{\omega}_{tt}}\left(\hat{\boldsymbol{\Omega}}-\frac{\hat{\boldsymbol{\omega}}_t\hat{\boldsymbol{\omega}}_t'}{\hat{\omega}_{tt}}+\hat{\boldsymbol{\Sigma}}_j\right)^{-1}\frac{\hat{\boldsymbol{\omega}}_t}{\hat{\omega}_{tt}}} \\
&= \lim_{N_t\to\infty}\frac{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}\boldsymbol{\beta}_j}{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}} \\
&= \lim_{A\to 0,\boldsymbol{B}\to\mathbf{0}}\frac{\begin{bmatrix}1\\ \boldsymbol{Y}\end{bmatrix}'\begin{bmatrix}A & \boldsymbol{B}'\\ \boldsymbol{B} & \boldsymbol{C}\end{bmatrix}^{-1}\boldsymbol{\beta}_j}{\begin{bmatrix}1\\ \boldsymbol{Y}\end{bmatrix}'\begin{bmatrix}A & \boldsymbol{B}'\\ \boldsymbol{B} & \boldsymbol{C}\end{bmatrix}^{-1}\begin{bmatrix}1\\ \boldsymbol{Y}\end{bmatrix}}.
\end{aligned}$$

By the properties of block matrices, we have

$$\begin{aligned}
\operatorname*{plim}_{N_t\to\infty}\hat{\beta}_{\text{MTAG},\,j,t} &= \lim_{A\to 0,\boldsymbol{B}\to\mathbf{0}}\frac{\begin{bmatrix}1\\ \boldsymbol{Y}\end{bmatrix}'\begin{bmatrix}A^{-1}+A^{-1}\boldsymbol{B}'\left(\boldsymbol{D}-\boldsymbol{B}A^{-1}\boldsymbol{B}'\right)^{-1}\boldsymbol{B}A^{-1} & -A^{-1}\boldsymbol{B}'\left(\boldsymbol{D}-\boldsymbol{B}A^{-1}\boldsymbol{B}'\right)^{-1}\\ -\left(\boldsymbol{D}-\boldsymbol{B}A^{-1}\boldsymbol{B}'\right)^{-1}\boldsymbol{B}A^{-1} & \left(\boldsymbol{D}-\boldsymbol{B}A^{-1}\boldsymbol{B}'\right)^{-1}\end{bmatrix}\boldsymbol{\beta}_j}{\begin{bmatrix}1\\ \boldsymbol{Y}\end{bmatrix}'\begin{bmatrix}A^{-1}+A^{-1}\boldsymbol{B}'\left(\boldsymbol{D}-\boldsymbol{B}A^{-1}\boldsymbol{B}'\right)^{-1}\boldsymbol{B}A^{-1} & -A^{-1}\boldsymbol{B}'\left(\boldsymbol{D}-\boldsymbol{B}A^{-1}\boldsymbol{B}'\right)^{-1}\\ -\left(\boldsymbol{D}-\boldsymbol{B}A^{-1}\boldsymbol{B}'\right)^{-1}\boldsymbol{B}A^{-1} & \left(\boldsymbol{D}-\boldsymbol{B}A^{-1}\boldsymbol{B}'\right)^{-1}\end{bmatrix}\begin{bmatrix}1\\ \boldsymbol{Y}\end{bmatrix}} \\
&= \lim_{A\to 0}\frac{\begin{bmatrix}1\\ \boldsymbol{Y}\end{bmatrix}'\begin{bmatrix}A^{-1} & 0\\ 0 & \boldsymbol{D}^{-1}\end{bmatrix}\boldsymbol{\beta}_j}{\begin{bmatrix}1\\ \boldsymbol{Y}\end{bmatrix}'\begin{bmatrix}A^{-1} & 0\\ 0 & \boldsymbol{D}^{-1}\end{bmatrix}\begin{bmatrix}1\\ \boldsymbol{Y}\end{bmatrix}} \\
&= \lim_{A\to 0}\frac{A^{-1}\beta_{j,t}+\boldsymbol{Y}'\boldsymbol{D}^{-1}\boldsymbol{\beta}_{j,-t}}{A^{-1}+\boldsymbol{Y}'\boldsymbol{D}^{-1}\boldsymbol{Y}},
\end{aligned}$$

where $\boldsymbol{\beta}_{j,-t}$ is the vector $\boldsymbol{\beta}_j$ omitting the t^{th} element. Multiplying the numerator and denominator by A, we get

$$\begin{aligned}
\operatorname*{plim}_{N_t\to\infty}\hat{\beta}_{\text{MTAG},\,j,t} &= \lim_{A\to 0}\frac{\beta_{j,t}+A\boldsymbol{Y}'\boldsymbol{D}^{-1}\boldsymbol{\beta}_{j,-t}}{1+A\boldsymbol{Y}'\boldsymbol{D}^{-1}\boldsymbol{Y}} \\
&= \beta_{j,t}.
\end{aligned}$$

Hence MTAG is a consistent estimator.

1.4 Theoretical Properties of MTAG

In this section, we derive analytic formulas for three possible measures of MTAG's performance: the mean squared error (MSE), the estimator's statistical power to detect a true single-SNP association, and its false discovery rate (FDR). We begin by deriving a closed-form solution for the MSE of MTAG that is valid regardless of whether the homogeneous-$\boldsymbol{\Omega}$ assumption holds, and we prove that the MTAG estimates always have (weakly) lower MSE than the corresponding single-trait GWAS estimates. For the special case of $T = 2$, we also illustrate how the MSE depends in intuitive ways on model parameters. Next, under the assumption

that the SNPs' effect-size distribution is a mixture of mean-zero multivariate normal distributions, we derive closed-form solutions for MTAG's statistical power and FDR.

1.4.1 Mean Squared Error (MSE)

The mean squared error (MSE) of an estimator $\tilde{\beta}_{j,t}$ is defined as

$$\text{MSE}\left(\tilde{\beta}_{j,t}\right) \equiv \mathbb{E}\left[\left(\tilde{\beta}_{j,t} - \beta_{j,t}\right)^2\right]. \tag{15}$$

In deriving the genome-wide mean squared error (MSE) of the MTAG estimator, we will *not* impose the homogeneous-$\mathbf{\Omega}$ assumption. In a slight abuse of notation, we let $\mathbf{\Omega}$ denote the genome-wide (i.e., across-SNP) variance-covariance matrix of effect sizes. No assumptions are required about this variance-covariance matrix for any particular SNP because the genome-wide MSE is an expectation across all the SNPs.

In this calculation, however, we will make the simplifying assumption that $\mathbf{\Sigma}_j$ is the same for all SNPs j. This will be true if within the GWAS for each trait, each SNP effect size is estimated in a sample of the same size. Our simplifying assumption is often approximately true in practice (especially because among its SNP filters, MTAG drops SNPs that are estimated in a sample that is smaller than 75% of the 90th percentile of the sample size distribution for each trait).

Given the definition of MSE above, MTAG's MSE is the same quantity that we calculated in the derivation of the MTAG standard error in section 1.2.4:

$$\text{MSE}\left(\hat{\beta}_{\text{MTAG},\,j,t}\right) = \frac{1}{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\mathbf{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \mathbf{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}}. \tag{16}$$

Below, we show that MTAG always leads to an MSE that is weakly smaller than the MSE of the GWAS coefficients. Because both the GWAS and MTAG estimators have mean zero across SNPs, this result implies that MTAG will always lead to more precise estimates on average across the genome than a standard GWAS, regardless of the joint genetic architecture of the traits included.

To compare the MSEs, we first need to calculate the MSE of the GWAS coefficients for trait t. The GWAS MSE is simply:

$$\begin{aligned}
\text{MSE}\left(\hat{\beta}_{j,t}\right) &= \mathbb{E}\left[\left(\hat{\beta}_{j,t} - \beta_{j,t}\right)^2\right] \\
&= \mathbb{E}\left[\left(\beta_{j,t} + e_{j,t} - \beta_{j,t}\right)^2\right] \\
&= \mathbb{E}\left(e_{j,t}^2\right) \\
&= \sigma_{j,tt},
\end{aligned}$$

where $\sigma_{j,tt}$ is the $(t,t)^{\text{th}}$ element of $\mathbf{\Sigma}_j$. Hence we need to show that:

$$\sigma_{j,tt} \geq \frac{1}{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\mathbf{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \mathbf{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}}. \tag{17}$$

We will prove that this inequality holds by induction. Without loss of generality, order the traits so that the trait of interest is trait 1. We will show that (17) holds when only one trait is used in MTAG (i.e., when $T = 1$). We will then show that if (17) holds for $T = T_0$, then it must hold for $T = T_0 + 1$.

We start by establishing some notation. $\mathbf{\Omega}^{(T)}$ denotes the variance-covariance matrix of the effects sizes for the first T traits. $\boldsymbol{\omega}_t^{(T)}$ is a column vector of the covariance of the effect size of trait t with the first T traits. (If $t \leq T$, then this will be the t^{th} column of $\mathbf{\Omega}^{(T)}$.) As a reminder, ω_{tt} is the variance of the effect sizes for trait t. We define $\mathbf{\Sigma}_j^{(T)}$, $\boldsymbol{\sigma}_{j,t}^{(T)}$, and $\sigma_{j,tt}$ in a parallel way. Lastly, we use $\hat{\beta}^{(T)}_{\text{MTAG},\,j,t}$ to denote the MTAG estimate of the effect of SNP j on phenotype t based on the GWAS estimates for the first T traits.

For the $T = 1$ case, the MTAG estimate is simply equal to the GWAS estimate and thus has the same MSE. Formally, $\mathbf{\Omega}^{(1)} = \boldsymbol{\omega}_1^{(1)} = \omega_{11}$ and $\mathbf{\Sigma}_j^{(1)} = \sigma_{j,11}$, and so

$$\begin{aligned}
\text{MSE}\left(\hat{\beta}^{(1)}_{\text{MTAG},\,j,t}\right) &= \frac{1}{\frac{\boldsymbol{\omega}_1^{(1)\prime}}{\omega_{11}}\left(\mathbf{\Omega}^{(1)} - \frac{\boldsymbol{\omega}_1^{(1)}\boldsymbol{\omega}_1^{(1)\prime}}{\omega_{11}} + \mathbf{\Sigma}_j^{(1)}\right)^{-1}\frac{\boldsymbol{\omega}_1^{(1)}}{\omega_{tt}}} \\
&= \frac{1}{\frac{\omega_{tt}}{\omega_{tt}}\left(\omega_{tt} - \frac{\omega_{tt}^2}{\omega_{tt}} + \sigma_{j,tt}\right)^{-1}\frac{\omega_{tt}}{\omega_{tt}}} \\
&= \frac{1}{\left(\sigma_{j,tt}\right)^{-1}} \\
&= \sigma_{j,tt} \\
&= \text{MSE}\left(\hat{\beta}_{\text{GWAS},\,j,t}\right).
\end{aligned}$$

Thus, (17) holds (with equality) when $T = 1$. Next, assume that (17) holds when $T = T_0$:

$$\frac{1}{\frac{\boldsymbol{\omega}_1^{(T_0)\prime}}{\omega_{11}}\left(\mathbf{\Omega}^{(T_0)} - \frac{\boldsymbol{\omega}_1^{(T_0)}\boldsymbol{\omega}_1^{(T_0)\prime}}{\omega_{11}} + \mathbf{\Sigma}^{(T_0)}\right)^{-1}\frac{\boldsymbol{\omega}_1^{(T_0)}}{\omega_{11}}} \quad \leq \quad \sigma_{j,tt}. \tag{18}$$

For the $T = T_0 + 1$ case, we have

$$\text{MSE}\left(\hat{\beta}^{(T_0+1)}_{\text{MTAG},\,j,t}\right) = \frac{1}{\frac{\boldsymbol{\omega}_1^{(T_0+1)\prime}}{\omega_{11}}\left(\mathbf{\Omega}^{(T_0+1)} - \frac{\boldsymbol{\omega}_1^{(T_0+1)}\boldsymbol{\omega}_1^{(T_0+1)\prime}}{\omega_{11}} + \mathbf{\Sigma}_j^{(T_0+1)}\right)^{-1}\frac{\boldsymbol{\omega}_1^{(T_0+1)}}{\omega_{tt}}}.$$

In block matrix notation, we have that:

$$\begin{aligned}
&\text{MSE}\left(\hat{\beta}^{(T_0+1)}_{\text{MTAG},\,j,t}\right) \\
= &\frac{1}{\begin{bmatrix} \frac{\boldsymbol{\omega}_1^{(T_0)\prime}}{\omega_{11}} & \frac{\omega_{1T_0+1}}{\omega_{11}} \end{bmatrix}\begin{bmatrix} \mathbf{\Omega}^{(T_0)} - \frac{\boldsymbol{\omega}_1^{(T_0)}\boldsymbol{\omega}_1^{(T_0)\prime}}{\omega_{11}} + \mathbf{\Sigma}^{(T_0)} & \boldsymbol{\omega}_{T_0+1}^{(T_0)} - \frac{\boldsymbol{\omega}_1^{(T_0)}\omega_{1T_0+1}}{\omega_{11}} + \boldsymbol{\sigma}_{T_0+1}^{(T_0)} \\ \boldsymbol{\omega}_{T_0+1}^{(T_0)\prime} - \frac{\omega_{1T_0+1}\boldsymbol{\omega}_1^{(T_0)\prime}}{\omega_{11}} + \boldsymbol{\sigma}_{T_0+1}^{(T_0)\prime} & \omega_{T_0+1,T_0+1} - \frac{\omega_{1T_0+1}^2}{\omega_{11}} + \sigma_{T_0+1,T_0+1} \end{bmatrix}^{-1}\begin{bmatrix} \frac{\boldsymbol{\omega}_1^{(T_0)}}{\omega_{11}} \\ \frac{\omega_{1T_0+1}}{\omega_{11}} \end{bmatrix}} \\
= &\frac{1}{\begin{bmatrix} \boldsymbol{X}' & y \end{bmatrix}\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}' & C \end{bmatrix}^{-1}\begin{bmatrix} \boldsymbol{X} \\ y \end{bmatrix}},
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{X} &\equiv \frac{\boldsymbol{\omega}_1^{(T_0)}}{\omega_{11}} \\
y &\equiv \frac{\omega_{1T_0+1}}{\omega_{11}} \\
\boldsymbol{A} &\equiv \boldsymbol{\Omega}^{(T_0)} - \frac{\boldsymbol{\omega}_1^{(T_0)}\boldsymbol{\omega}_1^{(T_0)\prime}}{\omega_{11}} + \boldsymbol{\Sigma}^{(T_0)} \\
\boldsymbol{B} &\equiv \boldsymbol{\omega}_{T_0+1}^{(T_0)} - \frac{\boldsymbol{\omega}_1^{(T_0)}\omega_{1T_0+1}}{\omega_{11}} + \boldsymbol{\sigma}_{T_0+1}^{(T_0)} \\
C &\equiv \omega_{T_0+1,T_0+1} - \frac{\omega_{1T_0+1}^2}{\omega_{11}} + \sigma_{T_0+1,T_0+1}.
\end{aligned}$$

By the properties of block matrices, we have

$$\begin{aligned}
&\mathrm{MSE}\left(\hat{\beta}_{\mathrm{MTAG},\,j,t}^{(T_0+1)}\right) \\
&= \left(\begin{bmatrix} \boldsymbol{X}' & y \end{bmatrix}\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}' & C \end{bmatrix}^{-1}\begin{bmatrix} \boldsymbol{X} \\ y \end{bmatrix}\right)^{-1} \\
&= \left(\begin{bmatrix} \boldsymbol{X}' & y \end{bmatrix}\begin{bmatrix} \boldsymbol{A}^{-1}+\boldsymbol{A}^{-1}\boldsymbol{B}\left(C-\boldsymbol{B}'\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1}\boldsymbol{B}'\boldsymbol{A}^{-1} & -\boldsymbol{A}^{-1}\boldsymbol{B}\left(C-\boldsymbol{B}'\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1} \\ -\left(C-\boldsymbol{B}'\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1}\boldsymbol{B}'\boldsymbol{A}^{-1} & \left(C-\boldsymbol{B}'\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1} \end{bmatrix}\begin{bmatrix} \boldsymbol{X} \\ y \end{bmatrix}\right)^{-1} \\
&= \Big(\boldsymbol{X}'\boldsymbol{A}^{-1}\boldsymbol{X} \\
&\quad + \begin{bmatrix} \boldsymbol{X}' & y \end{bmatrix}\begin{bmatrix} \boldsymbol{A}^{-1}\boldsymbol{B}\left(C-\boldsymbol{B}'\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1}\boldsymbol{B}'\boldsymbol{A}^{-1} & -\boldsymbol{A}^{-1}\boldsymbol{B}\left(C-\boldsymbol{B}'\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1} \\ -\left(C-\boldsymbol{B}'\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1}\boldsymbol{B}'\boldsymbol{A}^{-1} & \left(C-\boldsymbol{B}'\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1} \end{bmatrix}\begin{bmatrix} \boldsymbol{X} \\ y \end{bmatrix}\Big)^{-1}.
\end{aligned}$$

Since $\left(C-\boldsymbol{B}'\boldsymbol{A}^{-1}\boldsymbol{B}\right)$ is a scalar, we have

$$\begin{aligned}
&\mathrm{MSE}\left(\hat{\beta}_{\mathrm{MTAG},\,j,t}^{(T_0+1)}\right) \\
&= \left(\boldsymbol{X}'\boldsymbol{A}^{-1}\boldsymbol{X} + \left(C-\boldsymbol{B}'\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1}\begin{bmatrix} \boldsymbol{X}' & y \end{bmatrix}\begin{bmatrix} \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{B}'\boldsymbol{A}^{-1} & -\boldsymbol{A}^{-1}\boldsymbol{B} \\ -\boldsymbol{B}'\boldsymbol{A}^{-1} & 1 \end{bmatrix}\begin{bmatrix} \boldsymbol{X} \\ y \end{bmatrix}\right)^{-1} \\
&= \left(\boldsymbol{X}'\boldsymbol{A}^{-1}\boldsymbol{X} + \left(C-\boldsymbol{B}'\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1}\begin{bmatrix} \boldsymbol{X}' & y \end{bmatrix}\begin{bmatrix} \boldsymbol{A}^{-1}\boldsymbol{B} \\ -1 \end{bmatrix}\begin{bmatrix} \boldsymbol{B}'\boldsymbol{A}^{-1} & -1 \end{bmatrix}\begin{bmatrix} \boldsymbol{X} \\ y \end{bmatrix}\right)^{-1} \\
&= \left[\boldsymbol{X}'\boldsymbol{A}^{-1}\boldsymbol{X} + \left(C-\boldsymbol{B}'\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{A}^{-1}\boldsymbol{B}-y\right)^2\right]^{-1}.
\end{aligned}$$

Next, note that by (18),

$$\begin{aligned}
\boldsymbol{X}'\boldsymbol{A}^{-1}\boldsymbol{X} &= \frac{\boldsymbol{\omega}_1^{(T_0)\prime}}{\omega_{11}}\left(\boldsymbol{\Omega}^{(T_0)} - \frac{\boldsymbol{\omega}_1^{(T_0)}\boldsymbol{\omega}_1^{(T_0)\prime}}{\omega_{11}} + \boldsymbol{\Sigma}^{(T_0)}\right)^{-1}\frac{\boldsymbol{\omega}_1^{(T_0)}}{\omega_{11}} \\
&= \left(\left[\frac{\boldsymbol{\omega}_1^{(T_0)\prime}}{\omega_{11}}\left(\boldsymbol{\Omega}^{(T_0)} - \frac{\boldsymbol{\omega}_1^{(T_0)}\boldsymbol{\omega}_1^{(T_0)\prime}}{\omega_{11}} + \boldsymbol{\Sigma}^{(T_0)}\right)^{-1}\frac{\boldsymbol{\omega}_1^{(T_0)}}{\omega_{11}}\right]^{-1}\right)^{-1} \\
&= \left[\mathrm{MSE}\left(\hat{\beta}^{(T_0)}_{\mathrm{MTAG},\,j,t}\right)\right]^{-1} \\
&\geq \frac{1}{\sigma_{j,11}}.
\end{aligned}$$

Therefore,

$$\mathrm{MSE}\left(\hat{\beta}^{(T_0+1)}_{\mathrm{MTAG},\,j,t}\right) \leq \left(\frac{1}{\sigma_{j,11}} + \left(C - \boldsymbol{B}'\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{A}^{-1}\boldsymbol{B} - y\right)^2\right)^{-1}.$$

Furthermore, since

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}' & C \end{bmatrix}$$

is the variance-covariance matrix of $\left(\hat{\boldsymbol{\beta}}_j - \frac{\boldsymbol{\omega}_t}{\omega_{tt}}\beta_{j,t}\right)$ (see section 1.2.4), it must be positive definite. Therefore,

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}' & C \end{bmatrix}^{-1}$$

must be positive definite, and therefore all of the diagonal elements of

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}' & C \end{bmatrix}^{-1}$$

must be positive, including the bottom-right element, which we have shown to be $\left(C - \boldsymbol{B}'\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1}$. Therefore,

$$\left(C - \boldsymbol{B}'\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{A}^{-1}\boldsymbol{B} - y\right)^2 \geq 0,$$

implying that

$$\begin{aligned}
\mathrm{MSE}\left(\hat{\beta}^{(T_0+1)}_{\mathrm{MTAG},\,j,t}\right) &\leq \left(\frac{1}{\sigma_{j,11}} + \left(C - \boldsymbol{B}'\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{A}^{-1}\boldsymbol{B} - y\right)^2\right)^{-1} \\
&\leq \left(\frac{1}{\sigma_{j,11}}\right)^{-1} \\
&= \sigma_{j,11} \\
&= \mathrm{MSE}\left(\hat{\beta}_{\mathrm{GWAS},\,j,t}\right).
\end{aligned}$$

Since (17) holds for $T = 1$, and since (17) holds for $T = T_0 + 1$ whenever it holds for $T = T_0$, by induction (17) must hold for all natural numbers T. We have shown for any joint distribution of effect sizes, MTAG will produce estimates that, on average across SNPs, reduce the MSE relative to the MSE of the initial GWAS results.

1.4.2 Properties of MSE with Two Traits

In this section, to help build intuition about how the parameters $\mathbf{\Omega}$ and $\mathbf{\Sigma}_j$ interact, we prove some additional properties of the MSE in a bivariate MTAG analysis and then plot some illustrative calculations.

In the two-trait case, it is helpful for intuition to reparameterize $\mathbf{\Omega}$ and $\mathbf{\Sigma}_j$ (without loss of generality) as follows:

$$\mathbf{\Omega} \equiv \begin{bmatrix} 1 & r_\beta \\ r_\beta & 1 \end{bmatrix}$$

and

$$\mathbf{\Sigma}_j \equiv \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}.$$

We will prove two claims under this parameterization which we make in the main text of this paper. Namely, (a) *when $r_\beta = 1$, MSE is decreasing in the sample size of each trait*, and (b) *when the mean χ^2-statistic of the GWAS summary statistics for each trait is the same, MSE is decreasing in the absolute difference between r_β and $r_\varepsilon \equiv \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}$*.

To prove each of these, we first express the MSE in terms of the above parameterization. We have:

$$\begin{aligned}
\text{MSE} &= \frac{1}{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\mathbf{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \mathbf{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}} \\
&= \frac{1}{\begin{bmatrix} 1 & r_\beta \end{bmatrix}\left(\begin{bmatrix} 1 & r_\beta \\ r_\beta & 1 \end{bmatrix} - \begin{bmatrix} 1 & r_\beta \\ r_\beta & r_\beta^2 \end{bmatrix} + \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}\right)^{-1}\begin{bmatrix} 1 \\ r_\beta \end{bmatrix}} \\
&= \frac{1}{\begin{bmatrix} 1 & r_\beta \end{bmatrix}\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & 1 - r_\beta^2 + \sigma_{22} \end{bmatrix}^{-1}\begin{bmatrix} 1 \\ r_\beta \end{bmatrix}} \\
&= \frac{1}{\frac{1}{\sigma_{11}\left(1-r_\beta^2+\sigma_{22}\right)-\sigma_{12}^2}\begin{bmatrix} 1 & r_\beta \end{bmatrix}\begin{bmatrix} 1 - r_\beta^2 + \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix}\begin{bmatrix} 1 \\ r_\beta \end{bmatrix}} \\
&= \frac{1}{\frac{1}{\sigma_{11}\left(1-r_\beta^2+\sigma_{22}\right)-\sigma_{12}^2}\left(1 - r_\beta^2 + \sigma_{22} - 2r_\beta\sigma_{12} + r_\beta^2\sigma_{11}\right)} \\
&= \frac{\sigma_{11} - r_\beta^2\sigma_{11} + \sigma_{11}\sigma_{22} - \sigma_{12}^2}{1 - r_\beta^2 + \sigma_{22} - 2r_\beta\sigma_{12} + r_\beta^2\sigma_{11}}
\end{aligned}$$

Claim (a): When $r_\beta = 1$, MSE is decreasing in the sample size of each trait. First, consider the case where $r_\beta = 1$. Then

$$\text{MSE} = \frac{\sigma_{11}\sigma_{22} - \sigma_{12}^2}{\sigma_{22} - 2\sigma_{12} + \sigma_{11}}.$$

Increasing the sample size for Trait 1 corresponds to decreasing σ_{11}. (It also corresponds to increasing the mean χ^2-statistic, which is $\frac{1+\sigma_{11}}{\sigma_{11}}$). To evaluate the effect of increasing the sample size for Trait 1, we

differentiate

$$\begin{aligned}
\frac{\partial \text{MSE}}{\partial \sigma_{11}} &= \frac{\partial}{\partial \sigma_{11}}\left(\frac{\sigma_{11}\sigma_{22}-\sigma_{12}^2}{\sigma_{11}-2\sigma_{12}+\sigma_{22}}\right) \\
&= \frac{\left(\sigma_{11}-2\sigma_{12}+\sigma_{22}\right)\sigma_{22}-\left(\sigma_{11}\sigma_{22}-\sigma_{12}^2\right)}{\left(\sigma_{11}-2\sigma_{12}+\sigma_{22}\right)^2} \\
&= \frac{\sigma_{22}^2-2\sigma_{12}\sigma_{22}+\sigma_{12}^2}{\left(\sigma_{11}-2\sigma_{12}+\sigma_{22}\right)^2} \\
&= \frac{\left(\sigma_{22}-\sigma_{12}\right)^2}{\left(\sigma_{11}-2\sigma_{12}+\sigma_{22}\right)^2} \\
&> 0.
\end{aligned}$$

This positive derivative implies that as σ_{11} decreases (or rather as the sample size for the GWAS of Trait 1 increases), MSE decreases. By symmetry, the same will hold when we increase the sample size for Trait 2. Therefore, whenever we increase the sample size for either trait, MSE will decrease if $r_\beta = 1$.

Claim (b): When the mean χ^2-statistic of the GWAS summary statistics for each trait is the same, MSE is decreasing in the absolute difference between r_β and r_ε. Next, suppose the power of the GWASs for each trait is equal. That is, $\sigma_{11} = \sigma_{22}$. Then $\sigma_{12} = r_\varepsilon\sqrt{\sigma_{11}\sigma_{22}} = r_\varepsilon\sigma_{11}$, giving us

$$\begin{aligned}
\text{MSE} &= \frac{\sigma_{11}-r_\beta^2\sigma_{11}+\sigma_{11}^2-r_\varepsilon^2\sigma_{11}^2}{1-r_\beta^2+\sigma_{11}-2r_\beta r_\varepsilon\sigma_{11}+r_\beta^2\sigma_{11}} \\
&= \sigma_{11}\frac{1-r_\beta^2+\sigma_{11}\left(1-r_\varepsilon^2\right)}{1-r_\beta^2+\sigma_{11}\left(1-2r_\beta r_\varepsilon+r_\beta^2\right)}.
\end{aligned}$$

We will show that MSE is decreasing in $|r_\beta - r_\varepsilon|$ by first showing that MSE rises with r_β when $r_\beta < r_\varepsilon$ and falls with r_β when $r_\beta > r_\varepsilon$. We therefore differentiate

$$\begin{aligned}
\frac{\partial \text{MSE}}{\partial r_\beta} &= \frac{\partial}{\partial r_\beta}\left(\sigma_{11}\frac{1-r_\beta^2+\sigma_{11}\left(1-r_\varepsilon^2\right)}{1-r_\beta^2+\sigma_{11}\left(1-2r_\beta r_\varepsilon+r_\beta^2\right)}\right) \\
&= \sigma_{11}\frac{\left[1-r_\beta^2+\sigma_{11}\left(1-2r_\beta r_\varepsilon+r_\beta^2\right)\right](-2r_\beta)-\left[1-r_\beta^2+\sigma_{11}\left(1-r_\varepsilon^2\right)\right](-2r_\beta-2r_\varepsilon\sigma_{11}+2r_\beta\sigma_{11})}{\left[1-r_\beta^2+\sigma_{11}\left(1-2r_\beta r_\varepsilon+r_\beta^2\right)\right]^2} \\
&= 2\sigma_{11}\frac{\left(r_\varepsilon\sigma_{11}\right)r_\beta^2+\left(r_\varepsilon^2\sigma_{11}^2-r_\varepsilon^2\sigma_{11}-\sigma_{11}^2-\sigma_{11}\right)r_\beta+\left(-r_\varepsilon^3\sigma_{11}^2+r_\varepsilon\sigma_{11}^2+r_\varepsilon\sigma_{11}\right)}{\left[1-r_\beta^2+\sigma_{11}\left(1-2r_\beta r_\varepsilon+r_\beta^2\right)\right]^2}.
\end{aligned}$$

Note that the numerator is quadratic in r_β, with strictly positive expressions outside the fraction and in the denominator. Thus, the sign of this derivative is the same as the sign of

$$\left(r_\varepsilon\sigma_{11}\right)r_\beta^2+\left(r_\varepsilon^2\sigma_{11}^2-r_\varepsilon^2\sigma_{11}-\sigma_{11}^2-\sigma_{11}\right)r_\beta+\left(-r_\varepsilon^3\sigma_{11}^2+r_\varepsilon\sigma_{11}^2+r_\varepsilon\sigma_{11}\right).$$

We therefore use the quadratic equation to find the roots of this expression:

$$\begin{aligned} r_\beta &= \frac{-r_\varepsilon^2\sigma_{11}^2 + r_\varepsilon^2\sigma_{11} + \sigma_{11}^2 + \sigma_{11} \pm \sqrt{\left(r_\varepsilon^2\sigma_{11}^2 - r_\varepsilon^2\sigma_{11} - \sigma_{11}^2 - \sigma_{11}\right)^2 - 4r_\varepsilon\sigma_{11}\left(-r_\varepsilon^3\sigma_{11}^2 + r_\varepsilon\sigma_{11}^2 + r_\varepsilon\sigma_{11}\right)}}{2r_\varepsilon\sigma_{11}} \\ &= \frac{-r_\varepsilon^2\sigma_{11}^2 + r_\varepsilon^2\sigma_{11} + \sigma_{11}^2 + \sigma_{11} \pm \sqrt{\left(\sigma_{11}^2 + \sigma_{11} - r_\varepsilon^2\sigma_{11}^2 - r_\varepsilon^2\sigma_{11}\right)^2}}{2r_\varepsilon\sigma_{11}} \\ &= \frac{r_\varepsilon^2\sigma_{11}^2 - r_\varepsilon^2\sigma_{11} - \sigma_{11}^2 - \sigma_{11} \pm \left(\sigma_{11}^2 + \sigma_{11} - r_\varepsilon^2\sigma_{11}^2 - r_\varepsilon^2\sigma_{11}\right)}{-2r_\varepsilon\sigma_{11}} \end{aligned}$$

For the positive root, we have

$$\begin{aligned} r_\beta &= \frac{r_\varepsilon^2\sigma_{11}^2 - r_\varepsilon^2\sigma_{11} - \sigma_{11}^2 - \sigma_{11} + \left(\sigma_{11}^2 + \sigma_{11} - r_\varepsilon^2\sigma_{11}^2 - r_\varepsilon^2\sigma_{11}\right)}{-2r_\varepsilon\sigma_{11}} \\ &= \frac{2r_\varepsilon^2\sigma_{11}}{2r_\varepsilon\sigma_{11}} \\ &= r_\varepsilon. \end{aligned}$$

For the negative root, we have

$$\begin{aligned} r_\beta &= \frac{r_\varepsilon^2\sigma_{11}^2 - r_\varepsilon^2\sigma_{11} - \sigma_{11}^2 - \sigma_{11} - \left(\sigma_{11}^2 + \sigma_{11} - r_\varepsilon^2\sigma_{11}^2 - r_\varepsilon^2\sigma_{11}\right)}{-2r_\varepsilon\sigma_{11}} \\ &= \frac{\left(1 - r_\varepsilon^2\right)\sigma_{11} + 1}{r_\varepsilon}. \end{aligned}$$

This term will be smallest when $\sigma_{11} \to 0$, but note that in such a case, $|r_\beta| \geq 1$. This implies that the sign of the derivative changes only once for the feasible values of r_β. When $r_\beta = -1$, the sign of the derivative is

$$\begin{aligned} \text{sign}\left(\frac{\partial \text{MSE}}{\partial r_\beta}\right) &= \text{sign}\left[\left(r_\varepsilon\sigma_{11}\right)r_\beta^2 + \left(r_\varepsilon^2\sigma_{11}^2 - r_\varepsilon^2\sigma_{11} - \sigma_{11}^2 - \sigma_{11}\right)r_\beta + \left(-r_\varepsilon^3\sigma_{11}^2 + r_\varepsilon\sigma_{11}^2 + r_\varepsilon\sigma_{11}\right)\right] \\ &= \text{sign}\left(-r_\varepsilon^3\sigma_{11}^2 - r_\varepsilon^2\sigma_{11}^2 + r_\varepsilon^2\sigma_{11} + r_\varepsilon\sigma_{11}^2 + 2r_\varepsilon\sigma_{11} + \sigma_{11}^2 + \sigma_{11}\right) \\ &= \text{sign}\left(\sigma_{11}\left(1 - r^2\right)\left(r_\varepsilon + 1\right) + \left(r_\varepsilon + 1\right)^2\right), \end{aligned}$$

which is positive. Therefore, MSE rises with r_β when $r_\beta < r_\varepsilon$ and falls with r_β when $r_\beta > r_\varepsilon$. It follows that MSE is decreasing in $|r_\beta - r_\varepsilon|$.

We illustrate calculation of the MSE formula in **Supplementary Figure 1a**. Holding the traits' estimation-error correlation r_ε and Trait 1's expected χ^2-statistic fixed, the figure plots Trait 1's MSE as a function of the effect-size correlation r_β and the expected χ^2-statistic of Trait 2. For Trait 2's GWAS, we consider expected χ^2-statistics of 1.1 (low power), 1.4 (medium power), or 2.0 (high power). (For comparison, our GWAS of DEP ($h^2 = 0.064$) would have these expected χ^2-statistics if estimated in effective sample sizes of 81,189, 324,758, and 568,326 individuals, respectively.) Both of the claims we have proven are evident in the figure.

1.4.3 Power and False Discovery Rate (FDR)

There is no general formula for MTAG's statistical power or FDR that holds for all joint distributions of effect sizes across traits. We can derive formulas, however, when the effect-size distribution can be characterized

as a mixture of mean-zero normals. This class of distributions includes a large number of thick-tailed distributions, including non-infinitesimal models where some fraction of SNPs are null for a subset of traits. Specifically, we assume that the joint distribution of $\boldsymbol{\beta}_j$ is

$$\boldsymbol{\beta}_j \sim \begin{cases} N(\mathbf{0},\, \boldsymbol{\Omega}_1) & \text{with probability } p_1 \\ \vdots & \vdots \\ N(\mathbf{0},\, \boldsymbol{\Omega}_C) & \text{with probability } p_C, \end{cases}$$

where $\boldsymbol{\Omega}_c$ is the variance-covariance matrix for the c^{th} component (out of C) making up the mixture distribution and the p_c's are the mixture weights (which sum to one). Whenever there is more than one component in the mixture distribution, the homogeneous-$\boldsymbol{\Omega}$ assumption is violated.

The z-statistic of the MTAG estimator is

$$\begin{aligned} Z_{j,t} &= \frac{\hat{\boldsymbol{\beta}}_{\text{MTAG},j,t}}{\text{SE}\left(\hat{\boldsymbol{\beta}}_{\text{MTAG},j,t}\right)} \\ &= \frac{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1}\hat{\boldsymbol{\beta}}_j}{\sqrt{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}}}. \end{aligned}$$

Since the vector of GWAS estimates for a SNP that is drawn from the c^{th} component of the mixture distribution is distributed

$$\hat{\boldsymbol{\beta}}_j \,|\, c \sim N(\mathbf{0},\, \boldsymbol{\Omega}_c + \boldsymbol{\Sigma}_j),$$

the z-statistics for those SNPs will be normally distributed with mean zero and variance

$$\begin{aligned} \text{Var}\left(Z_{j,t} \,|\, c\right) &= \text{Var}\left(\frac{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1}\hat{\boldsymbol{\beta}}_j}{\sqrt{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}}} \,|\, c\right) \\ &= \frac{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1}\text{Var}\left(\hat{\boldsymbol{\beta}}_j \,|\, c\right)\left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}}{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}} \\ &= \frac{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1}(\boldsymbol{\Omega}_c + \boldsymbol{\Sigma}_j)\left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}}{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}}. \end{aligned} \tag{19}$$

To calculate power and FDR, first define the subset of components, D, in which $\beta_{j,t}$ follows a degenerate distribution with full mass on zero (i.e., SNP j is null for trait t), or equivalently the set of components where all elements of the t^{th} row and the t^{th} column of $\boldsymbol{\Omega}_c$ are zero. Power is defined as the probability of genome-wide significance conditional on the SNP effect size being drawn from a component not in D (i.e., components where SNP j is non-null for trait t). Therefore, letting z_0 denote the z-statistic that corresponds

to genome-wide significance in our data, we can calculate

$$\begin{aligned}\text{Power} &\equiv \mathbb{P}\left(Z_{j,t} > z_0 \,|\, \beta_{j,t} \neq 0\right) \\ &= \sum_{c \notin D} \mathbb{P}\left(Z_{j,t} > z_0 \,|\, c\right) \frac{p_c}{\sum_{c \notin D} p_c} \\ &= \frac{\sum_{c \notin D} \mathbb{P}\left(Z_{j,t} > z_0 \,|\, c\right) p_c}{\sum_{c \notin D} p_c}. \end{aligned} \tag{20}$$

Since we know the distribution of $(Z_{j,t} \,|\, c)$, we can directly calculate $\mathbb{P}\left(|Z_{j,t}| > z_0 \,|\, c\right)$ and therefore power.

To illustrate results from calculating the power formula, **Supplementary Figure 1b** plots power for the same parameter values as **Supplementary Figure 1a** for the special case of a multivariate normal distribution. We use this special case to illustrate the power formula for two reasons: it is fully parameterized by $\mathbf{\Omega}$, and it generates a lower bound for power relative to any mixture distribution. Not surprisingly, MTAG's power is greater precisely when its MSE is smaller.

FDR is defined as the probability that a SNP is null for the trait of interest conditional on the SNP being genome-wide significant. To calculate FDR, we first need to know that probability that any arbitrary SNP is genome-wide significant:

$$\begin{aligned} p_{\text{hit}} &\equiv \mathbb{P}\left(|Z_{j,t}| > z_0\right) \\ &= \sum_{c} \mathbb{P}\left(|Z_{j,t}| > z_0 \,|\, c\right) p_c. \end{aligned}$$

Using the calculation, the false discovery rate is defined as

$$\begin{aligned}\text{FDR} &\equiv \mathbb{P}\left(\text{null} \,|\, |Z_{j,t}| > z_0\right) \\ &= \frac{\mathbb{P}\left(|Z_{j,t}| > z_0 \,|\, \text{null}\right) \mathbb{P}\left(\text{null}\right)}{\mathbb{P}\left(|Z_{j,t}| > z_0\right)} \\ &= \frac{\left[\sum_c \mathbb{P}\left(|Z_{j,t}| > z_0 \,|\, \text{null},\, c\right) \mathbb{P}\left(c \,|\, \text{null}\right)\right] \left(\sum_{c \in D} p_c\right)}{p_{\text{hit}}} \\ &= \frac{\left[\sum_{c \in D} \mathbb{P}\left(|Z_{j,t}| > z_0 \,|\, c\right) \frac{p_c}{\sum_{c \in D} p_c}\right] \left(\sum_{c \in D} p_c\right)}{p_{\text{hit}}} \\ &= \frac{\sum_{c \in D} \mathbb{P}\left(|Z_{j,t}| > z_0 \,|\, c\right) p_c}{p_{\text{hit}}}. \end{aligned} \tag{21}$$

Since each term in equation 21 is known analytically, the FDR can be calculated directly as a function of $\{\mathbf{\Omega}_c\}$ and $\mathbf{\Sigma}_j$.

The formula holds for any number of traits, but to illustrate calculation of the formula, we consider a simple setting with two traits, A and B, and four classes of SNPs: those which are non-null for both A and B, those that are null for both A and B, those that are null only for A, and those that are null only for B. We denote these states as TT, FF, FT, and TF, respectively.

We assume the following variance-covariance matrices for the distribution of SNP effects in each state:

$$\mathbf{\Omega}_{TT} = \begin{bmatrix} 1 & 0.7 \\ 0.7 & 1 \end{bmatrix}, \quad \mathbf{\Omega}_{FF} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{\Omega}_{FT} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{\Omega}_{TF} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},$$

and we set the variance-covariance matrix of the estimation error to be

$$\boldsymbol{\Sigma}_j = \begin{bmatrix} 1 & 0.35 \\ 0.35 & 1 \end{bmatrix}.$$

for all SNPs j. Lastly, we set the mixing probabilities to

$$\boldsymbol{p} = \begin{bmatrix} p_{TT} \\ p_{FF} \\ p_{FT} \\ p_{TF} \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.2 \\ 0.2 \\ 0.1 \end{bmatrix}.$$

Under this parameterization, it is straightforward to calculate the average value of $\boldsymbol{\Omega}$:

$$\begin{aligned}
\boldsymbol{\Omega} &= \text{Var}\left(\boldsymbol{\beta}_j\right) \\
&= \mathbb{E}\left(\boldsymbol{\beta}_j \boldsymbol{\beta}_j'\right) \\
&= \sum_c \mathbb{E}\left(\boldsymbol{\beta}_j \boldsymbol{\beta}_j' \,|\, c\right) p_c \\
&= \boldsymbol{\Omega}_{TT} p_{TT} + \boldsymbol{\Omega}_{FF} p_{FF} + \boldsymbol{\Omega}_{FT} p_{FT} + \boldsymbol{\Omega}_{TF} p_{TF} \\
&= \begin{bmatrix} 1 & 0.7 \\ 0.7 & 1 \end{bmatrix} 0.5 + \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} 0.2 + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} 0.2 + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} 0.1 \\
&= \begin{bmatrix} .6 & 0.35 \\ 0.35 & .7 \end{bmatrix}.
\end{aligned}$$

Substituting this matrix into (19) for trait A, we have for each of the four states:

$$\text{Var}\left(Z_{j,A} \,|\, TT\right) = 2.08, \quad \text{Var}\left(Z_{j,A} \,|\, FF\right) = 0.99, \quad \text{Var}\left(Z_{j,A} \,|\, FT\right) = 1.01, \quad \text{Var}\left(Z_{j,A} \,|\, TF\right) = 1.83.$$

In our power calculations, we use the conventional P value threshold of 5×10^{-8}, which corresponds to a z-statistic threshold of

$$z_0 = \Phi^{-1}\left(1 - 5 \times 10^{-8}/2\right) = 5.45.$$

Using this threshold we see that

$$\begin{aligned}
\mathbb{P}\left(|Z_{j,t}| > z_0 \,|\, TT\right) &= 1.57 \times 10^{-4} \\
\mathbb{P}\left(|Z_{j,t}| > z_0 \,|\, FF\right) &= 4.04 \times 10^{-8} \\
\mathbb{P}\left(|Z_{j,t}| > z_0 \,|\, FT\right) &= 6.18 \times 10^{-8} \\
\mathbb{P}\left(|Z_{j,t}| > z_0 \,|\, TF\right) &= 5.77 \times 10^{-5}.
\end{aligned}$$

Finally, plugging these values into (20) and (21) give us

$$\begin{aligned}
\text{Power} &= 1.40 \times 10^{-4}, \\
\text{FDR} &= 2.42 \times 10^{-4}.
\end{aligned}$$

1.4.4 Bounding the FDR (maxFDR)

For any joint distribution of effect sizes that can be written as a mixture of normals, Equation (21) gives the FDR. Even if the exact mixing parameters are not known, the equation may yield an informative upper bound on the FDR. Here, we describe a framework for calculating this upper bound, which we refer to as maxFDR. The maxFDR calculations we propose are intended to supplement, not replace, other follow-up analyses, such as replication and prediction analyses in independent samples, that may be used by investigators to probe the credibility of any MTAG-identified loci.

The calculations that follow require two additional assumptions. First, we assume that the effect-size distribution is a mixture of 2^T multivariate normals, where T is the number of traits. These components correspond to the complete set of states in which the SNP is null for some subset of traits and non-null for the others.[5] Second, we assume that the variance of effect sizes for each trait is equal across all states in which the SNP is non-null for the trait, and that the covariance of effect sizes for each pair of traits is equal across all states in which the SNP is non-null for both traits. We use $\tilde{\mathbf{\Omega}}$ to denote the variance-covariance matrix of effect sizes in the state that the SNP is non-null for all traits. Under our assumptions, the variance-covariance matrix of effects sizes for component c, denoted $\mathbf{\Omega}_c$, is equal to $\tilde{\mathbf{\Omega}}$ with the rows and columns zeroed out that correspond to null traits in component c. We do not make any assumptions about the mixing probabilities of the mixture distribution (i.e., the frequencies of SNPs corresponding to each state).

As shown in section 1.4.3, the FDR is a function of $\{\mathbf{\Omega}_c\}$, $\mathbf{\Sigma}_j$, and the vector of mixing probabilities $\boldsymbol{p}$. There are simple relationships between $\mathbf{\Omega}$, $\{\mathbf{\Omega}_c\}$, and $\tilde{\mathbf{\Omega}}$. To relate $\mathbf{\Omega}$ and $\{\mathbf{\Omega}_c\}$, observe that:

$$\begin{aligned}\mathbf{\Omega} &= \text{Var}\left(\boldsymbol{\beta}_j\right) \\ &= \mathbb{E}\left(\boldsymbol{\beta}_j \boldsymbol{\beta}_j'\right) \\ &= \sum_c \mathbb{E}\left(\boldsymbol{\beta}_j \boldsymbol{\beta}_j' \,|\, c\right) p_c \\ &= \sum_c \mathbf{\Omega}_c p_c.\end{aligned}$$

Next, we relate $\mathbf{\Omega}$ and $\tilde{\mathbf{\Omega}}$. Considering each element separately, we see that the $(t,s)^{\text{th}}$ element of $\tilde{\mathbf{\Omega}}$ is

$$\tilde{\omega}_{t,s} = \frac{\omega_{t,s}}{q_{t,s}}, \tag{22}$$

where $q_{t,s}$ is the sum of the elements of $\boldsymbol{p}$ that correspond to states where the SNP is non-null for both trait t and s.[6] Note that sometimes putting together the elements of $\tilde{\mathbf{\Omega}}$ from this formula will generate a non-positive-semi-definite matrix $\tilde{\mathbf{\Omega}}$. If so, the assumed mixing probabilities make it impossible to attain the correlations in true effect sizes described by $\mathbf{\Omega}$.

Under the above assumptions, we can use (21), (22), and our estimates $\hat{\mathbf{\Omega}}$ and $\hat{\mathbf{\Sigma}}_j$ to calculate the FDR for *any* vector of probabilities $\boldsymbol{p}$. This allows us to assess the credibility of single-SNP associations identified by MTAG in two ways. First and most simply, we can evaluate the FDR under a wide range of assumptions about the vector $\boldsymbol{p}$. If such sensitivity analyses yield low values of FDR even under very pessimistic assumptions, then we may conclude that the MTAG findings are robust. As a rule of thumb, the FDR tends to be highest

[5] For example, in the $T = 2$ case, there will be four components: one for the SNP being null for both traits, one for the SNP being non-null for both traits, and two more for the cases where the SNP is null for one trait and not the other.

[6] If $q_{t,s} = 0$, then any value of $\tilde{\omega}_{t,s}$ is equivalent, so we arbitrarily pick $\tilde{\omega}_{t,s} = 1$ if $t = s$ and $\tilde{\omega}_{t,s} = 0$ if $t \neq s$ in those cases.

when a moderate fraction of the probability mass is on states where the SNP is null for the trait of interest but non-null for the other traits in the analysis.

Second, in cases with a small number of traits, it is possible to calculate the value of $\boldsymbol{p}$ that maximizes the FDR, possibly subject to some constraints on the vector $\boldsymbol{p}$. Such calculations can be used to generate worst-case-scenario values of each trait's FDR, which we call *maxFDR*. Without any constraints on $\boldsymbol{p}$, other than that its elements must sum to one, the $\boldsymbol{p}$ that maximizes the FDR is usually one where almost all SNPs (>99%) are null for all traits. In practice, the maxFDR may therefore fail to yield an informative upper bound in settings where the investigator has no prior information about the genetic architecture of the traits that can be used to constrain the search space. But in many realistic applications, it may be possible to incorporate information about the fraction of non-null SNPs. To illustrate, **Supplementary Figure 2a** and **2b** illustrate the maxFDR in a two-trait MTAG analysis of non-overlapping samples where the trait of interest is assumed to be Trait 1. The calculations are based on the assumption that at least 10% of SNPs are non-null for both traits. We consider this a conservative bound for many polygenic traits, but in applications of MTAG to less polygenic traits, the calculation would of course need to be conducted allowing for fewer than 10% of SNPs to be causal for both traits.

Supplementary Figure 2a plots maxFDR for a spectrum of effect-size correlations when one trait is low-powered ($\mathbb{E}\left(\chi^2\right) = 1.1$) and the other has different levels of power ($\mathbb{E}\left(\chi^2\right) \in \{1.1, 1.4, 2.0\}$). This figure highlights two important points. First, as the correlation of the effect sizes goes to one, maxFDR falls toward zero. This is because the inflation of FDR arises as a result of SNPs that are null for the trait of interest and non-null for other traits. If $r_\beta = 1$, such SNPs do not exist. Second, as expected, the FDR is moderate (less than 8%) when the secondary trait has limited power, but the maximum FDR can explode to concerning levels (greater than 15%) at certain levels of r_β when the mean χ^2-statistic is 2.0.

Supplementary Figure 2b shows values from MTAG applied to summary statistics from two equally-powered GWASs. When both traits are low-powered, maxFDR never exceeds 8% and decreases monotonically in r_β. In the moderate- and high-powered cases considered, however, the FDR is much smaller than 1% for all values of r_β.

1.5 Computational Run-time.

MTAG is computationally quick because all its steps have closed-form solutions. In the real-data application described in this paper—with three traits and 6.1M SNPs—the median run time across five identical runs using one core of a 2.20 GHz Intel(R) Xeon(R) CPU E5-2650 v4 processor was approximately 28 minutes. Of course, run time may vary as a function of the computing environment.

2 Simulations

2.1 Background

In this section we provide additional evidence, from simulations and real data, on MTAG's performance. We begin (section 2.2) with a general description of the data-generating process used throughout our simulations. Next, we verify that the analytic formulas for MSE, statistical power, and FDR derived in section 1.4 hold in our simulated data (as of course they must). Our remaining analyses test for biases in MTAG standard errors. Our theoretical derivation of the feasible MTAG estimator and its standard error (sections 1.2.4-1.2.5) did not account for the sampling variance in $\hat{\mathbf{\Omega}}$ and $\hat{\mathbf{\Sigma}}_j$. Theoretically, the size of the bias should increase with the number of traits included in the MTAG analysis. In section 2.4.1, we therefore explore quantitatively the biases resulting from unaccounted-for sampling variance as MTAG is progressively applied to a larger number (up to 20) of traits. Finally, we used individual-level data from UK Biobank to create "synthetic" cohorts with known overlap. We apply MTAG to summary statistics from genome-wide association analyses conducted in these cohorts and test for biases by comparing MTAG standard errors to those from a single GWAS conducted in the full sample.

2.2 Data-Generating Process in Simulations

The analyses in sections 2.3 and 2.4.1 were based on GWAS summary statistics simulated using the following procedure.

1. Generate 100,000 length-two "true" effect size vectors, $\{\boldsymbol{\beta}_j\}$, corresponding to the effects of 100,000 independent SNPs on two traits.
2. For each SNP, generate z-statistics, $\{\boldsymbol{Z}_j\}$. To do so, for each SNP j, we generate effect-size estimates by adding estimation error to the true effect size vectors $\boldsymbol{\beta}_j$, and then we divide the resulting effect-size estimate by the standard deviation of the estimation error. The estimation error is drawn from a bivariate normal distribution with mean zero and variance-covariance matrix $\mathbf{\Sigma}$.
3. Generate an "estimate" of $\mathbf{\Sigma}$, $\hat{\mathbf{\Sigma}}$, by adding independent, normally distributed noise to $\mathbf{\Sigma}$.
4. Using $\{\boldsymbol{Z}_j\}$ and $\hat{\mathbf{\Sigma}}$, generate an estimate of $\hat{\mathbf{\Omega}}$ using the method of moments procedure described in section (9.1.2) .
5. Using $\{\boldsymbol{Z}_j\}$, $\hat{\mathbf{\Sigma}}$, and $\hat{\mathbf{\Omega}}$, generate MTAG estimates and standard errors for each SNP and each trait.

These summary statistics are then used to conduct the intended test, and these steps are replicated many times for each simulation scenario. The number of replications varies across tests, depending on how much precision is needed.

Across our simulation scenarios, effect sizes can be distributed as multivariate normal or multivariate spike-and-slab. Here we describe how we generate the true effect sizes from each of these distributions. In all cases, we generate the effect-size distribution such that it has variance one (so that the effect sizes are all on a comparable scale) and correlation r_β (a parameter that we fix to different values in different scenarios).

For scenarios with normally distributed effect sizes, we simply draw bivariate normally distributed effect sizes with mean zero and variance-covariance matrix

$$\mathbf{\Omega}_{\text{normal}} = \begin{bmatrix} 1 & r_\beta \\ r_\beta & 1 \end{bmatrix}.$$

To simulate a scenario in which some fraction of SNPs has no association with each trait, we generate data from a multivariate spike-and-slab distribution. To do so, we first draw potential effect sizes from a bivariate normal distribution with mean zero and variance-covariance matrix $\mathbf{\Omega}_{\text{SS}}$. Then for each SNP, we set the effect of both SNPs to exactly zero with probability p_{FF}, we set the effect of just the first trait to zero with probability p_{FT}, and we set the effect of just the second SNP to exactly zero with probability p_{TF}. This means that there is probability $p_{TT} = 1 - p_{FF} - p_{FT} - p_{TF}$ that the SNP is non-null for both traits.

We let $\boldsymbol{p}$ to denote the vector

$$\boldsymbol{p} \equiv \begin{bmatrix} p_{FF} \\ p_{FT} \\ p_{TF} \\ p_{TT} \end{bmatrix}.$$

Setting

$$\mathbf{\Omega}_{\text{SS}} = \begin{bmatrix} \frac{1}{p_{TT}+p_{TF}} & \frac{r_\beta}{p_{TT}} \\ \frac{r_\beta}{p_{TT}} & \frac{1}{p_{TT}+p_{FT}} \end{bmatrix},$$

ensures that the effect sizes will have variance one and correlation r_β.[7]

To speed up computations, we directly generate a GWAS estimate vector $\hat{\boldsymbol{\beta}}_j$ in each replication rather than simulate data and estimate the effect sizes. Since standard GWAS estimators are asymptotically normally distributed conditional on the true effect size, we assume that $\left(\hat{\boldsymbol{\beta}}_j \,|\, \boldsymbol{\beta}_j\right)$ is bivariate normally distributed. Assuming that the estimator that produced $\hat{\boldsymbol{\beta}}_j$ is unbiased, the distribution of the estimation error has mean zero. Defining the correlation matrix of the estimation error to be[8]

$$\mathbf{\Sigma}_{LD} \equiv \begin{bmatrix} 1 & r_\varepsilon \\ r_\varepsilon & 1 \end{bmatrix},$$

the variance-covariance matrix of the estimation error will be $\mathbf{\Sigma} = \boldsymbol{C}\mathbf{\Sigma}_{LD}\boldsymbol{C}$, where

$$\boldsymbol{C} = \begin{bmatrix} \sqrt{\frac{1}{\chi_1^2 - 1}} & 0 \\ 0 & \sqrt{\frac{1}{\chi_2^2 - 1}} \end{bmatrix},$$

using χ_t^2 to denote the expected χ^2-statistic for trait t. This parameterization allows us to model the correlation of the estimation error and the power of the GWAS separately. Higher heritability and larger estimation sample size for trait t, and therefore greater power, are modeled by a larger value for the parameter χ_t^2. Greater phenotypic correlation (either positive or negative) and greater sample overlap are modeled by a value for $r_\varepsilon \in [-1, 1]$ further from zero.[9]

[7] For certain values of r_β and $\boldsymbol{p}$, the matrix $\mathbf{\Omega}_{\text{SS}}$ will not be positive definite, in which case it is not a valid variance-covariance matrix. In such cases, it is impossible to achieve a correlation of true genetic effects equal to r_β given the probabilities $\boldsymbol{p}$, so we will not able to perform simulations corresponding to those parameters.

[8] The notation $\mathbf{\Sigma}_{LD}$ is used because this matrix is equivalent to the matrix $\mathbf{\Sigma}_{LD}$ described in section (1.3.1).

[9] In principle, the value of r_ε would also capture the degree to which the biases in the estimated effect on one trait covary

After simulating the effect-size estimates for each trait, $\hat{\boldsymbol{\beta}}_j$, we calculate z-statistics. To do so, we divide the estimated effect sizes by the standard deviation of the sampling variance:

$$Z_{j,t} \equiv \frac{\hat{\beta}_{j,t}}{\sqrt{\sigma_{tt}}}.$$

When we apply MTAG to real data, we would use LD score regression to estimate $\hat{\boldsymbol{\Sigma}}$. In our simulations, however, all of the SNPs are independent, so there is no variation in the LD scores of each simulated SNP. Nonetheless, we would like our simulation to capture the effects of estimation error on $\hat{\boldsymbol{\Sigma}}$. To do so, we add independent noise to each element of the matrix $\boldsymbol{\Sigma}$, drawn from a normal distribution with mean zero and variance proportional to the sample size implied by the noncentrality parameter of the expected χ^2-statistic. In the empirical data in this paper, the traits have a mean χ^2-statistic of about 1.4, and the sampling variance of each of the elements of $\boldsymbol{\Sigma}_j$ is roughly 10^{-4}. Using these values to calibrate our simulations, we set the variance of the noise term added to the $(t,s)^{\text{th}}$ element (where t may be equal to s) of the true value of $\boldsymbol{\Sigma}_j$ to $10^{-4} \times \frac{0.4}{\sqrt{\mathbb{E}(\chi_t^2-1)\mathbb{E}(\chi_s^2-1)}}$.

2.3 Verification of Analytic Approximations for MSE, Power, and FDR

In section 1.4, we derived analytic formulas for MSE, statistical power, and FDR. The formulas are approximations when evaluated using $\hat{\boldsymbol{\Omega}}$ and $\hat{\boldsymbol{\Sigma}}_j$ instead of the true matrices $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}_j$. Our first set of simulations examined the accuracy of the approximations across a set of two-trait simulation scenarios; we examine multi-trait simulations in section 2.4.1 below.

In all simulations, we considered a two-trait MTAG analysis under varying assumptions about the following parameters: χ_1^2, χ_2^2, r_β, r_ε, and $\boldsymbol{p}$ (where $\boldsymbol{p}$ is relevant for the statistical power and FDR calculation but not the MSE formula). For the expected χ^2-statistics, we consider both a "low power" ($\chi_t^2 = 1.1$) and "high power" ($\chi_t^2 = 2.0$) setting. For r_β, we consider "high-correlation cases" where $r_\beta = 0.7$, such as what we observe in the empirical data in this paper for DEP, NEUR, and SWB, and "zero-correlation cases," where $r_\beta = 0$. We similarly use these values for r_ε. The high-correlation case for r_ε could correspond to a setting in which a pair of traits has a phenotypic correlation of 0.7 and the GWAS summary statistics are estimated in a perfectly overlapping sample. The no-correlation case would occur if the GWAS summary statistics are estimated from non-overlapping samples.

We consider three cases for $\boldsymbol{p}$. First,

$$\boldsymbol{p} = \begin{bmatrix} p_{FF} \\ p_{FT} \\ p_{TF} \\ p_{TT} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

This is the infinitesimal case where all SNPs are drawn from the same bivariate normal distribution. Note

with the biases in the estimated effect on the other trait. We ignore this source of covariance because we assume that the efforts to reduce the biases (e.g., quality control, inclusion of PCs as controls in the GWAS) render it negligible relative to the impact of phenotypic correlation and sample overlap.

that in this case, the FDR is always zero by definition because every SNP is non-null. Second,

$$\boldsymbol{p} = \begin{bmatrix} p_{FF} \\ p_{FT} \\ p_{TF} \\ p_{TT} \end{bmatrix} = \begin{bmatrix} 0.3 \\ 0 \\ 0 \\ 0.7 \end{bmatrix}.$$

In this setting, 30% of SNPs are null for both traits, but there are no SNPs that are null for one trait and non-null for the other. The value of 0.7 was chosen because [8] find that approximately 30% of SNPs are null for each trait DEP, NEUR, and SWB. The final case is

$$\boldsymbol{p} = \begin{bmatrix} p_{FF} \\ p_{FT} \\ p_{TF} \\ p_{TT} \end{bmatrix} = \begin{bmatrix} 0.09 \\ 0.21 \\ 0.21 \\ 0.49 \end{bmatrix}.$$

Note that these are the mixing probabilities that correspond to 30% of SNPs being null for each trait, but the event that a SNP is null for one trait is independent of the event that the SNP is null for the other.

In our simulations, we compare the true MSE, power, and FDR (i.e., calculated using $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}$) to the simulated MSE, power, and FDR (i.e., calculated using $\hat{\boldsymbol{\Omega}}$ and $\hat{\boldsymbol{\Sigma}}$) for every combination of these parameters over 1,000 replications, each including 100,000 SNPs. This gives us $2 \times 2 \times 2 \times 2 \times 3 = 48$ simulations. The results are reported in **Supplementary Table 1** and show a close correspondence between the true and simulated values. Since the non-centrality parameter of the χ^2-statistic scales linearly with sample size and heritability, we can use the mean χ^2-statistics to calibrate what these values correspond to in real-data applications. For example, an expected χ^2-statistic of 1.1 corresponds to a GWAS of height conducted in a sample of approximately 40,000 individuals, and an expected χ^2 statistic of 2.0 corresponds to the same analysis conducted in a discovery sample approximately ten times larger.

2.4 MTAG Standard Errors

2.4.1 Sampling variance in $\hat{\boldsymbol{\Omega}}$ and $\hat{\boldsymbol{\Sigma}}_j$ for $T \in [1, 20]$ traits

As described in section 1.2.4, the MTAG estimator and standard errors do not account for sampling variance in the estimates $\hat{\boldsymbol{\Omega}}$ and $\hat{\boldsymbol{\Sigma}}_j$. This can lead to inflation of MTAG test statistics. Since adding more traits increases the number of estimated elements in $\hat{\boldsymbol{\Omega}}$ and $\hat{\boldsymbol{\Sigma}}_j$ (with the number of elements increasing at the rate T^2), the bias is expected to be larger as the number of traits T increases. The simulations in this section assess the magnitude of this bias.

In these simulations, for each replication, we fix some value $r_\beta \in \{0.0, 0.7\}$, expected χ^2_t-statistic in $\{1.1, 1.4, 2.0\}$ for all traits t, and some total number of traits $T \in \{1, 2, 3, 5, 7, 10, 15, 20\}$.[10] We draw effects sizes for 100,000 SNPs from a multivariate normal distribution with mean zero and variance-covariance matrix $\boldsymbol{\Omega}$, where each diagonal element of $\boldsymbol{\Omega}$ is equal to one and each off-diagonal element of $\boldsymbol{\Omega}$ is equal to

[10] We restrict r_β to be positive because for any negative value of r_β, there is a T_0 such that $\boldsymbol{\Omega}$ is non-positive definite for any $T > T_0$.

r_β. We then simulate z-statistics, $\hat{\mathbf{\Omega}}$, and $\hat{\mathbf{\Sigma}}_j$ based on the procedure described in section 2.2. For each set of simulation parameters, we conduct 10,000 replications.

To measure the inflation of the MTAG test statistics, we first calculate the MTAG estimates in the standard way, using the estimated values $\hat{\mathbf{\Omega}}$ and $\hat{\mathbf{\Sigma}}_j$. We also calculate a set of "oracle" MTAG estimates, denoted $\hat{\beta}_{\text{oracle},j,t}$, which implement the MTAG formula but using the true, known values of $\mathbf{\Omega}$ and $\mathbf{\Sigma}_j$. We then calculate the mean χ^2-statistic of $\hat{\beta}_{\text{MTAG},j,t}$ and of $\hat{\beta}_{\text{oracle},j,t}$ across all SNPs and replications. Since under the null, the expected χ^2-statistic is 1, we compare these two estimates by taking the ratio

$$R_t = \frac{\overline{\chi^2_{\text{MTAG},t}} - 1}{\overline{\chi^2_{\text{oracle},j,t}} - 1} - 1,$$

where $\overline{\chi^2_{\text{MTAG},t}}$ and $\overline{\chi^2_{\text{oracle},t}}$ are the mean χ^2-statistics for the MTAG and oracle results, respective. This value is a measure of the fraction of inflation of the MTAG summary statistics due to over-fitting that results from treating the estimates $\hat{\mathbf{\Omega}}$ and $\hat{\mathbf{\Sigma}}_j$ as the true values.

Figs 1a-1b shows the results when the data are simulated without overlap (with $r_\varepsilon = 0$), but we verified that the findings are very similar if the data are simulated with moderate overlap ($r_\varepsilon = 0.35$). The inflation of the MTAG test statistics increases approximately linearly in the number of traits, but in most scenarios, the bias is quantitatively small. In well-powered MTAG analyses of five or fewer traits, the inflation never exceeds 0.5% across all the scenarios we considered. We observe the greatest inflation (up to 3%) when each GWAS has low power and the number of traits is large.

2.4.2 MTAG Standard Errors with Sample Overlap

For reasons of computational speed, the simulation evidence shown so far was based on simulations in which we directly drew SNPs' true effect sizes and z-statistics, rather than using real data. Among other things, this approach allows us to avoid estimating LD score regressions in each simulation run. A limitation of the approach, however, is that we cannot use it to test how well MTAG corrects for overlapping GWAS samples because MTAG's correction relies on LD score regression—and by extension, the assumptions about genotype and phenotype data that underlie LD score regression. In practice, it is important to know whether violations of these assumptions introduce biases in applications of MTAG to real data.

To address that question, we conducted analyses using real genotypic and phenotypic data from the initial release of the UK Biobank (UKB). We created synthetic UKB subcohorts with known overlap by drawing three equally sized subsamples, chosen so that the pairwise overlap is always equal to 50%. More precisely, we divided the data into thirds, and each cohort was made up of two of the thirds. In each cohort, we conducted genome-wide association analyses of six traits. We then applied MTAG to the summary statistics for each trait and compared MTAG z-statistics to z-statistics from a conventional GWAS of the complete UKB sample (from which the synthetic subcohorts were drawn).

All genome-wide association analyses were conducted in "White-British ancestry" subjects in the interim release of the UK Biobank[11] using association models described in section 3. We considered six typical GWAS traits: height ($N = 112,151$), body mass index (BMI, $N = 112,031$), educational attainment

[11] Marchini, J. et al. Genotype Imputation and Genetic Association Studies of UK Biobank: Interim Data Release. Tech. Rep. (2015).

(EA, $N = 111,349$), depressive symptoms (DEP, $N = 101,615$), neuroticism (NEUR, $N = 104,439$), and subjective well-being (SWB, $N = 40,603$). We constructed our DEP, NEUR and SWB phenotypes using the same procedures as in the single-trait GWASs described in section 3, and our measure of EA was constructed exactly as in [9].

If MTAG standard errors accurately correct for the sample overlap, the MTAG effect size estimates, their standard errors, and the z-statistics should be approximately equal to those from a conventional GWAS conducted in the original UKB sample from which the synthetic cohorts were drawn. And indeed, for all six trait considered, we found that the correlation between MTAG and GWAS z-statistics was very high (R^2's are in the range 0.964 to 0.999). **Figs 2a-b** show results for height and DEP. Analogous figures for the remaining four traits are shown in **Supplementary Figure** 3. Each figure also reports the slope of the regression line from a regression of the MTAG z-statistics on the GWAS z-statistics. As expected, the estimated slopes are consistently close to 1 (range 0.982 to 1.021). We also verified that MTAG correctly recovers the degree of overlap. In each MTAG analysis, the estimated values of the off-diagonal elements of $\hat{\mathbf{\Sigma}}_j$ are always approximately 0.5.

3 GWAS Meta-Analyses of Depression, Neuroticism and Subjective Well-Being

3.1 Background

We jointly analyze three traits: depression (DEP), neuroticism (NEUR), and subjective well-being (SWB). We selected these three traits because previous work has found them to be highly genetically correlated [8] and because summary statistics from published GWA studies with large discovery samples were available for all three traits when we launched the study [8, 10]. The traits' high genetic correlations with each other imply that the benefits from joint analyses may be substantial.

MTAG requires one input file for each trait included in the analysis. Each input file contains association results from conventional (single-trait) association analyses. Each row in an input file contains results from a test of association between a specific genetic variant and the trait in question. In our setting, the three files contain results from GWAS meta-analysis. In this section, we describe the meta-analyses used to generate the three input files and summarize key findings.

3.2 Summary Overview of Single-Trait Analyses

Our single-trait input files are the results of GWAS meta-analyses based on association analyses of 1000G-imputed variants from the 22 autosomal chromosomes. For a given meta-analysis, each file that enters into the meta-analysis is either obtained from previously published GWASs or from new genome-wide analyses. **Fig. 3** gives a schematic overview of the three meta-analyses (see also **Supplementary Table** 2):

- Depression (DEP; $N = 465,337$; $N_{\text{eff}} = 354,862$)
- Neuroticism (NEUR; $N = 168,105$)
- Subjective Well-Being (SWB; $N = 388,538$).

All of our analyses rely heavily on the summary statistics from [8], a previous GWAS conducted by the Social Science Genetic Association Consortium (SSGAC). The SSGAC study analyzed subjective well-being, depressive symptoms, and neuroticism from association results in a large number of cohorts. In the SSGAC study, cohort-level association results from the UK Biobank contributed to the meta-analyses of all three phenotypes, and cohort-level results from 23andMe contributed to the meta-analysis of SWB.

For the purposes of this study, we reran all three original SSGAC meta-analyses with updated association results from the UK Biobank (interim release)[12]. The new UK Biobank results are based on a slightly revised analysis protocol. The revisions include changes to phenotype definition and, in the association analyses, a more comprehensive set of controls for the genotype-measurement batch. The revisions are all minor, and it is hence unsurprising that they yield substantively (but not numerically) identical results. Most importantly, we expanded the meta-analysis of [8] of SWB using updated association results from a substantially increased discovery sample from the 23andMe cohort ($N = 93,454$ in the earlier sample versus $N = 252,053$ in the new

[12] Marchini, J. et al. Genotype Imputation and Genetic Association Studies of UK Biobank: Interim Data Release. Tech. Rep. (2015).

sample), and we expanded Okbay *et al.*'s meta-analysis of DEP by adding results from a recently published study of self-diagnosed major depression conducted in a sample composed primarily of 23andMe research participants[10].

To summarize, our DEP meta-analysis combines summary statistics from three sources: (i) a recent study of major depressive disorder in a large sample of 23andMe research participants[10] ($N = 307,354$ in total, with 75,607 cases), (ii) the SSGAC analysis of major depression in the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort[13] ($N = 56,368$ in total, with 7,231 cases), and (iii) an analysis of an index of depressive symptoms in the interim release of the UK Biobank ($N = 101,615$) that is revised relative to the previous SSGAC analysis of the same data. Each of these are population-based samples and are therefore not recruited based on depression status. Details on the phenotypes used in the final analysis are shown in **Supplementary Table 3 Panel A**.[14] Despite the heterogeneity in the measures used, estimates of genetic overlap derived from linkage disequilibrium (LD) score regression [6] suggest that the genetic overlap between the measures of depression is high: $\hat{r}_g = 0.99$ ($s.e. = 0.19$) for 23andMe/GERA, $\hat{r}_g = 0.73$ ($s.e. = 0.04$) for 23andMe/UKB, and $\hat{r}_g = 0.94$ ($s.e. = 0.19$) for GERA/UKB.

In our NEUR meta-analysis, we closely follow [8] and meta-analyze association results obtained from a genome-wide analysis of neuroticism in the UK Biobank ($N = 104,439$) with results from the published GWAS on neuroticism conducted by the Genetics of Personality Consortium (GPC) ($N = 63,666$) [11]. The neuroticism measures in each study are described in **Supplementary Table 3 Panel B**. LD score regression estimates suggest that the genetic correlation between the GPC and UKB neuroticism measures is very high: $\hat{r}_g = 1.12$ ($s.e. = 0.14$).

In our meta-analysis of SWB ($N = 388,542$), we use results from the "post hoc" GWAS of SWB reported by the SSGAC (the main analysis of SWB was based on HapMap2-imputed variants, but the study also reported results from a "post hoc" meta-analysis conducted in the subset of cohorts with 1000G-imputed data). Omitting the UKB and 23andMe cohorts from this meta-analysis leaves 21 cohorts with a combined sample size of 95,886. Adding association results from the UK Biobank based on the revised analysis protocol (resulting in an addition of $N = 40,603$ individuals) and association results from the substantially increased 23andMe sample ($N = 93,454$ in the earlier sample versus $N = 252,053$ in the new sample) gives a combined sample size of 388,542 individuals.

To maximize statistical power, the analysis of SWB in [8] included cohorts with association results for life satisfaction (LS) and cohorts with association results for positive affect (PA), two facets of SWB that are typically distinguished in the literature.[15] Following [8], we use association results from analyses of a combined measure of LS and PA (generally constructed by averaging LS and PA phenotypes) when available, LS if both LS and PA association results are available in separate files, and PA otherwise. The phenotype measures used in our SWB analysis are summarized in **Supplementary Table 3 Panel C**.

[13] GERA. Resource for Genetic Epidemiology Research on Adult Health and Aging (2015).

[14] Note that in [8], the GWAS of DEP included summary statistics from the Psychiatric Genetics Consortium. In this paper, however, we restricted ourselves only to cohorts that have been imputed to the 1000 Genome Reference panel. As a results, these data were omitted.

[15] PA refers to the frequency and intensity of positive emotions and feeling happy. Typical survey questions used to gauge PA include "During the past week, I was happy?" and "How would you rate your emotional well-being at present?" LS refers to a longer-term evaluation of one's life. A typical survey question would be "How satisfied are you with your life as a whole?".

3.2.1 Genotyping and Imputation

Details on the genotyping and imputation procedures applied by the cohorts in the SSGAC study have previously been reported in Supplementary Table 4 of a previously published study [8]. Our new analyses of UK Biobank and 23andMe, as well as the analysis in [10] of depression in a sample of 23andMe research participants, were also conducted using the cohort-specific filters listed there.

3.2.2 Association Analyses

In our new analyses of UK Biobank and 23andMe samples, we used the following regression equation for each SNP:

$$Y = \beta_0 + \beta_1 SNP + \boldsymbol{PC}\,\boldsymbol{\gamma} + \boldsymbol{B}\,\boldsymbol{\alpha} + \boldsymbol{X}\,\boldsymbol{\theta} + \varepsilon,$$

where Y is an unstandardized outcome variable; SNP is the allele dose of the SNP; $\boldsymbol{PC}$ is a vector of principal components of the variance-covariance matrix of the genotypic data, estimated after the removal of genetic outliers; $\boldsymbol{B}$ is a vector of standardized controls, including sex, age and their interactions; and $\boldsymbol{X}$ is a set of cohort-specific controls.

In the new UK Biobank analyses, we control for 15 principal components and indicator variables for all year-of-birth and sex category combinations. We also control for indicators for the genotype-measurement batch.

In our new updated analyses of SWB in the 23andMe cohort, we control for ten principal components, sex, a cubic polynomial of age, and interactions between sex and the age variables. We also control for indicator variables for the genotyping platform used by 23andMe.

3.2.3 Reference Panel

Our description of our quality control and pruning procedures makes frequent reference to a reference file, whose construction we define here (for a detailed account, see **Supplementary Note section 1.I** in [8]). Briefly, our reference file is constructed by processing publicly available data on CEU (Utah Residents (CEPH) with Northern and Western European Ancestry), TSI (Toscani in Italia), or GBR (British in England and Scotland) individuals in two releases (Phase 1 and Phase 3) of data from the 1000 Genomes Project. The final reference sample is restricted to 294 approximately unrelated individuals of European ancestry (as defined above). We impose the restrictions described in [8], giving us a reference file with 14,680,555 autosomal and biallelic SNPs. In the reference file, each variant has an rsID that maps to a unique ChrPosID (a concatenation of a SNP's chromosome number, a colon, and the SNP's base pair position). We do not include any variants for which, between Phases 1 and 3, the alleles are not consistent or the base pair coordinates changed.

3.3 Quality-Control Protocol

Our three meta-analyses are based on results files that have been cleaned using the exact same set of quality-control filters and diagnostic checks used in cleaning the files from cohorts with 1000 Genomes imputed data in the post hoc meta-analysis of subjective well-being in [8]. In this section, we provide a summary overview

of these checks and describe how they were implemented in the new results files not used in the analyses of [8]. For a detailed account, we refer readers to **Supplementary Note 1** in the SSGAC study

3.3.1 SNP-Level Quality Control

Using EasyQC [12], we filtered out SNPs from each of the uploaded results files in the following order (see **Supplementary Table 4** for the exact parameter thresholds used in some steps):

Misalignment. In cohorts with genotype data imputed against the September or December 2013 releases of the 1000 Genomes Phase 1 haplotypes provided by the software IMPUTE2, we dropped the 929 SNPs whose strands are known to have been incorrectly aligned in these releases.[16]

Indels and Structural Variants. We dropped a SNP if neither an effect nor other allele was supplied, or if either of them takes values other than "A", "C", "G", or "T".

Missing from Reference File. We dropped the SNP if its rsID was not available in the reference file.

Variable Quality. We dropped a SNP if any of the following variables were missing: P value, a coefficient estimate (beta) and its standard error, effect allele frequency, sample size (N), or imputation accuracy (for imputed SNPs). We also dropped SNPs if any of the variables reported for the SNP were outside the permissible range of the variable (for example, P values greater than 1 or negative standard errors).

Minor Allele Count. We dropped SNPs with minor allele count below 30 (as specified in **Supplementary Table** 4).

Imputation Accuracy. We filtered out SNPs with low imputation accuracy. The definition of the imputation accuracy metric varies by imputation software. If the cohort supplied us with the "*Rsq*" variable generated by MaCH [13], we dropped SNPs with $Rsq < 0.6$. If they uploaded the "INFO" variable generated by IMPUTE [14], we applied a threshold of 0.7. If PLINK's "info" variable was supplied, we applied a threshold of 0.8.

Hardy-Weinberg Equilibrium. We dropped genotyped SNPs with low Hardy-Weinberg equilibrium (HWE) P value (see **Supplementary Table 4** for exact cutoffs used).

Call Rate. We dropped SNPs with call rate below 95%.

Duplication or Allele Mismatch. If multiple SNPs in a results file were mapped to an identical ChrPosIDs in the reference file, we dropped the SNPs. We also dropped SNPs that could not be successfully aligned because the reference and other allele in the results file did not match those in the reference file.

Supplementary Table 5 shows, separately for every results file contributing to each our three meta-analyses, the number of SNPs dropped in each of the filtering steps. The table also lists the estimated genomic control factors from each cohort. All cleaned results files except those from the new analyses of UK Biobank (DEP, NEUR and SWB) and the 23andMe cohorts (DEP and LS) are identical to those used by the SSGAC. For previously used files, the filtering numbers are identical to those reported in the previous SSGAC study. For the new cohorts—DEP, NEUR, and SWB in UK Biobank and DEP and SWB in 23andMe—none of the filtering steps result in an unusual number of SNPs being dropped. Across cohorts, the estimated genomic control factors exhibit a strong relationship with sample size, as expected under polygenicity [15]. In small cohorts, the estimated inflation factors are all close to one.

[16] 730 SNP were corrected in the December 2013 release and 199 were corrected in the June 2014 release. The announcement is available on https://mathgen.stats.ox.ac.uk/impute/impute_v2.html#whats_new.

3.3.2 Other Diagnostics

Having processed the data through these filters, we prepared and inspected several diagnostic plots.

1. *Allele Frequency Plots* (*AF Plots*): We looked for errors in allele frequencies and strand orientations by inspecting a plot of sample allele frequencies against the allele frequency in a European reference sample.
2. *P-Z Plots*: We checked that reported P values are consistent with the reported coefficient estimates and their standard errors (SE's).
3. *Q-Q Plots*: We visually inspected the cohort-level Q-Q plots to look for evidence of unaccounted-for stratification.

We also verified that the SE's predicted from the N's and SD's supplied in the descriptive statistics matched the SE's in the results files. [12] proposes a similar diagnostic (the *SE-N Plots*), which is based on following approximation to the standard error of a coefficient estimated by OLS:

$$SE_j \approx \frac{\hat{\sigma}_Y}{\sqrt{N}} \cdot \frac{1}{\sqrt{MAF_j\,(1 - MAF_j)}}, \tag{23}$$

where $\hat{\sigma}_Y$ is the standard deviation of the dependent variable (equal to 1 in cohorts that reported standardized regression coefficients), MAF_j is the minor allele frequency of SNP j, and N is the sample size. We used Equation (23) to generate a predicted standard error for the 50K SNP set, and we then plotted these predicted standard errors against the reported standard errors. We used an analogous equation for cohorts with binary dependent variables that ran logistic regressions. These plots, which we refer to as *50K plots* in what follows, were used to check for systematic discrepancies between the predicted and reported standard errors and for outlier SE's.

3.4 Meta-Analyses

All meta-analyses were conducted in the software Metal [16] and are based exclusively on results files that have passed the diagnostic tests described in the previous sections. We do not apply cohort-level genomic control [17] to adjust the standard errors for non-independence. Instead, we meta-analyze the unadjusted cohort-level summary statistics, and we subsequently inflate the standard errors from the meta-analysis by the square root of the estimated intercept from an LD score regression [1].

In the meta-analysis of depression, we weight the UKB results (of a continuous index of depressive symptoms) by the size of the estimation sample. We weight the two case-control studies by effective sample size, defined as

$$N_{\text{eff}} = \frac{4}{N_{\text{cases}}^{-1} + N_{\text{controls}}^{-1}}, \tag{24}$$

as recommended by [16].

In our meta-analysis of neuroticism in UK Biobank and GPC, we elected against sample-size weighting. The reason is that even though the genetic correlation between UK Biobank and GPC is very high, the SNP-based heritability in GPC is substantially lower. Under such conditions, sample-size weighting is inefficient.

Intuitively, that is because association results from a sample with a noisier phenotype measure are less informative, implying that it is appropriate to weight GPC results less than proportionately to their sample size.

To generate the optimal weights, we proceeded in two steps. We first used a standard approximation to calculate the estimated effect (and associated standard error) of an additional copy of the reference allele on the dependent variable measured in standard-deviation units. The approximation is:

$$\hat{\beta}_j \approx z_j \frac{1}{\sqrt{2\,N_j\,MAF_j\,(1-MAF_j)}}. \tag{25}$$

In the second step, we transformed the standardized coefficients and SE's of [11] (but not the UKB) by the factor $\sqrt{h^2_{\text{UKB}}/h^2_{\text{GPC}}} \approx \sqrt{0.131/0.035} = 1.93$. The resulting coefficients and SE's (transformed in GPC and untransformed in UKB) were then meta-analyzed using inverse-variance weighting. It can be shown that this weighting scheme is equivalent to the MTAG estimator for the special case where the samples do not overlap, the traits are perfectly correlated genetically, and the two traits have different heritabilities.

3.5 Results

The QQ plots from the three analyses are shown in **Supplementary Figures 4a**, **b**, and **c**. All show strong evidence of inflation: $\lambda_{GC} = 1.36$ for DEP, $\lambda_{GC} = 1.24$ for NEUR, and $\lambda_{GC} = 1.28$ for SWB. The estimated LD score regression intercepts are 1.013 (DEP), 0.990 (NEUR) and 1.016 (SWB), suggesting that nearly all of the observed inflation is due to polygenicity; see **Supplementary Figures 5a**, **b**, and **c** and **Supplementary Table 6**. Bivariate LD score estimates of genetic correlation and bivariate LD score intercept estimates can be found in **Supplementary Tables 7** and **8**, respectively.

Supplementary Table 9 reports the set of approximately independent SNPs that reached nominal significance ($P < 10^{-5}$) in each of our three meta-analyses. To define independent loci, we used the following algorithm (the same as that described in the **Online Methods**). First, the SNP with the smallest P value is identified in the pooled meta-analysis results. This SNP is the lead SNP of clump 1. Second, we identified all SNPs on that chromosome whose LD with the lead SNP exceeds $R^2 = 0.1$ and assigned them to the clump. We calculate LD using the reference file whose construction was described in section **3.2.3**. To generate the second clump, the SNP with lowest P value among the SNPs that remain after removal of clump 1 is identified, and the same steps are applied to identify the set of SNPs comprising clump 2. The process is repeated until no SNPs remain with P values below 10^{-5} (or whatever is the desired P value threshold).

Supplementary Table 9 reports the number of loci identified in our genome-wide analyses at a P value threshold of 5×10^{-8} ("lead SNPs"). Our pruning algorithm yields 32 lead SNPs in the DEP meta-analysis (248 SNPs at $P < 10^{-5}$), 12 lead SNPs in the NEUR meta-analysis (165 SNPs at $P < 10^{-5}$), and 13 lead SNPs in the SWB meta-analysis (165 SNPs at $P < 10^{-5}$). Since 3 of the lead SNPs in the NEUR GWAS are located in the inversion region identified by [8], we collapse them into a single locus, giving us a total of 10 lead SNPs. (Note that when we estimate MTAG,we drop all SNPs in the inversion region. As a result, when we compare the number of hits between the GWAS and MTAG results, we only report 9 lead SNPs in the GWAS of NEUR.) Since our NEUR meta-analysis is very similar to the SSGAC study, it is unsurprising that the findings are nearly identical to those reported in the SSGAC GWAS (10 lead SNPs in our analyses, all of which reached genome-wide significance in the previous study).

Our meta-analyses of DEP and NEUR are based on substantially larger discovery samples than previously published GWAS of these traits [8, 10, 18, 19]. Consistent with what has been found for other complex traits as discovery samples have increased [20], the number of genetic associations identified at genome-wide significance in our analyses is larger than in previous studies. In our DEP meta-analyses, we identify 32 lead SNPs ($N_{\text{eff}} = 354,862$), compared to 4 in the discovery-stage analysis of [10] ($N_{\text{eff}} = 228,032$). In our SWB meta-analysis, increasing the sample size from $N = 229,883$ to $N = 388,538$ increases the number of lead SNPs from 2 to 13.

Manhattan plots with each of the approximately independent genome-wide significant hits colored in yellow are shown in **Supplementary Figures 6a**, **b**, and **c**.

4 MTAG Analysis of DEP, NEUR and SWB

4.1 Analyses

Using the procedures in **Online Methods**, we applied MTAG to the three sets of summary statistics from the single-trait GWASs described in section 3 and **Fig. 3**. See Panels A in **Supplementary Tables 10-13** for details on our main MTAG analyses, which were restricted to a set of $\sim 6M$ SNPs that passed MTAG SNP filters in each of the single-trait GWASs (panels B-D contain analogous results for MTAG analyses used in the validation replication analyses summarized in sections 5 and 6).

4.2 Results

We applied our clumping algorithm (**Online Methods**) to each set of MTAG association statistics, using a P value threshold of 10^{-5}. Q-Q plots comparing the GWAS and MTAG results for each trait are in **Supplementary Figure** 7. The association results are shown in **Supplementary Table 14**, with the set of approximately independent genome-wide significant loci marked in boldface.

To probe the robustness of the single-SNP findings, we used the procedures in section 1.4.3 to calculate the maxFDR of each trait. In these calculations, we assumed that at least 10% of SNPs are non-null for each of the three traits (a much lower degree of polygenicity than is implied by the results in **Online Methods**, where we report maximum-likelihood estimates above 50% for all three traits). Procedurally, we used a grid-search procedure that evaluated every feasible vector of mixing probabilities $\boldsymbol{p}$ in 0.1 unit increments. We found that in our application the maxFDR bound is very low for all three traits: 0.0014 for DEP, 0.0080 for NEUR, and 0.0044 for SWB. For all three traits, the FDR is maximized at

$$\boldsymbol{p} = \begin{bmatrix} p_{FFF} \\ p_{FFT} \\ p_{FTF} \\ p_{FTT} \\ p_{TFF} \\ p_{TFT} \\ p_{TTF} \\ p_{TTT} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0.2 \\ 0 \\ 0.2 \\ 0.2 \\ 0.4 \end{bmatrix}.$$

5 Replication of MTAG-Identified Loci

Following procedures described in **Online Methods**, we sought to replicate the loci identified at genome-wide significance in our empirical application of MTAG to DEP, NEUR, and SWB summary statistics. Below, we provide details regarding these replication analyses.

5.1 MTAG-identified Loci Included in Replication Analysis

Our replication analyses were conducted in two samples: the Health and Retirement Study (HRS, [21]) and the National Longitudinal Study of Adolescent to Adult Health (Add Health).[17] Because the HRS was included in the main analysis of section 3we reran the main MTAG analysis omitting HRS. Panels B of **Supplementary Tables 10-13** contain some summary information about this auxiliary MTAG analysis.

5.2 Phenotype Measures

5.2.1 HRS

The HRS contains high-quality measures of subjective well-being (SWB), depressive symptoms (DEP), and neuroticism (NEUR). We constructed the final phenotypes analyzed in our prediction analyses using data from the survey waves of 2006, 2008, 2010 and 2012. If a phenotype was not always measured consistently across waves (e.g., using the same psychometric measurement scale), we used the measurement scale available in the largest number of waves.

DEP. In the 2006, 2008, 2010 and 2012 waves, the HRS administered a mental health screening battery known as the World Health Organization Composite International Diagnostic Interview Short Form (CIDI-SF) scale [22] to a subset of respondents. The CIDI-SF begins with three screening questions designed to identify participants who are statistically more likely to satisfy diagnostic criteria for major depression. Respondents selected for additional surveying based on their responses to the screening questions are asked seven more follow-up questions about specific symptoms.

Responses to these seven follow-up questions can be mapped to a summary score equal to the number of symptoms that a respondent endorses. Sometimes, researchers use a cutoff of the summary score to determine whether or not a respondent meets clinical requirements for major depression. Rather than pursue this strategy, we instead use the methodology introduced by Nelson (1998)[18] to assign a probability of "caseness" to each subject based on their profile of symptoms. Conceptually, the probability is an estimate of the probability that the respondent would be categorized as having the disorder if they were administered the full CIDI. Respondents who, on the basis of their responses to the screening questions, are not selected for follow-up questions are assigned a probability of zero. Our final measure of DEP is calculated by averaging

[17] We had special permission to use the latest release of the Add Health data, which at the time we ran the analyses in this paper was in a "Freeze 1" hold. In Freeze 1, 7,598 individuals were genotyped using the Illumina Human Omni 1 chip, and 2,098 individuals were genotyped on the Illumina Human Omni 2.5 chip. Because nearly 85% of subjects genotyped using the Omni 2.5 chip were of non-European ancestry, we excluded individuals on this chip both to reduce bias introduced by batch effects and to avoid an unnecessary loss of SNPs when taking the intersection of SNPs common to both chips. (Harris, K. M. The Add Health Study : Design and Accomplishments. Chapel Hill: Carolina Population Center, University of North Carolina at Chapel Hill 122 (2013).)

[18] Nelson, C. B., Kessler, R. C. & Mroczek, D. Scoring the World Health Organization's Composite International Diagnostic Interview Short Form (CIDI-SF; v1.0 NOV98) (1998).

the subject's inferred probability of depression from all waves with non-missing data. This phenotype is available for $N = 8,307$ subjects.

NEUR. Personality was measured in the four waves between 2006 and 2012 using the Midlife Development Inventory personality scales [23]. An individual's neuroticism score in a wave is calculated from their level of agreement with four claims about themselves ("Moody," "Worrying," "Nervous," and "Calm"). For each claim, the respondent is asked to choose one of our response categories, ranging from "A lot" to "Not at all." We use the numerical coding in the HRS variable documentation [24] to map the four categorical responses into a single neuroticism score. We set the variable to missing whenever at least one of the four items is missing. Our final measure of NEUR is defined as the average neuroticism score across waves with non-missing data. The final NEUR variable is available for $N = 8,197$ subjects.

SWB. We obtain the SWB phenotype by combining measures of life satisfaction (LS) and positive affect (PA). The HRS does not contain a measure of PA that was consistently administered across all four waves between 2006 and 2012. In 2008, 2010 and 2012, however, PA and LS were measured consistently, and we therefore construct our SWB variable using data from these three waves. Our PA variable is derived from responses to thirteen items, eleven of which are from the Positive and Negative Affect Schedule—Expanded Form (PANAS-X; [25]). The remaining two are from other studies [26, 27]. LS was measured using the Satisfaction with Life Scale [28], which asks respondents to indicate their level of agreement with five statements (e.g., "In most ways my life is close to ideal").

We construct our wave-level measure of LS by mapping the subject's response to each LS item to a numerical value as described in the HRS manual. The wave-level measure of LS is constructed by averaging the scores across the five items [29]. We similarly construct our wave-level measure of PA by mapping responses to the 13 PA items to numerical values and averaging the numerical values across the items. In each wave, we create a SWB score by standardizing the LS and PA measures and taking a simple average of them. Our final SWB variable is the mean SWB score across waves with non-missing data. This final SWB phenotype is available for $N = 6,857$ subjects.

5.2.2 Add Health

Add Health also contains comparable measures of DEP, NEUR, and SWB, this time measured in a nationally-representative group of US adolescents and young adults. We constructed the final phenotypes analyzed in our prediction analyses using data from questions asked in survey Waves III and IV, which were conducted in 2001-2002 and 2008, respectively.

DEP. In Wave IV, Add Health created a constructed variable made up of questions related to depressive symptoms. This variable is an additive, shortened version of the CES-D scale, which is a self-report scale designed to measure depressive symptomatology in the general population [30]. Using the question prompt "How often was the following true during the past seven days?" respondents were asked to report on: (1) being bothered by things that usually don't bother them; (2) feeling unable to "shake off the blues", even with help from family and friends; (3) having trouble keeping their minds on what they were doing; (4) feeling depressed; and (5) feeling sad. Response options for each of the five questions included "never or rarely", "sometimes", "a lot of the time", and "most of the time or all of the time", and were coded 0, 1, 2, and 3, respectively. The final measure of DEP was obtained by adding responses for these five questions, giving a total range of 0-15 for $N = 4,334$ subjects.

NEUR. Also in Wave IV, Add Health created a constructed variable, this time made up of questions related to neuroticism. This variable is an additive, shortened version of the Mini-International Personality Item Pool (Mini-IPIP; [31]), which is itself a shortened version of the full International Personality Item Pool developed by Goldberg *et al.* ([32]) to measure the Big Five personality traits. Using the question prompt "How much do you agree with each statement about you as you generally are now, not as you wish to be in the future?" respondents were asked to report on: (1) having frequent mood swings; (2) being relaxed most of the time; (3) getting upset easily; and (4) seldom feeling blue. Response options for each of the four questions included "strongly agree", "agree", "neither agree nor disagree", "disagree", and "strongly disagree", and were coded 1, 2, 3, 4, and 5, respectively. Add Health reverse coded the first three questions, and the final measure of NEUR was obtained by adding responses for these four questions, giving a total range of 4-20 for $N = 4,332$ subjects.

SWB. In Add Health, the SWB phenotype is a measure of life satisfaction (LS), assessed in Wave III. Respondents were asked "How satisfied are you with your life as a whole?" Response options included "very satisfied", "satisfied", "neither satisfied nor dissatisfied", "dissatisfied", and "very dissatisfied", and were coded 1, 2, 3, 4, and 5, respectively. To obtain the final measure of SWB, we reverse coded this variable, giving a total range of 1-5 for $N = 3,673$ subjects.

5.3 Winner's Curse Correction

MTAG estimates are corrected for winner's curse following procedures previously described [8]. Briefly, for each trait, we use maximum likelihood to fit the MTAG results to a (univariate) spike-and-slab distribution such that

$$\beta_j \sim \begin{cases} 0 & \text{with probability } \pi \\ N\left(0, \tau^2\right) & \text{otherwise.} \end{cases}$$

For DEP, NEUR, and SWB, we estimate $\hat{\pi}$to be 0.598, 0.652, and 0.633 and $\hat{\tau}^2$ to be 3.12×10^{-6}, 5.05×10^{-6}, and 2.15×10^{-6}, respectively. We then use these estimates as the parameters of the prior distribution and calculate the posterior distribution of the effect size β_j given the estimate $\hat{\beta}_{\text{MTAG},j}$ for each SNP as

$$\hat{\beta}_{\text{adj},j} = (1 - \pi_{\text{post},j}) \frac{\hat{\tau}^2}{\hat{\tau}^2 + \hat{s}_j^2} \hat{\beta}_{\text{MTAG},j},$$

where $\pi_{\text{post},j}$ is the posterior probability that $\beta_j = 0$ and $\hat{s}_j^2$ is the squared standard error of the MTAG estimate.

5.4 Results

Supplementary Table 15 reports the bivariate LD score regression intercept estimates for each discovery and replication cohort. In each case, the intercept is insignificant with a standard error of approximately 0.005 (with all point estimates smaller than 0.007 in absolute value), consistent with there being no overlap (or minimal overlap) between any discovery and replication cohort. By the properties of LD score regression, the intercept for a pair of GWAS summary statistics for the same trait is equal to $\frac{N_s}{\sqrt{N_1 N_2}}$, where N_s is the overlapping sample size, N_1 is the sample size of the first GWAS and N_2 is the sample size of the second. So for each trait, substituting the sample size of the discovery and replication GWAS, this intercept test would

be significant if it implied that more than 5.2%, 3.7%, or 6.1% of the replication sample were found in the cohorts used in the discovery sample for DEP, NEUR, and SWB, respectively. We therefore believe that what follows may be interpreted as out-of-sample predictive power.

Using procedures described above, we calculated winner's-curse-adjusted estimates of the effect size of each lead SNP in an MTAG analysis (with the HRS omitted). Then, in each validation cohort and for each trait, for the set of lead SNPs, we regressed the effect sizes estimated in the validation cohort on the MTAG winner's-curse-adjusted effect sizes. In these regressions, we constrain the intercept to equal zero. The estimated slope coefficients are shown in **Supplementary Table 16.** The table separately reports the results for Add Health, HRS, and a sample-size weighted average of the two estimates. **Fig. 5** plots the pooled estimates. If the loci replicate successfully, then we expect to see a slope of one. In all cases, we cannot reject the null hypothesis that the slope is one (and we strongly reject the null hypothesis that the slope is zero), indicating that the lead SNPs taken as a whole replicate well for each trait.

6 Prediction

6.1 Introduction

This section provides additional details on the prediction analyses described in **Online Methods**. A major goal of our prediction analyses was to compare the predictive power of MTAG-based polygenic scores (PGSs), relative to conventional GWAS PGSs, to the theoretically expected gains (**Online Methods**). A secondary goal was to examine whether, in the context of our application, MTAG PGSs based on trait-specific association statistics (e.g., using MTAG association statistics for DEP to predict DEP) have greater predictive power than PGSs based on trait-specific associations for one of the other traits (e.g., using the MTAG association statistics for either NEUR or SWB to predict DEP). Below, we provide additional details on the analyses that underlie the results reported in the main text. We also describe the results from some additional robustness analyses.

6.2 SNP Selection for Polygenic-Score Construction

All comparisons of predictive power reported in this section and in the main text are based on PGSs constructed using an identical set of SNPs (but weighted using either GWAS-based or MTAG-based association statistics). Our main analyses are based on a subset of HapMap3 SNPs selected according to procedures described in **Online Methods**. See **Panel C** of **Supplementary Tables 10-13** for additional information about the process by which we arrived at the final list of SNPs. As in our replication analyses, the PGSs are constructed using weights from analyses that omit HRS from the discovery-stage analysis.

To probe robustness, we also analyzed PGSs based on a subset of the HapMap3 SNPs selected using an algorithm that ensures the final SNP list does not contain any two SNPs whose pairwise linkage disequilibrium exceeds $R^2 = 1$. Procedurally, we selected these SNPs using Plink's pruning algorithm [33]. We used the reference file described in section 3.2.3 to estimate linkage disequilibrium. In our application, the algorithm leaves 54,238 SNPs (see Panel D in **Supplementary Tables 10-13** for further details**).**

6.3 Phenotypes

Our prediction analyses are based on phenotypes defined using exactly as in our replication analyses in section 5.2.

6.4 Results

6.4.1 GWAS vs MTAG Polygenic Scores

Supplementary Tables 17 and **18** summarize the findings from our analyses of the predictive power of MTAG and GWAS PGSs based, respectively, on the full and pruned sets of HapMap3 SNPs (see **Online Methods** and the table captions for additional details). We report both cohort-level estimates and a combined estimate calculated by sample-size weighting the two cohort-level estimates. The numbers reported in **Supplementary Table 17** are the data underlying **Figs. 6a and 6b**. **Fig. 6c** compares the observed MTAG gains (relative to GWAS) to the theoretically predicted ones.

For DEP and SWB, the observed gains in predictive power are very close to the theoretical projections (**Online Methods**). For NEUR, the observed gain is smaller than predicted, but not significantly so. The theoretical projections are based on several simplifying assumptions, one of which is that the heritability of each trait is the same in the discovery and validation samples. In practice, this assumption is unlikely to hold exactly for NEUR, which is measured with a 4-item battery in our two validation cohorts and a 12-item battery in the UKB discovery sample (which makes up the bulk of the NEUR GWAS). The coarser measure of NEUR in the validation samples is likely to have lower re-test reliability and therefore also lower heritability. To quantify this potential bias, we constructed a 4-item measure of NEUR in UKB and calculated its re-test reliability. We selected the four items most similar to those included in the validation sample batteries. We found that the re-test reliability of the 12-item score is 0.822, compared to 0.734 for the 4-item score. A post hoc modification of the theoretical framework that assumes the NEUR measures used in our validation samples have 11% lower re-test reliability leads to a predicted gain that is very close to what we observe empirically.

6.4.2 Trait-Specific Association Statistics

Supplementary Table 19 reports estimates of the predictive power of MTAG PGSs based on either own-trait association statistics (e.g., MTAG-DEP weights are used to construct the PGS when the dependent variable is DEP) or other-trait association statistics (e.g., MTAG-NEUR or MTAG-SWB weights are used to construct the PGS). The results for the full set of HapMap3 SNPs are reported in **Supplementary Table 19**. Despite the very high genetic correlation between the traits considered in our application, own-trait PGSs always do better than either of the other-trait PGSs. When we meta-analyze the two cohort-level estimates, the increase is statistically significant in five out of six cases; the exception is that we cannot reject the null that the MTAG-SWB and MTAG-NEUR PGSs are equally predictive of NEUR. These results are displayed in **Figs. 6c and 6d**.

Based on our theoretical and empirical results, it is tempting to interpret the PGS for a trait based on MTAG summary statistics as uncontaminated by the traits with which it was analyzed. And indeed, if the homogeneous-$\mathbf{\Omega}$ assumption holds, such an interpretation would be appropriate. However, if the homogeneous-$\mathbf{\Omega}$ assumption is violated, then the resulting biases in the MTAG effect estimates can influence the interpretation of the PGS. Suppose, for example, that an investigator applies MTAG to GWAS summary statistics for two traits, A and B and would like to use the PGS for trait A to predict trait C is some other sample. Suppose—in violation of the homogeneous-$\mathbf{\Omega}$ assumption for the traits A and B being analyzed by MTAG—that there are a set of SNPs that are null for trait A but not for traits B and C. Then even if traits A and C are genetically uncorrelated, the PGS for trait A will in general be associated with trait C due to the bias in the MTAG effect-size estimates for those SNPs that are truly null for trait A. However, we conjecture that the biases in a trait's PGS become small when the GWAS for the trait is either well powered or comparably powered to the GWASs for the other traits in the MTAG analysis (just as the FDR is low in these cases).

6.4.3 MTAG vs. Naïve Meta-Analysis

The analyses in section 6.4.1 benchmarked the predictive power of MTAG-based PGSs against PGSs constructed using weights from a single-trait GWAS. In many realistic empirical settings, we believe the GWAS

benchmark is the most relevant one. Here we performed additional comparisons against alternative benchmarks that shed light on which features of MTAG are driving its gains relative to the GWAS-based PGSs.

To motivate the additional analyses, note that relative to simply meta-analyzing the various GWAS results together, MTAG differs by (i) adjusting for sample overlap and (ii) leveraging information about correlation in the SNPs' effects to optimally weight the GWAS results from different traits. Our primary goal in this section is to examine quantitatively the importance of each channel in the context of our specific application of MTAG to the three well-being phenotypes.

In some settings, it may be straightforward to conduct a meta-analysis of GWAS results from different phenotypes with weights derived under the simplifying assumption that $r_\beta = 1$. If the true r_β's are close to one, then a predictor based on such weights may perform comparably to an MTAG predictor. The analyses we report here thus serve the auxiliary purpose of allowing us to explore this conjecture.

As a preliminary, we reran our three single–trait GWAS meta-analyses in a set of non-overlapping cohorts. We "stacked the deck" against MTAG by retaining cohort-level summary statistics for the trait with higher SNP heritability whenever a cohort contributed to more than one of the single-trait GWASs. Applying this criterion left us with three GWAS meta-analyses.

1. DEP (N = 354,862; $N_{eff} = 253,247$) [Omitted cohort: UKB]
2. NEUR (N = 104,439) [Omitted cohort: GPC]
3. SWB (N = 85,944) [Omitted cohorts: UKB, 23andMe, HRS (which is omitted anyway due to being a validation cohort)]

Below, we refer to these as our "restricted" meta-analyses. We refer to the original meta-analyses as our "unrestricted" meta-analyses. As before, our two validation cohorts, HRS and Add Health, are excluded from all of the restricted and unrestricted meta-analyses to ensure that the prediction analyses are conducted in cohorts whose overlap with the discovery cohorts is minimal.

We subsequently compared the results of four MTAG analyses:

1. No Overlap ($r_\beta = 1$)
2. No Overlap ($r_\beta = \hat{r}_\beta$)
3. Overlap ($r_\beta = 1$)
4. Overlap ($r_\beta = \hat{r}_\beta$)

Analyses (1) and (2) are conducted using the restricted meta-analyses. Since (by construction) there is no known overlap between the cohorts included in the three underlying meta-analyses, we label these our "No Overlap" analyses. Analyses (3) and (4) are instead conducted using the unrestricted single-trait GWASs, which are conducted in overlapping samples.

In our "$r_\beta = 1$" analyses, we fix all pairwise genetic correlations to equal one; that is, we fix the off-diagonal terms in the $\hat{\mathbf{\Omega}}$ matrix such that the correlation of effect sizes between pairs of traits is one (i.e., the covariance between a pair of traits u and v, ω_{uv}, is set equal to $\sqrt{\omega_{uu}\omega_{vv}}$). However, we do not impose the restriction

that all traits have identical heritabilities. (Imposing that restriction would yield greater MTAG gains than those reported below.) In our "$r_\beta = \hat{r}_\beta$" analyses, we instead treat all elements of $\hat{\mathbf{\Omega}}$ as free parameters that are estimated, as is the default for MTAG. Analysis (4) is simply our standard MTAG analysis. In our "No Overlap" analyses, we use the restricted meta-analyses and restrict the off-diagonal entries of the $\hat{\mathbf{\Sigma}}_j$ matrix to equal zero for all SNPs. "Overlap" analyses were conducted on the full data with no restrictions on $\hat{\mathbf{\Sigma}}_j$. With the exception of the restrictions on $\hat{\mathbf{\Omega}}$ and $\hat{\mathbf{\Sigma}}_j$ listed above, MTAG analyses (1) through (3) were conducted following procedures identical to those in our full MTAG analyses.

As we explain below, differences between the results of these four MTAG analyses are informative about what factors generate MTAG gains in our specific application (and whether a naïve $r_\beta = 1$ analysis performs comparably to a full-fledged MTAG analysis in terms of prediction accuracy). To illustrate the logic, suppose that the predictive power of an MTAG-generated PGS based on the "Overlap ($r_\beta = 1$)" analysis was found to be near-identical to that of an MTAG-generated PGS constructed with weights from the "No Overlap ($r_\beta = 1$)" analysis. Suppose further that MTAG-generated PGSs from the remaining two analyses—"No Overlap ($r_\beta = \hat{r}_\beta$)" and "Overlap ($r_\beta = \hat{r}_\beta$)"—were also found to have similar predictive power to each other. Under these conditions, the appropriate inference would be that most of the observed MTAG gains are driven by channel (ii) and that the return to exploiting information about sample overlap is minimal. Of course, the relative importance of the two channels will vary across applications.

In our empirical application, we generated PGSs using the output from each of the four MTAG analyses. For MTAG analyses with $r_\beta = 1$, all weights are the same, and we generated a single PGS. For analyses with $r_\beta = \hat{r}_\beta$, we generate three PGSs: one based on each set of trait-specific association statistics. To maximize comparability, each PGS was based on an identical set of SNPs, defined as the intersection of HapMap3 SNPs and the set of SNPs passing recommended MTAG filters in each unrestricted and restricted meta-analysis. In total, there were 1,015,895 SNPs satisfying these conditions. We subsequently compared the predictive power of the four PGSs.

The results are summarized in **Supplementary Tables 20**. As in our main prediction analyses, we report the incremental R^2's of each PGS when it is added to a specification with baseline controls of age, sex, their interactions, and 10 principal components. Each reported R^2 is calculated as the sample-size weighted average in the two validation cohorts. We obtain 95% confidence intervals using the bootstrap with 1,000 iterations. For each trait, the gains from using the estimated r_β are always small relative to the apparent gains of allowing sample overlap. This finding is not surprising given the high pairwise genetic correlations between the traits in our application.

Nevertheless, MTAG PGSs from analyses with $r_\beta = \hat{r}_\beta$ are typically more predictive than analogous MTAG analyses with $r_\beta = 1$. **Supplementary Figure 8a** compares the predictive power of PGSs based on each of the four MTAG analyses. In five out of six cases, the $r_\beta = \hat{r}_\beta$ PGSs perform better than $r_\beta = 1$ PGSs, holding constant the set of cohorts included in the underlying single-trait GWASs used as input. **Supplementary Figure 8b** shows the differences in the predictive power of the PGSs rather than absolute level of predictive power. Overall, the results suggest that in our application, most observed MTAG gains appear to be come from the estimator's ability to accommodate summary statistics from (larger but) overlapping samples.

6.5 Concluding Discussion

We draw two main conclusions from the results reported in this section. First, the observed gains in prediction accuracy are consistent with our theoretical expectation based on the observed increase in the mean χ^2-statistic for each trait. Second, most MTAG gains observed in our specific application are explained by MTAG's ability to efficiently incorporate summary statistics from overlapping samples. Nevertheless, despite the very high pairwise genetic correlations in our specific application, the trait-specific association statistics are informative and contribute modestly to improving prediction accuracy.

7 Biological Annotation

7.1 Background

We used the bioinformatics tool DEPICT [34] (downloaded February 2016) to prioritize likely causal genes within loci defined by lead SNPs, to detect enrichment of gene sets (e.g., pathways), and to pinpoint the tissues of action. To determine the extent of additional biological information afforded by MTAG, we applied DEPICT twice to each trait, using as inputs both the single-trait GWAS results and the corresponding MTAG results. In both cases, we applied the default DEPICT P value threshold, 10^{-5}, for including SNPs. We identify the prioritized genes, enriched gene sets, and enriched tissues at the false discovery rate (FDR) threshold of 5%.

A gene is missing from the DEPICT inventory if it lacks high-quality Affymetrix expression data in the Gene Expression Omnibus. To obtain some coverage of these genes, we added any non-DEPICT protein-coding gene with a status of 'known' in GENCODE (downloaded February 26, 2015) to our results if it either encompasses one of the lead SNPs in a DEPICT-defined locus or has the transcription start site closest to such a SNP.

7.2 Results

The complete set of findings for each trait can be found in **Supplementary Tables 21** to **29**. Results based on MTAG coefficients are substantially more informative than results based on GWAS coefficients—identifying more prioritized genes, enriched gene sets, and enriched tissues—as summarized in **Table 1**, **Fig. 7**, and **Supplementary Figures 9** and **10**.

Some of this improved signal may be a result of using many more SNPs in the MTAG analyses, since many more SNPs pass the P value threshold of 10^{-5}. To verify that our results are not driven by this effect, for each trait, we find the P value threshold such that there are an equal number of clumped SNPs in the GWAS results below the threshold as there are clumped SNPs below the P value threshold of 10^{-5} in the MTAG results. We then run DEPICT on the GWAS results using the relaxed threshold corresponding to each trait. We find that even by this procedure, we still see an increase in signal of the MTAG-based results over the constant-number-of-SNPs GWAS results for DEP and SWB. For NEUR, the MTAG results find nearly identical amounts of enrichment to the GWAS results with a relaxed P value. This comparison can be found in **Supplementary Table 30**.

Because the DEPICT results from the three traits are overlapping (and correlated with each other), we focus our discussion on the results for just one of the traits, DEP. Although we identify 347 enriched gene sets for DEP, many of the significantly enriched gene sets are highly correlated and thus do not represent independent biology. To facilitate the interpretation of the gene-set results, we applied the Affinity Propagation Algorithm [35] to segregate the gene sets into clusters. The algorithm names each cluster after an exemplary member. The input to the algorithm consists of the correlations between gene sets. We include in the gene sets only those genes prioritized by DEPICT (i.e., genes with FDR < 0.05). These results can be found in **Fig. 7b**.

Genes whose products are involved in synaptic signaling between neurons are strongly over-represented in our results, as can be seen in many of the clusters: 'synapse,' 'synapse assembly,' 'regulation of synaptic transmission,' and 'regulation of postsynaptic membrane potential' (**Fig 7b**). The genes *PCLO* (piccolo),

BSN (bassoon), and *SNAP25* encode components of the presynaptic matrix that docks neurotransmitter-loaded vesicles close to voltage-gated calcium channels and to the release machinery [36]. The subunit of the voltage-gated calcium channel encoded by *CACNA1E* can trigger the release of neurotransmitter, although its role is still not well understood [37, 38]. Once the neurotransmitter glutamate is released into the synaptic cleft, it may bind to a variety of ionotropic and metabotropic receptors that are also encoded by prioritized genes (*GRIA1, GRIK3*, *GRM1*, *GRM5*, *GRM8*). Whereas *GRIA1* is a subunit of an AMPA-type receptor responsible for the immediate response of the postsynaptic neuron to arriving input, the other receptors affect the release rates and spiking dynamics of pre- and postsynaptic neurons on a longer timescale. This longer timescale may suggest the importance of learning mechanisms in some of the genetic factors underlying DEP. That hypothesis is reinforced by some of the enriched gene set results, which are defined by altered reactions to stress and novelty in mice: 'decreased exploration in a new environment,' 'abnormal contextual conditioning behavior,' 'increased anxiety-related response,' and 'behavioral fear response' (**Supplementary Table 22**).

7.3 Relationship to Earlier Work

It is interesting that our results implicate glutamate-based transmission so prominently (e.g., the gene set 'extracellular-glutamate-gated ion channel activity'), since hypotheses regarding major depression and related traits have tended to focus on modulatory monoamine transmitters [39, 40]. We do identify one gene encoding a subtype of the dopamine receptor, *DRD2*, which previous studies have linked to the genetically correlated trait of schizophrenia [41].

A small study found lower levels of the receptor *GRM5* in the brains of patients diagnosed with major depression [42]. At least three of our other DEPICT-prioritized genes—*SLCA6A15* [43, 10], *NEGR1*[10], and *SORCS3* [10]—have also been implicated in previous genetic studies of major depression, although our data overlaps that of [10].

8 Comparison to Other Multi-trait Methods

8.1 Introduction

This section compares MTAG to several other multi-trait methods. We limit the comparison to methods that can be applied in the specific setting for which MTAG was developed: the only available inputs are summary statistics from an arbitrary number of genome-wide analyses conducted in samples with unknown overlap. We identified three methods satisfying these criteria: a multi-trait method proposed by Bolormaa *et al.* [44] and the two CPASSOC ("cross-phenotype association by using summary statistic") methods, S_{Hom} and S_{Het}, proposed by Zhu *et al.* [45]. In what follows, we refer to the three methods as Bolormaa, S_{Hom} and S_{Het}.

MTAG produces trait-specific effect estimates for each SNP, whereas the alternative methods all test the joint null hypothesis that a SNP is not associated with any of the traits. Thus, in settings where the purpose of the multi-trait analysis is to test for association between a SNP and a *single* trait or to improve the prediction accuracy of a polygenic score, the alternative methods are not readily applicable. In such settings, a more natural benchmark for the MTAG results are the single-trait GWASs in section 3. For comparability with the other multi-trait methods, all comparative analyses reported here focus on testing the joint null hypothesis that a SNP is not associated with any of the traits.

The section is organized as follows. We begin in section 8.2 with a summary overview of the three methods, highlighting some key distinguishing theoretical features of each estimator. An important upshot from this discussion is that, unlike the other methods, MTAG distinguishes between the genetic and estimation-error sources of correlation between test statistics from the single-trait GWASs. Failure to draw this distinction can cause the other estimators to lose power relative to MTAG. section 8.3 corroborates this prediction in simulated data.

In section 8.2, we compare MTAG to the other estimators using summary statistics from published GWAS results. In our first such empirical application, we performed Bolormaa, S_{Hom}, and S_{Het} analyses on summary statistics from the three single-trait GWASs (DEP, NEUR, and SWB) described in section 8.3 and contrasted the results to those produced by MTAG. In our second application, we applied MTAG and each of the three alternative methods to summary statistics from sex-stratified genome-wide analyses of height, body mass index (BMI), and waste-to-hip-ratio adjusted for BMI (WHRadjBMI) conducted by the GIANT consortium[46]. In both the application to SSGAC and GIANT summary statistics, we found that MTAG compares favorably to the other methods in terms of the number of loci identified.

8.2 Theoretical Discussion

8.2.1 Bolormaa's Approach

Bolormaa *et al.* [44] motivate their method with a modified version of what they ultimately propose. In this modified version, for each SNP, they consider calculating $\boldsymbol{t}'\boldsymbol{D}^{-1}\boldsymbol{t}$, where $\boldsymbol{t}$ is the length-T vector of GWAS z-statistics (one per trait) and $\boldsymbol{D}$ is the variance-covariance matrix of these test statistics across traits, which equals $\boldsymbol{D} = \boldsymbol{\Omega} + \boldsymbol{\Sigma}_j$ in our notation. $\boldsymbol{D}$ is assumed to be homogeneous across SNPs, which is analogous to the MTAG assumption that $\boldsymbol{\Omega}$ is homogeneous across SNPs (formally, it would follow from both $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}_j$

being homogeneous). The null hypothesis is that the SNP has no genetic effect on any of the traits, i.e., $\boldsymbol{\Omega} = \mathbf{0}$. Under that null hypothesis, $\boldsymbol{t}'\boldsymbol{D}^{-1}\boldsymbol{t}$ is asymptotically distributed χ^2 with T degrees of freedom. In the derivation of this hypothesis test, there is no need to distinguish between $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}_j$ because under the null hypothesis, $\boldsymbol{D} = \boldsymbol{\Sigma}_j$ only.

However, in order to apply this method, $\boldsymbol{D}$ would need to be estimated. If the observed variance-covariance matrix of the GWAS test statistics is used, however, then $\hat{\boldsymbol{D}}$ is an estimate of $\boldsymbol{\Omega} + \boldsymbol{\Sigma}_j$. Consequently, under the alternative hypothesis that $\boldsymbol{\Omega} > \mathbf{0}$, the hypothesis test based on $\boldsymbol{t}'\hat{\boldsymbol{D}}^{-1}\boldsymbol{t}$ would be too conservative. This can be seen most straightforwardly in the case of a single trait, where the test statistic is simply t^2/d. Here d should equal the scalar 1, which is the variance of the z-statistic under the null hypothesis of no association. However, it is instead estimated to be the observed variance of the z-statistics across SNPs, $\hat{d}$. If the null is false—i.e., there is true genetic signal—then $\hat{d} > 1$, and the test statistic $t^2/\hat{d}$ will always be smaller than the correct test statistic, t^2. Thus, in the case of a single trait, the test statistic is biased downward and hence the test is too conservative.[19] (This is analogous to applying genomic control to adjust GWAS results. Doing so deflates the GWAS test statistics by a measure of the total amount of inflation in the test statistics, which is an over-correction if true genetic signal contributes to the inflation. Instead, adjusting GWAS results using the LD score intercept aims to adjust only for the component of the inflation in the GWAS test statistics that is due to bias, rather than the component that is due to true genetic signal.)

Recognizing this issue, Bolormaa *et al.* [44] conduct a hypothesis test based on $\boldsymbol{t}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{t}$, where $\hat{\boldsymbol{V}}$ is the observed *correlation* matrix between the GWAS test statistics (rather than the observed variance-covariance matrix $\hat{\boldsymbol{D}}$). Doing so ensures that the diagonal elements of $\hat{\boldsymbol{V}}$ are equal to 1, as the diagonal elements of $\boldsymbol{D}$ would be under the null hypothesis. However, the off-diagonal elements of $\hat{\boldsymbol{V}}$ are biased estimates of the off-diagonal elements of $\boldsymbol{D}$ under the null hypothesis, and therefore $\boldsymbol{t}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{t}$ does not follow a χ^2 distribution under the null hypothesis. For these reasons, [44] refer to their method as an "approximate analysis."[20]

8.2.2 CPASSOC

For brevity and simplicity, we focus most of the discussion here on S_{Hom}, but the same conclusions carry over to S_{Het}. While S_{Hom} is a more complicated test statistic than the Bolormaa test statistic, the basic issues are the same. Indeed, as Zhu *et al.* [45] point out, S_{Hom} is in the same family of tests as Bolormaa, but it is the most powerful test when it is assumed that, under the alternative hypothesis, the effects of the SNP on each trait is equal. Using the notation from Bolormaa, $\boldsymbol{t}$ is the length-T vector of GWAS test statistics and $\boldsymbol{D}$ is the variance-covariance matrix of these test statistics across traits. $\boldsymbol{D}$ is assumed to be homogeneous across SNPs. For each SNP, the test statistic is

$$\frac{\mathbf{1}'\left(\boldsymbol{D}\boldsymbol{W}\right)^{-1}\boldsymbol{t}\left[\mathbf{1}'\left(\boldsymbol{D}\boldsymbol{W}\right)^{-1}\boldsymbol{t}\right]'}{\mathbf{1}'\left(\boldsymbol{W}\boldsymbol{D}\boldsymbol{W}\right)^{-1}\mathbf{1}}$$

[19] For a general proof for an arbitrary number of traits, we use the fact that if $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{A} - \boldsymbol{B}$ are all positive definite matrices, then $\boldsymbol{B}^{-1} - \boldsymbol{A}^{-1}$ is also positive definite (see fact 9(a) on p. 8.10 in [47]). Define $\boldsymbol{A} = \hat{\boldsymbol{D}} = \boldsymbol{\Omega} + \boldsymbol{\Sigma}_j$ and $\boldsymbol{B} = \boldsymbol{\Sigma}_j$. Then, $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{A} - \boldsymbol{B} = \boldsymbol{\Omega}$ are all positive definite. Therefore $\boldsymbol{\Sigma}_j^{-1} - \hat{\boldsymbol{D}}^{-1}$ is positive definite. By the properties of positive definite matrices, for any vector $\boldsymbol{t}$, $\boldsymbol{t}'\left(\boldsymbol{\Sigma}_j^{-1} - \hat{\boldsymbol{D}}^{-1}\right)\boldsymbol{t} > 0$ and therefore $\boldsymbol{t}'\boldsymbol{\Sigma}_j^{-1}\boldsymbol{t} > \boldsymbol{t}'\hat{\boldsymbol{D}}^{-1}\boldsymbol{t}$. Therefore, the test statistic is inefficient under the alternative hypothesis that $\boldsymbol{\Omega}$ is non-zero.

[20] An alternative way to pursue the $\boldsymbol{t}'\boldsymbol{D}^{-1}\boldsymbol{t}$ test statistic (or the CPASSOC test statistics discussed below) would be to estimate $\hat{\boldsymbol{\Sigma}}_j$ using LD score regressions (as in MTAG) and then conduct the χ^2-test with $\hat{\boldsymbol{\Sigma}}_j$ in place of $\boldsymbol{D}$. Developing this idea is beyond the scope of this paper.

where $\mathbf{1}$ is a matrix of ones and $\boldsymbol{W}$ is a diagonal matrix whose elements are the square root of the sample size of each single-trait GWAS included in the multivariate analysis. Under the null hypothesis that $\boldsymbol{\Omega} = \mathbf{0}$, this test statistic is asymptotically distributed χ^2 with 1 degree of freedom.

Just like with the Bolormaa test statistic, the S_{Hom} test statistic is implemented by replacing $\boldsymbol{D}$ by $\hat{\boldsymbol{V}}$, where $\hat{\boldsymbol{V}}$ is the observed *correlation* matrix of the GWAS test statistics. The potential problems are the same as in Bolormaa: the off-diagonal elements of $\hat{\boldsymbol{V}}$ are biased estimates of what the off-diagonal elements of $\boldsymbol{D}$ would be under the null and therefore, the test statistic no longer follows a χ^2 distribution under the null.

In an attempt to reduce the impact of true genetic variance and covariance (i.e., $\boldsymbol{\Omega}$) on $\hat{\boldsymbol{V}}$, when applying S_{Hom} to real data from the GIANT consortium, Park *et al.*[48] estimated $\hat{\boldsymbol{V}}$ in a subset of approximately independent SNPs for which the absolute value of the test statistic does not exceed 1.96 for any of the SNPs. This adjustment eliminates any influence of true genetic variance and covariance if among the SNPs with test statistic less than or equal to 1.96, there is no true genetic signal.

S_{Het} builds on S_{Hom} in two ways. First, each element of $\boldsymbol{W}$ is multiplied by the sign of the test statistic from the single-trait GWAS. An alternative, equivalent implementation is to switch $\boldsymbol{t}$ from being the vector of GWAS test statistics to the vector of their absolute values. Second, the multi-trait test statistic is evaluated many times, each time restricting the set of SNPs used to those with (absolute value of) GWAS test statistics exceeding some threshold τ. The value of the S_{Het} test statistic equals the maximum value over all $\tau > 0$. The S_{Het} test is computationally intensive, largely because the distribution of S_{Het} under the null hypothesis must be simulated. S_{Het} inherits from S_{Hom} the potential problems from using the correlation matrix rather than the variance-covariance matrix and using the observed correlation matrix of the test statistics rather than just the component due to estimation error. However, because S_{Het} uses the absolute values of the test statistics, it can be better powered than S_{Hom} or MTAG to detect SNPs that have effects on the traits that are opposite in direction.

8.2.3 Comparison to MTAG

MTAG produces trait-specific association statistics, whereas the other multi-trait methods are tests of the null hypothesis of no association with any trait. In order to adapt MTAG to test this null hypothesis, we use the trait-specific MTAG P values to construct a classical Bonferroni test of the joint null hypothesis. For example, in the case of three traits, we use $P = 3 \times \min\{p_1, p_2, p_3\}$, where p_1, p_2, and p_3 are the P values of the MTAG output for each trait. Since the MTAG-generated P values are correlated across traits, this Bonferroni test is a conservative test that disadvantages MTAG relative to the other methods (each of which produce a single P value based on an omnibus test statistic). For all of the methods, we test the joint null hypothesis at the genome-wide significance P value threshold.

8.2.4 Other Related Methods

Two other multi-trait methods are closely related to MTAG, MTGBLUP [49] and PleioPred [50], but we did not include them in the comparative analyses that follow, for reasons we explain here. The three methods all use a similar random-effects framework to boost the power to estimate SNP effects on each of the traits. In contrast to MTAG's GMM framework, however, MTGBLUP and PleioPred use Bayesian frameworks that

rely on distributional assumptions. MTGBLUP is built on a normally distributed prior distribution for effect sizes, and PleioPred is built on a spike-and-slab prior.

An important difference from MTAG is that MTGBLUP and PleioPred are designed primarily to improve polygenic prediction, and therefore neither report SNP-level effect sizes or P values. Since neither method produces P values, we are unable to compare MTAG to these other methods using the same approach described below.

Additionally, MTGBLUP requires individual-level data, which is not the setting for which MTAG was designed. Although PleioPred operates on GWAS summary statistics, it only can analyze two traits at a time and is too computationally burdensome (approximately 2 hours per replication) to compare its performance to MTAG in a simulation setting. [50] compare MTAG to PleioPred in terms of polygenic prediction in three two-trait empirical settings: Crohn's disease and Ulcerative Colitis; celiac disease and Ulcerative Colitis; and type-II diabetes with coronary artery disease. They find that PleioPred outperforms MTAG as measured by AUC, presumably because in those settings, the true effect-size distribution is well approximated by the spike-and-slab distribution assumed by PleioPred.

8.3 Comparative Analysis: Simulation Evidence

The theoretical discussion in section 8.2 highlighted that Bolormaa, S_{Hom}, and S_{Het} do not distinguish as cleanly as MTAG does between the component of the variance-covariance matrix of GWAS test statistics that is due to true genetic effects versus the component that is due to estimation error. This effect may reduce the power of these multi-trait methods relative to MTAG when the genetic component is non-negligible. On the other hand, the Bonferroni adjustment that we apply to the trait-specific MTAG results reduces the power of MTAG relative to the other methods. Moreover, as noted above, S_{Het} can have more power than MTAG in cases where a SNP has opposite effects on the traits. To assess the net result of the combination of these effects, we conduct simulations that compare MTAG with Bolormaa, S_{Hom}, and S_{Het}.

We perform 1000 replications each of the following set of simulations. Following the data-generating procedure outlined in section 2.2, we produce summary statistics for two traits where either (1) there is no effect-size nor estimation-error correlation ($r_\beta = r_\varepsilon = 0$), (2) there is high estimation-error correlation but no effect-size correlation ($r_\beta = 0$, $r_\varepsilon = 0.7$), or (3) there is high effect-size correlation but no estimation-error correlation ($r_\beta = 0.7$, $r_\varepsilon = 0$). To vary the strength of the true genetic effects, for each of these correlation profiles, we produce summary statistics allowing the mean χ^2-statistic be some value in $\chi^2_t \in \{1.1, 1.4, 2.0\}$ for each trait t. This gives us nine simulations for each correlation profile.

Because the potential inefficiency of the other methods relative to MTAG comes from implicitly assuming that $\mathbf{\Omega} = \mathbf{0}$, we might expect the relative performance of MTAG over the alternatives to be better as $\chi^2_1 = \chi^2_2$ increases. However, the degree of conservativeness of the MTAG Bonferroni test depends on the effective number of traits. The effective number of traits varies with r_β, and it also varies with χ^2_t whenever $r_\varepsilon \neq 0$. Therefore, the cleanest test of this claim is in the setting where $r_\beta = 0.7$ and $r_\varepsilon = 0$ since in that setting, the effective number of tests is held constant as the mean χ^2-statistic varies.

The results of these simulations are found in **Supplementary Figure 11**. In the scenarios where $r_\beta \neq 0$ or $r_\varepsilon \neq 0$, MTAG almost always finds more hits than the alternative methods. This is especially striking given the conservativeness of the Bonferroni adjustment used to obtain the MTAG P value. In the simulations

with $r_\beta = 0.7$ and $r_\varepsilon = 0$, as expected, the relative improvement from MTAG over the other methods is increasing as the power of each GWAS grows.

8.4 Comparative Analysis: Real Summary Statistics

We next to turn to empirical applications that use real genotype and phenotype data. In our first application, we performed Bolormaa, S_{Hom}, and S_{Het} analyses on summary statistics from the three single-trait GWASs (DEP, NEUR, and SWB) described in Supplementary Note section 3 and contrasted the results to those produced by MTAG. In our second comparison, we applied MTAG and each of the three alternative methods to summary statistics from sex-stratified genome-wide analyses of height, body mass index (BMI), and waste-to-hip-ratio adjusted for BMI (WHRadjBMI) from the GIANT consortium. The six files with the GIANT summary statistics (three phenotypes; two sexes) were examined in a previous study using the CPASSOC estimators to identify new loci for anthropometric traits [48].

Our CPASSOC analyses were conducted using publicly available software accessed on May 10, 2017, from http://hal.case.edu/zhu-web/. As described in section 8.2.2, the S_{Hom} and S_{Het} test statistics for a given SNP can be calculated from two inputs: a vector of z-statistics for the SNP's association with each trait and a variance-covariance matrix of z-statistics. As recommended by Park *et al.* and Zhu *et al.* [45, 48], we estimated the variance-covariance matrix $\boldsymbol{V}$ using a subset of approximately independent SNPs and omitting SNPs whose z-statistics exceeded (in absolute value) 1.96. To identify the set of approximately independent SNPs, we used the reference sample described in section 3.2.3 to estimate linkage disequilibrium between SNPs. We subsequently applied our standard clumping algorithm with no P value threshold, but to avoid selecting the most significant SNP in each clump, we clumped our results after falsely setting the P values of all SNPs to be the same value (we arbitrarily picked the value 0.5).

For comparability with MTAG results, we un-did the genomic control adjustment that had been applied to the anthropometric summary statistics before applying any CPASSOC analyses (failure to do so would complicate the comparison to MTAG). Procedurally, we do this by estimating a LD score regression [1] for each trait and subsequently calculating adjusted z-statistics. Each adjusted z-statistic is defined as the original z-statistic multiplied by the inverse of the square root of the estimated intercept. As expected, the LD score intercepts for the GWAS results from all GIANT traits are substantially below 1, so the adjustment increases the power of the alternative methods relative to MTAG.

For each of our two classes of traits—anthropometric or well-being—we used our standard locus definition and clumping algorithm to identify the number of genome-wide significant loci identified by each method. As described in section 8.2.3, since MTAG outputs trait-specific P values for each SNP, we conservatively defined the MTAG P value for SNP j as the Bonferroni-corrected minimum P value produced by MTAG across the T traits for SNP j.

Below, we describe these analyses and summarize the key findings.

8.4.1 Well-Being Traits

For purposes of comparability, we restricted the Bolormaa, S_{Hom}, and S_{Het} analyses to the same ~6.1M SNPs that passed recommended filters applied in our main MTAG analysis. Applying our standard locus definition, S_{Hom} identified 2 jointly associated loci at genome-wide significance (5×10^{-8}), S_{Het} identified 34

loci, and Bolormaa identified 43 loci. By comparison, MTAG yielded 67 independent loci(using the P value that Bonferroni-adjusts for 3 traits).

8.4.2 Anthropometric Traits

In our second application, we instead considered a setting in which the S_{Hom} and S_{Het} estimators have been shown in prior work to deliver gains relative to single-trait GWAS results. Park *et al.* [48] used CPASSOC on publicly available summary statistics from the GIANT consortium's sex-stratified analyses of height, BMI, and WHRadjBMI. Park *et al.*'s [48] S_{Hom} and S_{Het} analyses of six publicly available results files with summary statistics (3 phenotypes; 2 sexes) yielded seven loci that had not previously reached genome-wide significance in either sex-stratified or pooled meta-analysis of any of the three traits (see their Table 3).

For our comparative analyses, we applied our recommended MTAG filters to all cohort-level results files. These restrictions left us with ~2.29M SNPs. We subsequently applied the Bolormaa, S_{Hom}, and S_{Het} methods to this set of SNPs, and we compared the results to MTAG. Applying our clumping algorithm, we found that S_{Hom} identified 86 loci at genome-wide significance, S_{Het} identified 240 loci, and Bolormaa identified 161 loci. By comparison, our MTAG analysis identified 264 approximately independent loci (using the P value that Bonferroni-adjusts for 6 traits).

(The number of loci we report for S_{Hom} and S_{Het} (86 and 240) exceeds the numbers reported in the Park *et al.* [48] analysis of these data (55 and 129). Our analysis differs from that of Park *et al.* [48] in three ways. First, we used our recommended MTAG SNP filters to select the set of SNPs. Second, our reference panel (see section 3.2.3) is different than that of Park *et al.* [48] (who use the ARIC sample), and we therefore do not use the same set of approximately independent SNPs to estimate $\hat{\boldsymbol{V}}$. We do, however, follow the same procedures to select the set of approximately independent SNPs. Third, we use estimated intercepts from LD score regressions to "undo" the genomic control applied to the original GIANT summary statistics before applying any of the mult-trait methods. In practice, the fact that we identify more loci than Park *et al.* [48] using S_{Hom} and S_{Het} is driven entirely by this third difference. We verified that without the adjustment, our results are nearly identical to those reported by Park *et al.* [48].)

8.4.3 Comparative Results for Well-Being and Anthropometric Traits

Overall, MTAG appears to be better powered to detect associations in these two applications, despite our use of what is effectively a more stringent significance threshold for the MTAG analyses.

Supplementary Figures 12 show "inverted Manhattan plots" with pairwise comparisons of MTAG to alternative estimators applied to the anthropometric summary statistics; see **Supplementary Figures 13** for analogous comparisons for the well-being analyses. In each figure, MTAG P values (on a $-\log_{10}$ scale) are plotted below the x-axis and P values from the other estimator (Bolormaa, S_{Hom}, or S_{Het}) above the x-axis. Lead SNPs identified by one method but not the other are labeled. For example, if a SNP reaches genome-wide significance in the MTAG analysis, it is only labeled in the pairwise comparison with Bolormaa if no SNP from the locus (as defined in **Online Methods**) reached genome-wide significance in the Bolormaa analysis. Conversely, above the x-axis, we only mark lead SNPs that are in loci that evaded detection by MTAG but were detected by the alternative method.

In the anthropometric application, MTAG identified 264 loci. Of these, 102 went undetected by Bolormaa, 43 went undetected by S_{Het}, and 181 went undetected by S_{Hom}, respectively. By contrast, MTAG missed 9 loci identified by Bolormaa, 20 loci identified by S_{Het}, and 12 identified by S_{Hom}. In the application to well-being traits, analogous comparisons appear to favor MTAG even more strongly. Of the 67 loci identified by MTAG, 29 evaded detection by Bolormaa, 19 evaded detection by S_{Het} and 66 evaded detection by S_{Hom}. In the other direction, MTAG missed only 8 of 43 significant hits found by Bolormaa, 1 of 2 found by S_{Hom}, and 14 of 68 found by S_{Het}.

We note that in these comparative analyses, we took several steps to ensure that any observed increase in the number of MTAG-associated loci, relative to the number of loci identified by the other methods, is likely to reflect real gains in statistical power. For example, we analyzed a fixed set of SNPs in each comparative analysis, and we used a conservative multiple-hypothesis adjustment to obtain an omnibus MTAG P value for each SNP.

8.4.4 Replication for Anthropometric Results

As an additional robustness test, we assessed replication rates of the loci identified by each estimator in our analyses of the GIANT phenotypes. These analyses exploit the fact that following the publication of the Park *et al.* [48], the GIANT consortium has published follow-up analyses of all three anthropometric traits based on substantially increased samples[51, 52, 53].

In these analyses, we use the summary statistics that have been corrected using the intercepts from LD score regressions as described in section 8.4.2 above. We then meta-analyzed the sex-stratified results files for each trait, without applying genomic control, to obtain three files with summary statistics: one for each trait. We similarly calculated adjusted z-statistics for each trait using publicly available summary statistics of the follow-up studies that have subsequently been published by the GIANT consortium[51, 52, 53]. We calculated an "incremental z-statistic" for each SNP and obtained a replication P value for each SNP using this incremental z-statistic. Omitting SNP and phenotype subscripts, the incremental z-statistic is defined [16] implicitly by $z_{\text{pooled}} = \left(\sqrt{\frac{N_1}{N}}Z_1 + \sqrt{\frac{N_2}{N}}Z_{\text{incr}}\right)$, where N_1 is the size of the (meta-analyzed) discovery sample in the first-stage analysis and N_2 is the size of the (meta-analyzed) discovery sample in the augmented, follow-up analysis. We used these incremental z-statistics and the P values associated with them to determine whether each association replicated. We defined the replication P value as the minimum P value across the three traits for each SNP multiplied by 3.

Results are presented in **Supplementary Table 31**. In the upper panel, we report the overall fraction of loci that replicated at three different significance thresholds: genome-wide significance (5×10^{-8}), 10^{-4}, and 10^{-3}. At genome-wide significance, 41% of MTAG-associated loci satisfied our replication criterion. This rate is comparable to that observed for S_{Het} (43%), though lower than those of Bolormaa (55%) and S_{Hom} (63%). The lower overall replication rate for MTAG and S_{Het}, relative to Bolormaa and S_{Hom}, is not surprising. MTAG and S_{Het} identified substantially more genome-wide significant loci (264 and 240, respectively) than Bolormaa (161) and S_{Hom} (86). Since the incremental loci identified by MTAG and/or S_{Het} are likely to have smaller effect sizes, the statistical power to replicate them should be lower. Indeed, of the top 86 hits for each method (the number of hits found by S_{Hom}), 81% of them replicate for MTAG, 80% for S_{Het}, and 79% for Bolormaa. Of the top 161 SNPs (the number found by Bolormaa), 57% of the MTAG hits replicate and 56% of the S_{Het}. Even among the 240 SNPs (the number found by S_{Het}), 44%

of the MTAG hits replicate. So in a fair comparison holding the number of SNPs constant, MTAG always replicates at higher rates than that alternative methods, though only marginally so.

The second panel therefore compares replication rates (at genome-wide significance) for the following subsets of loci: (i) those identified by MTAG and missed by the alternative method and (ii) loci detected by the alternative method and missed by MTAG. Of the 102 loci identified by MTAG but missed by Bolormaa, 17% replicated at genome-wide significance; of the 43 loci identified by MTAG but missed by S_{Het}, 12% replicated; and of the 181 loci identified by MTAG but missed by S_{Hom}, 29% replicated. None of the loci identified by the other methods but missed by MTAG replicated at genome-wide significance. A similar pattern is evident at weaker significance thresholds. From this panel, it appears that replication rates for distinct loci identified by MTAG are stronger than those for distinct loci identified by any of the other three methods. (We note that while the comparisons of replication rates is informative, it is difficult to interpret the replication-rate *levels* without a calculation of the expected replication rate, given the statistical power to replicate each locus.)

8.5 Conclusion

Both the results of our simulations and of our empirical applications suggest that MTAG has greater power to detect associations than the other multi-trait methods that we examine. Our theoretical analysis suggests that the main reason is that relative to MTAG, the other methods lose efficiency by not distinguishing between variance and covariance in the GWAS test statistics that is due to true genetic signal versus estimation error. We caution that our conclusions here almost surely do not generalize to all possible simulations and empirical applications. For example, because S_{Het} gains power in cases where SNPs have opposite effects on different traits, it is plausible that it would outperform MTAG in settings where such SNPs are common or where the scientific objective is to identify SNPs with effects whose signs are opposite to the direction that would be predicted from the overall genetic correlation.

9 Alternative MTAG Frameworks

9.1 Maximum Marginal Likelihood

In the original version of this paper posted on bioRxiv[21], the MTAG estimator was derived using a maximum likelihood (ML) framework for both $\hat{\beta}_{\text{MTAG},j,t}$ and $\hat{\mathbf{\Omega}}$. While the MTAG estimating equation for $\hat{\beta}_{\text{MTAG},j,t}$ is identical across the two frameworks, the ML derivation assumes that the true genetic effects are drawn from a normal distribution, while the GMM derivation does not require any particular distributional assumptions and in that sense is more general. The method-of-moments estimator for $\hat{\mathbf{\Omega}}$ is slightly different than the ML-based version, but the method-of-moments estimator again relies on weaker assumptions. Moreover, because the method-of-moments estimator for $\hat{\mathbf{\Omega}}$ has a closed-form solution, it is substantially faster to calculate than the ML estimator, which requires numerical optimization. In this section, we present the ML framework as a supplement to the GMM framework that is now emphasized in the paper.

9.1.1 The Model

The statistical model for the ML version of MTAG is identical to that described in section 1.2.1, except that we assume that the effect sizes, $\boldsymbol{\beta}_j$, are drawn independently across j from a multivariate normal distribution, such that

$$\boldsymbol{\beta}_j \sim N\left(\mathbf{0}, \mathbf{\Omega}\right)$$

for some variance-covariance matrix $\mathbf{\Omega}$ that is the same for all SNPs j.

We would like to produce an estimate of the effect of each SNP j on each trait t. Using the distributional assumption above, we construct a likelihood function that we can maximize to calculate an estimate of $\beta_{j,t}$. Since we only estimate the effect of the SNP on one trait at a time, effectively integrating out the effect of the SNP on other traits, the likelihood function that we will construct is called a "marginal likelihood function," and therefore MTAG is a maximum-marginal-likelihood estimator [54].

To construct the marginal likelihood function, we begin with the distribution of the random effects $\boldsymbol{\beta}_j$ given the true effect of SNP j on trait t, denoted $\beta_{j,t}$. By the properties of multivariate normal distributions [55] we have:

$$\left(\boldsymbol{\beta}_j \,|\, \beta_{j,t}\right) \sim N\left(\frac{\boldsymbol{\omega}_t}{\omega_{tt}}\beta_{j,t},\, \mathbf{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}\right),$$

where (consistent with the notation in section 1) $\boldsymbol{\omega}_t$ is a vector equal to the t^{th} column of $\mathbf{\Omega}$ and ω_{tt} is a scalar equal to the t^{th} element of $\boldsymbol{\omega}_t$ (or equivalently, the t^{th} diagonal element of $\mathbf{\Omega}$). It follows that

$$\left(\hat{\boldsymbol{\beta}}_j \,|\, \beta_{j,t}\right) = \left(\boldsymbol{\beta}_j + \boldsymbol{\varepsilon}_j \,|\, \beta_{j,t}\right) \sim N\left(\frac{\boldsymbol{\omega}_t}{\omega_{tt}}\beta_{j,t},\, \mathbf{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \mathbf{\Sigma}_j\right). \tag{26}$$

[21] Turley, P. et al. MTAG: Multi-Trait Analysis of GWAS. bioRxiv (2017).

We can re-express the marginal likelihood function as

$$\begin{aligned} L\left(\beta_{j,t};\hat{\boldsymbol{\beta}}_j\right) &\propto \exp\left\{-\frac{1}{2}\left(\hat{\boldsymbol{\beta}}_j-\frac{\boldsymbol{\omega}_t}{\omega_{tt}}\beta_{j,t}\right)'\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}\left(\hat{\boldsymbol{\beta}}_j-\frac{\boldsymbol{\omega}_k}{\omega_{tt}}\beta_{j,t}\right)\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left[\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}\beta_{j,t}^2\right.\right. \\ &\qquad \left.\left.-2\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1-1}\hat{\boldsymbol{\beta}}_j\beta_{j,t}\right]\right\}, \end{aligned}$$

where the operator $\propto$ denotes that the next line is proportional to the previous line with respect to $\beta_{j,k}$. For notational convenience, we define

$$\begin{aligned} \theta &\equiv \frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}\hat{\boldsymbol{\beta}}_j \\ \gamma &\equiv \frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}. \end{aligned}$$

In terms of this notation,

$$\begin{aligned} L\left(\beta_{j,t};\boldsymbol{Z}_j\right) &\propto \exp\left\{-\frac{1}{2}\left[\gamma\beta_{j,t}^2-2\theta\beta_{j,t}\right]\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left[\gamma\beta_{j,t}^2-2\theta\beta_{j,t}+\frac{\theta^2}{\gamma}\right]\right\} \\ &\propto \exp\left\{-\frac{1}{2}\frac{\left(\beta_{j,t}-\frac{\theta}{\gamma}\right)^2}{\frac{1}{\gamma}}\right\}. \end{aligned}$$

Note that this expression is the kernel for a normal distribution with mean

$$\frac{\theta}{\gamma}=\frac{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}}{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}}\hat{\boldsymbol{\beta}}_j.$$

Since a normal pdf is maximized at its mean, the MTAG estimate of the effect of the SNP on trait t is

$$\hat{\beta}_{\text{MLE},j,t}=\frac{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}}{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}}\hat{\boldsymbol{\beta}}_j. \tag{27}$$

Since this is the same estimator as the GMM version of MTAG, the standard error is the same:

$$\text{SE}\left(\hat{\beta}_{\text{MLE},j,t}\right)=\frac{1}{\sqrt{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega}-\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}}+\boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}}}. \tag{28}$$

9.1.2 Estimating $\boldsymbol{\Omega}$ by Maximum Likelihood

Applying the method described in section 1.3.1 to estimate $\hat{\boldsymbol{\Sigma}}_j$ for each SNP, we can estimate $\boldsymbol{\Omega}$ using ML. As our likelihood function, we use the marginal density of the estimates,

$$\hat{\boldsymbol{\beta}}_j \sim N\left(\mathbf{0},\, \boldsymbol{\Omega} + \boldsymbol{\Sigma}_j\right).$$

This implies a log-likelihood function

$$L\left(\boldsymbol{\Omega};\, \hat{\boldsymbol{\beta}}_j,\, \boldsymbol{\Sigma}_j\right) \propto \sum_j \left[-\frac{T}{2}\log\left(2\pi\right) - \frac{1}{2}\log\left|\boldsymbol{\Omega} + \boldsymbol{\Sigma}_j\right| - \frac{1}{2}\hat{\boldsymbol{\beta}}_j'\left(\boldsymbol{\Omega} + \boldsymbol{\Sigma}_j\right)^{-1}\hat{\boldsymbol{\beta}}_j\right], \tag{29}$$

where the operator $\propto$ implies that the two expressions are proportional with respect to the matrix $\boldsymbol{\Omega}$. The value $\hat{\boldsymbol{\Omega}}_{\text{MLE}}$ that maximizes (29) does not have a closed form in general. Therefore we find it numerically. Since $\boldsymbol{\Omega}$ is symmetric, it has $\frac{1}{2}T\left(T+1\right)$ elements. Therefore, finding the maximum is not too computationally burdensome as long as T is sufficiently small. For example, in the application of this paper ($T = 3$), using one core of a 2.20 GHz Intel(R) Xeon(R) CPU E5-2650 v4 processor, this maximization problem took approximately 3 hours. (By comparison, the method-of-moments estimation of $\boldsymbol{\Omega}$ that we now use for the results in the paper took 2 seconds. The two methods of estimation produce nearly identical $\hat{\boldsymbol{\Omega}}$.)

There is an alternative approach that can be used when T is too large for numerical optimization to be feasible. This alternative exploits the fact that when the estimation sample size is constant across SNPs, equation (29) has a closed-form solution which can be implemented if we restrict the set of SNPs in the analysis to those with a constant (or nearly constant) sample size. Of course, the downside of this faster approach is that the additional SNP filters may remove SNPs that would otherwise be genome-wide significant. (By comparison, the GMM estimator for $\boldsymbol{\Omega}$ has a closed-form solution that does not require the sample size to be constant across SNPs.)

We end this section by deriving the closed-form solution in the special case of a constant sample size across SNPs. It is clear that when $\boldsymbol{W}_j$ is constant across SNPs j, $\boldsymbol{\Sigma}_j \equiv \boldsymbol{\Sigma}$ is also constant across SNPs. Taking

the derivative of $L\left(\mathbf{\Omega};\, \hat{\boldsymbol{\beta}}_j,\, \mathbf{\Sigma}_j\right)$ with respect to $\mathbf{\Omega}$, setting it equal to zero, and rearranging:

$$
\begin{aligned}
0 &= \frac{\partial}{\partial \mathbf{\Omega}} L\left(\mathbf{\Omega};\, \hat{\boldsymbol{\beta}}_j,\, \mathbf{\Sigma}\right) \\
&= \frac{\partial}{\partial \mathbf{\Omega}} \sum_j \left[-\frac{T}{2}\log\left(2\pi\right) - \frac{1}{2}\log\left|\mathbf{\Omega}+\mathbf{\Sigma}\right| - \frac{1}{2}\hat{\boldsymbol{\beta}}_j'\left(\mathbf{\Omega}+\mathbf{\Sigma}\right)^{-1}\hat{\boldsymbol{\beta}}_j \right] \\
&= -\frac{1}{2}\frac{\partial}{\partial \mathbf{\Omega}} \left[T\log\left(2\pi\right) + \log\left|\mathbf{\Omega}+\mathbf{\Sigma}\right| + \frac{1}{M}\sum_j \hat{\boldsymbol{\beta}}_j'\left(\mathbf{\Omega}+\mathbf{\Sigma}\right)^{-1}\hat{\boldsymbol{\beta}}_j \right] \\
&= -\frac{1}{2}\frac{\partial}{\partial \mathbf{\Omega}} \left\{ T\log\left(2\pi\right) + \log\left|\mathbf{\Omega}+\mathbf{\Sigma}\right| + \mathrm{tr}\left[\frac{1}{M}\sum_j \hat{\boldsymbol{\beta}}_j'\left(\mathbf{\Omega}+\mathbf{\Sigma}\right)^{-1}\hat{\boldsymbol{\beta}}_j \right] \right\} \\
&= -\frac{1}{2}\frac{\partial}{\partial \mathbf{\Omega}} \left\{ T\log\left(2\pi\right) + \log\left|\mathbf{\Omega}+\mathbf{\Sigma}\right| + \frac{1}{M}\sum_j \mathrm{tr}\left[\hat{\boldsymbol{\beta}}_j'\left(\mathbf{\Omega}+\mathbf{\Sigma}\right)^{-1}\hat{\boldsymbol{\beta}}_j \right] \right\} \\
&= -\frac{1}{2}\frac{\partial}{\partial \mathbf{\Omega}} \left\{ T\log\left(2\pi\right) + \log\left|\mathbf{\Omega}+\mathbf{\Sigma}\right| + \frac{1}{M}\sum_j \mathrm{tr}\left[\left(\mathbf{\Omega}+\mathbf{\Sigma}\right)^{-1}\hat{\boldsymbol{\beta}}_j\hat{\boldsymbol{\beta}}_j' \right] \right\} \\
&= -\frac{1}{2}\frac{\partial}{\partial \mathbf{\Omega}} \left\{ T\log\left(2\pi\right) + \log\left|\mathbf{\Omega}+\mathbf{\Sigma}\right| + \mathrm{tr}\left[\left(\mathbf{\Omega}+\mathbf{\Sigma}\right)^{-1}\frac{1}{M}\sum_j \left(\hat{\boldsymbol{\beta}}_j\hat{\boldsymbol{\beta}}_j'\right) \right] \right\} \\
&= -\frac{1}{2}\left\{ \frac{\partial}{\partial \mathbf{\Omega}}\log\left|\mathbf{\Omega}+\mathbf{\Sigma}\right| + \frac{\partial}{\partial \mathbf{\Omega}}\mathrm{tr}\left[\left(\mathbf{\Omega}+\mathbf{\Sigma}\right)^{-1}\frac{1}{M}\sum_j \left(\hat{\boldsymbol{\beta}}_j\hat{\boldsymbol{\beta}}_j'\right) \right] \right\} \\
&= -\frac{1}{2}\left(\mathbf{\Omega}+\mathbf{\Sigma}\right)^{-1} + \frac{1}{2}\left(\mathbf{\Omega}+\mathbf{\Sigma}\right)^{-1}\left[\frac{1}{M}\sum_j \left(\hat{\boldsymbol{\beta}}_j\hat{\boldsymbol{\beta}}_j'\right) \right]\left(\mathbf{\Omega}+\mathbf{\Sigma}\right)^{-1} \\
\left(\mathbf{\Omega}+\mathbf{\Sigma}\right)^{-1} &= \left(\mathbf{\Omega}+\mathbf{\Sigma}\right)^{-1}\left[\frac{1}{M}\sum_j \left(\hat{\boldsymbol{\beta}}_j\hat{\boldsymbol{\beta}}_j'\right) \right]\left(\mathbf{\Omega}+\mathbf{\Sigma}\right)^{-1} \\
\mathbf{\Omega}+\mathbf{\Sigma} &= \frac{1}{M}\sum_j \left(\hat{\boldsymbol{\beta}}_j\hat{\boldsymbol{\beta}}_j'\right) \\
\mathbf{\Omega} &= \frac{1}{M}\sum_j \left(\hat{\boldsymbol{\beta}}_j\hat{\boldsymbol{\beta}}_j'\right) - \mathbf{\Sigma},
\end{aligned}
$$

where M is the number of SNPs in the MTAG analysis. Thus,

$$\hat{\mathbf{\Omega}}_{\mathrm{MLE}} = \frac{1}{M}\sum_j \left(\hat{\boldsymbol{\beta}}_j\hat{\boldsymbol{\beta}}_j'\right) - \mathbf{\Sigma}. \tag{30}$$

Substituting in our estimate of $\mathbf{\Sigma}$, equation (30) is the closed-form estimator for the case of a constant sample size across SNPs. Note that the method of moments estimator presented in section 1.3.2, $\hat{\mathbf{\Omega}} = \frac{1}{M}\sum_{j=1}^{M}\left(\hat{\boldsymbol{\beta}}_j\hat{\boldsymbol{\beta}}_j' - \mathbf{\Sigma}_j\right)$, also specializes to equation (30) when the sample size is constant across SNPs.

9.1.3 Non-Positive Definiteness

One problem that arises in the ML implementation of MTAG that does not occur in the GMM implementation is that the likelihood function (29) may not exist if $\mathbf{\Sigma}_j$ is not positive definite. There are many ways that one may transform a non-positive definite matrix into a similar matrix that is positive-definite. We describe two such options below.

Deflation of the off-diagonal. For this approach, we proceed in two steps. First, a necessary condition for the matrix to be positive definite is that the implied correlation between every pair of traits t and u, defined as

$$\frac{\Sigma_{LD,t,u}}{\sqrt{\Sigma_{LD,t,t}\Sigma_{LD,u,u}}},$$

must be between -1 and 1. If $|\Sigma_{LD,t,u}| > \sqrt{\Sigma_{LD,t,t}\Sigma_{LD,u,u}}$ for any pair, we shrink the two (symmetric) elements of the matrix corresponding to that pair of traits by a constant multiple toward zero until $|\Sigma_{LD,t,u}| = \sqrt{\Sigma_{LD,t,t}\Sigma_{LD,u,u}}$.

Second, since the first step only ensures that Σ_{LD} satisfies a necessary condition of being positive definite, the resulting transformed matrix may still not be positive definite. We therefore shrink all of the off-diagonal elements simultaneously by a constant multiple until the matrix is positive definite. This will with certainty produce a positive definite matrix for some shrinkage factor (since in the limit this shrinkage will result in a diagonal matrix with positive diagonal elements). We shrink the minimum amount necessary in order to preserve as much similarity between the original matrix Σ_{LD} and the transformed matrix used in the analysis.

Diagonalization. A characteristic of positive definite matrices is that they have all strictly positive eigenvalues. Since $\mathbf{\Sigma}_j$ is symmetric, it is diagonalizable. That is, there exist $T \times T$ matrices $\boldsymbol{P}$ and $\boldsymbol{A}$ such that $\mathbf{\Sigma}_j = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$, where $\boldsymbol{A}$ is a diagonal matrix containing the eigenvalues of $\mathbf{\Sigma}_j$. Define $\boldsymbol{A}^\star$ as a matrix equal to $\boldsymbol{A}$ except that each negative element is changed to a small positive number. Then $\mathbf{\Sigma}_j^\star = \boldsymbol{P}^{-1}\boldsymbol{A}^\star\boldsymbol{P}$ will be a positive definite matrix similar to $\mathbf{\Sigma}_j$.

9.2 Other Moment Conditions

We now return to the GMM framework for MTAG and discuss a possible extension. The MTAG moment condition is based on one of the first-order conditions of the best-linear-predictor problem outlined in section 1.2.3:

$$\mathbb{E}\left(\hat{\beta}_{j,s} - \frac{\omega_{ts}}{\omega_{tt}}\beta_{j,t}\right) = 0. \tag{31}$$

Note, however, that we could have also used the other first-order condition of best-linear-predictor problem,

$$\mathbb{E}\left[\beta_{j,t}\left(\hat{\beta}_{j,s} - \frac{\omega_{ts}}{\omega_{tt}}\beta_{j,t}\right)\right] = 0, \tag{32}$$

giving us an additional T moment conditions, one for each trait. The specification we used in our analysis can be thought of as a specification that uses both (31) and (32) but with a weight matrix that has zeros in every entry corresponding to a moment condition of the form (32). While this weighting will still give consistent estimates, it is not the most efficient estimator that uses both sets of moment equations since some gain in efficiency may be achieved by giving positive weight to the moment equations defined by (32).

We do not pursue this extension of MTAG because it would be substantially more computationally intensive. There are two reasons. First, to calculate the efficient GMM weight matrix, we would need to calculate the variance-covariance matrix of all moments. In this case, it would be a function of third and fourth moments of the data, which would either need to be estimated or additional assumptions would been to be made about the distribution of $\boldsymbol{\beta}_j$. Second, the GMM estimator based on (31) and (32) does not have a general, closed-form solution, so the objective function would need to be maximized numerically for each SNP, making it much more computationally intensive than the version of MTAG we focus on. If a method is developed that overcomes these concerns, it may be of interest to evaluate the gains of such an estimator over the version of MTAG developed here.

10 Extended Acknowledgements

This paper uses summary statistics from Okbay *et al.* [8]. Per SSGAC policy, below we list the authors of that paper:

Aysu Okbay, Jan-Emmanuel De Neve, Patrick Turley, Mark Alan Fontana, S Fleur W Meddens, Richard Karlsson Linnér, Cornelius A Rietveld, Jaime Derringer, Jacob Gratten, James J Lee, Jimmy Z Liu, Ronald de Vlaming, Tarunveer S Ahluwalia, Jadwiga Buchwald, Alana Cavadino, Alexis C Frazier-Wood, Nicholas A Furlotte, Victoria Garfield, Marie Henrike Geisel, Juan R Gonzalez, Saskia Haitjema, Robert Karlsson, Sander W van der Laan, Karl-Heinz Ladwig, Jari Lahti, Sven J van der Lee, Penelope A Lind, Tian Liu, Lindsay Matteson, Evelin Mihailov, Michael B Miller, Camelia C Minica, Ilja M Nolte, Dennis Mook-Kanamori, Peter J van der Most, Christopher Oldmeadow, Yong Qian, Olli Raitakari, Rajesh Rawal, Anu Realo, Rico Rueedi, Börge Schmidt, Albert V Smith, Evie Stergiakouli, Toshiko Tanaka, Kent Taylor, Gudmar Thorleifsson, Juho Wedenoja, Juergen Wellmann, Harm-Jan Westra, Sara M Willems, Wei Zhao, LifeLines Cohort Study, Najaf Amin, Andrew Bakshi, Sven Bergmann, Gyda Bjornsdottir, Patricia A Boyle, Samantha Cherney, Simon R Cox, Gail Davies, Oliver S P Davis, Jun Ding, Nese Direk, Peter Eibich, Rebecca T Emeny, Ghazaleh Fatemifar, Jessica D Faul, Luigi Ferrucci, Andreas J Forstner, Christian Gieger, Richa Gupta, Tamara B Harris, Juliette M Harris, Elizabeth G Holliday, Jouke-Jan Hottenga, Philip L De Jager, Marika A Kaakinen, Eero Kajantie, Ville Karhunen, Ivana Kolcic, Meena Kumari, Lenore J Launer, Lude Franke, Ruifang Li-Gao, David C Liewald, Marisa Koini, Anu Loukola, Pedro Marques-Vidal, Grant W Montgomery, Miriam A Mosing, Lavinia Paternoster, Alison Pattie, Katja E Petrovic, Laura Pulkki-Råback, Lydia Quaye, Katri Räikkönen, Igor Rudan, Rodney J Scott, Jennifer A Smith, Angelina R Sutin, Maciej Trzaskowski, Anna E Vinkhuyzen, Lei Yu, Delilah Zabaneh, John R Attia, David A Bennett, Klaus Berger, Lars Bertram, Dorret I Boomsma, Harold Snieder, Shun-Chiao Chang, Francesco Cucca, Ian J Deary, Cornelia M van Duijn, Johan G Eriksson, Ute Bültmann, Eco J C de Geus, Patrick J F Groenen, Vilmundur Gudnason, Torben Hansen, Catharine A Hartman, Claire M A Haworth, Caroline Hayward, Andrew C Heath, David A Hinds, Elina Hyppönen, William G Iacono, Marjo-Riitta Järvelin, Karl-Heinz Jöckel, Jaakko Kaprio, Sharon L R Kardia, Liisa Keltikangas-Järvinen, Peter Kraft, Laura D Kubzansky, Terho Lehtimäki, Patrik K E Magnusson, Nicholas G Martin, Matt McGue, Andres Metspalu, Melinda Mills, Renée de Mutsert, Albertine J Oldehinkel, Gerard Pasterkamp, Nancy L Pedersen, Robert Plomin, Ozren Polasek, Christine Power, Stephen S Rich, Frits R Rosendaal, Hester M den Ruijter, David Schlessinger, Helena Schmidt, Rauli Svento, Reinhold Schmidt, Behrooz Z Alizadeh, Thorkild I A Sørensen, Tim D Spector, John M Starr, Kari Stefansson, Andrew Steptoe, Antonio Terracciano, Unnur Thorsteinsdottir, A Roy Thurik, Nicholas J Timpson, Henning Tiemeier, André G Uitterlinden, Peter Vollenweider, Gert G Wagner, David R Weir, Jian Yang, Dalton C Conley, George Davey Smith, Albert Hofman, Magnus Johannesson, David I Laibson, Sarah E Medland, Michelle N Meyer, Joseph K Pickrell, Tõnu Esko, Robert F Krueger, Jonathan P Beauchamp, Philipp D Koellinger, Daniel J Benjamin, and David Cesarini.

23andMe, Inc. – 23andMe research participants provided informed consent to take part in this research under a protocol approved by the AAHRPP-accredited institutional review board, Ethical and Independent Review Services. We would like to thank the research participants and employees of 23andMe for making this work possible.

Add Health (National Longitudinal Study of Adolescent to Adult Health) — The National Longitudinal Study of Adolescent to Adult Health (Add Health) is supported by grant P01 HD031921 to Kathleen Mullan Harris

from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), with cooperative funding from 23 other federal agencies and foundations. Add Health GWAS data were funded by NICHD grants to Harris (R01 HD073342) and to Harris, Boardman, and McQueen (R01 HD060726). For information about access to the data from this study, contact addhealth@unc.edu.

UK Biobank – Informed consent was obtained from all subjects.

References

[1] Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics* **47**, 291–295 (2015).

[2] Vilhjálmsson, B. J. *et al.* Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *The American Journal of Human Genetics* **97**, 576–592 (2015).

[3] Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A tool for genome-wide complex trait analysis. *The American Journal of Human Genetics* **88**, 76–82 (2011).

[4] Hall, A. R. *Generalized Method of Moments* (Oxford University Press, 2005).

[5] Hansen, L. P. Large sample properties of generalised method of moments estimators. *Econometrica* **50**, 1029–1054 (1982).

[6] Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature Genetics* (2015).

[7] Speed, D. *et al.* Reevaluation of SNP heritability in complex human traits. *Nature Genetics* **49**, 986–992 (2017).

[8] Okbay, A. *et al.* Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature Genetics* **48**, 624–633 (2016). 15334406.

[9] Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542 (2016).

[10] Hyde, C. L. *et al.* Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nature Genetics* **48**, 1031–6 (2016).

[11] de Moor, M. H. M. *et al.* Meta-analysis of genome-wide association studies for neuroticism, and the polygenic association with Major Depressive Disorder. *JAMA Psychiatry* **72**, 642–650 (2015).

[12] Winkler, T. W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nature Protocols* **9**, 1192–1212 (2014).

[13] Li, Y., Willer, C. J., Ding, J., Scheet, P. & Abecasis, G. R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genetic Epidemiology* **34**, 816–834 (2010).

[14] Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature Genetics* **39**, 906–913 (2007).

[15] Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *European Journal of Human Genetics* **19**, 807–812 (2011).

[16] Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).

[17] Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).

[18] de Moor, M. H. M. *et al.* Meta-analysis of genome-wide association studies for personality. *Molecular Psychiatry* **17**, 337–349 (2012).

[19] Smith, D. J. *et al.* Genome-wide analysis of over 106 000 individuals identifies 9 neuroticism-associated loci. *Molecular Psychiatry* **21**, 749–57 (2016). `/dx.doi.org/10.1101/032417`.

[20] Visscher, P. M. M., Brown, M. A. A., McCarthy, M. I. I. & Yang, J. Five years of GWAS discovery. *The American Journal of Human Genetics* **90**, 7–24 (2012).

[21] Juster, T. F. & Suzman, R. An overview of the Health and Retirement Study. *Journal of Human Resources* **20**, 7–56 (1995).

[22] Kessler, R. C. *et al.* Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. *Archives of general psychiatry* **51**, 8–19 (1994).

[23] Lachman, M. E. & Weaver, S. L. The Midlife Development Inventory (MIDI) personality scales: Scale construction and scoring. *Waltham, MA: Brandeis University* 1–9 (1997).

[24] Clarke, Phillipa; Fisher, Swenith; House, Jim; Smith, Jacqui; Weir, D. Guide to Content of the HRS Psychosocial Leave-Behind Participant Lifestyle Questionnaires : 2004 & 2006. *HRS Psychosocial Working Group: Survey Research Center, Institute for Social Research, University of MIchigan* 1–48 (2008).

[25] Watson, D. & Clark, L. A. T HE PANAS-X Manual for the Positive and Negative Affect Schedule - Expanded Form. *Order A Journal On The Theory Of Ordered Sets And Its Applications* **277**, 1–27 (1994).

[26] Carstensen, L. L., Pasupathi, M., Mayr, U. & Nesselroade, J. R. Emotional experience in everyday life across the adult life span. *Journal of Personality and Social Psychology* **79**, 644–655 (2000).

[27] Ong, A. D., Edwards, L. M. & Bergeman, C. S. Hope as a source of resilience in later adulthood. *Personality and Individual Differences* **41**, 1263–1273 (2006).

[28] Diener, E., Emmons, R. A., Larsen, R. J. & Griffin, S. The satisfaction with life scale. *Journal of Personality Assessment* **49**, 71–75 (1985).

[29] Smith, J. *et al.* HRS psychosocial and lifestyle questionnaire 2006–2010: Documentation report. *Institute for Social Research, University of Michigan* (2013).

[30] Radloff, L. The CES-D scale a self-report depression scale for research in the general population. *Applied psychological measurement* (1977).

[31] Donnellan, M. B., Oswald, F. L., Baird, B. M. & Lucas, R. E. The Mini-IPIP Scales: Tiny-yet-effective measures of the Big Five Factors of Personality. *Psychological Assessment* **18**, 192–203 (2006).

[32] Goldberg, L. R. *et al.* The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality* **40**, 84–96 (2006).

[33] Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 1–16 (2015).

[34] Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nature Communications* **6**, 5890 (2015).

[35] Frey, B. J. & Dueck, D. Clustering by passing messages between data points. *Science* **315**, 972–976 (2007). 1401.2548.

[36] Südhof, T. C. The presynaptic active zone. *Neuron* **75**, 11–25 (2012).

[37] Dietrich, D. *et al.* Functional specialization of presynaptic Cav2.3 Ca2+ channels. *Neuron* **39**, 483–496 (2003).

[38] Ricoy, U. M. & Frerking, M. E. Distinct roles for Ca_v2.1-2.3 in activity-dependent synaptic dynamics. *Journal of Neurophysiology* **111**, 2404–13 (2014).

[39] Berton, O. & Nestler, E. J. New approaches to antidepressant drug discovery: Beyond monoamines. *Nature Reviews Neuroscience* **7**, 137–151 (2006).

[40] DeYoung, C. G. Personality neuroscience and the biology of traits. *Social and Personality Psychology Compass* **4**, 1165–1180 (2010).

[41] Pers, T. H. *et al.* Comprehensive analysis of schizophrenia-associated loci highlights ion channel pathways and biologically plausible candidate causal genes. *Human Molecular Genetics* **25**, 1247–1254 (2015).

[42] Deschwanden, A. *et al.* Reduced metabotropic glutamate receptor 5 density in major depression determined by [11C]ABP688 PET and postmortem study. *American Journal of Psychiatry* **168**, 727–734 (2011).

[43] Kohli, M. A. *et al.* The Neuronal Transporter Gene SLC6A15 Confers Risk to Major Depression. *Neuron* **70**, 252–265 (2011).

[44] Bolormaa, S. *et al.* A Multi-Trait, Meta-analysis for Detecting Pleiotropic Polymorphisms for Stature, Fatness and Reproduction in Beef Cattle. *PLoS Genetics* **10**, e1004198 (2014).

[45] Zhu, X. *et al.* Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. *American Journal of Human Genetics* **96**, 21–36 (2015).

[46] Randall, J. C. *et al.* Sex-stratified Genome-wide Association Studies Including 270,000 Individuals Show Sexual Dimorphism in Genetic Loci for Anthropometric Traits. *PLoS Genetics* **9**, e1003500 (2013).

[47] Barrett, W. Hermitian and Positive Definite Matrices. In Hogben, L. (ed.) *Handbook of Linear Algebra*, chap. 8, 8.1–8.12 (Chapman & Hall/CRC, Boca Raton, 2007).

[48] Park, H., Li, X., Song, Y. E., He, K. Y. & Zhu, X. Multivariate analysis of anthropometric traits using summary statistics of genome-wide association studies from GIANT consortium. *PLoS ONE* **11**, e0163912 (2016).

[49] Maier, R. *et al.* Joint analysis of psychiatric disorders increases accuracy of risk prediction for schizophrenia, bipolar disorder, and major depressive disorder. *American Journal of Human Genetics* **96**, 283–94 (2015).

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

[51] Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics* **46**, 1173–1186 (2014). `NIHMS150003`.

[52] Locke, A. E. A. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015). `arXiv:1011.1669v3`.

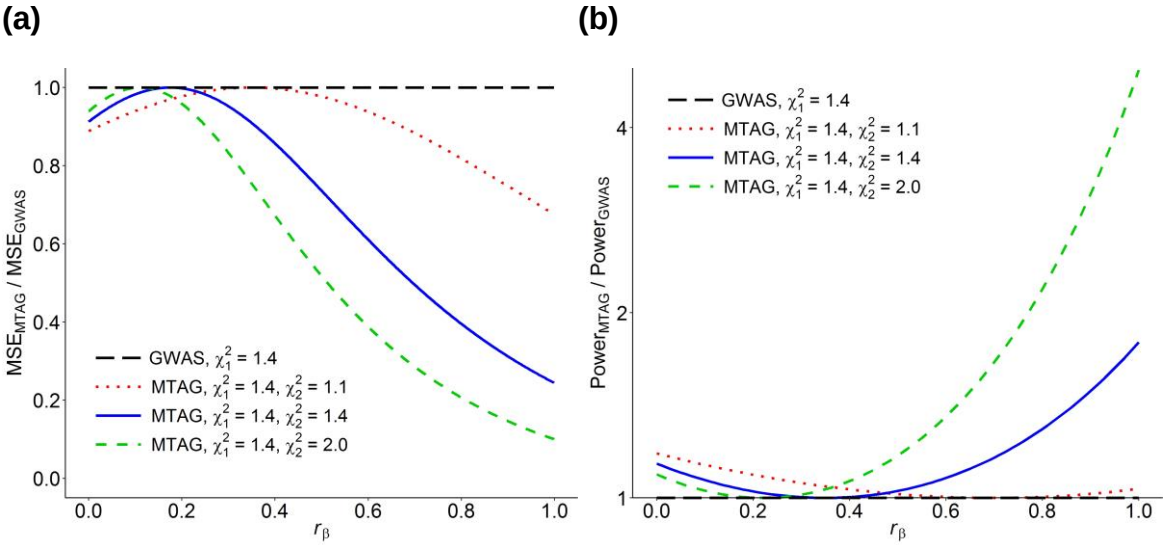
[53] Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187–196 (2015).

[54] Kass, R. & Raftery, A. Bayes Factors. *Journal of the American Statistical Association* **90**, 773–795 (1995).

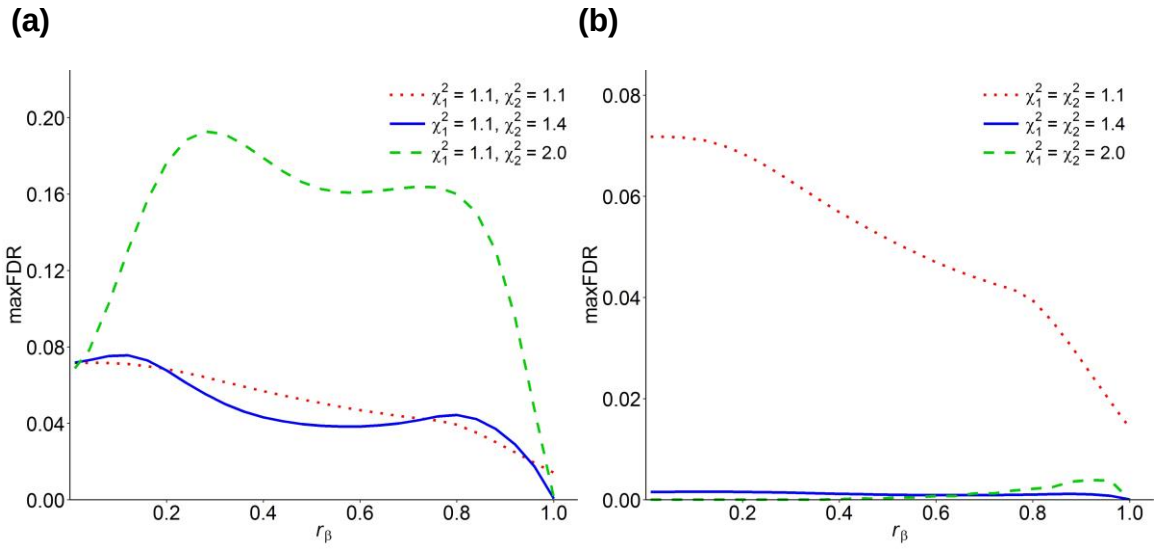
[55] Eaton, M. L. *Multivariate statistics: A vector space approach* (John Wiley {&} Sons, Inc., New York, 1983).

Supplementary Figures and Tables

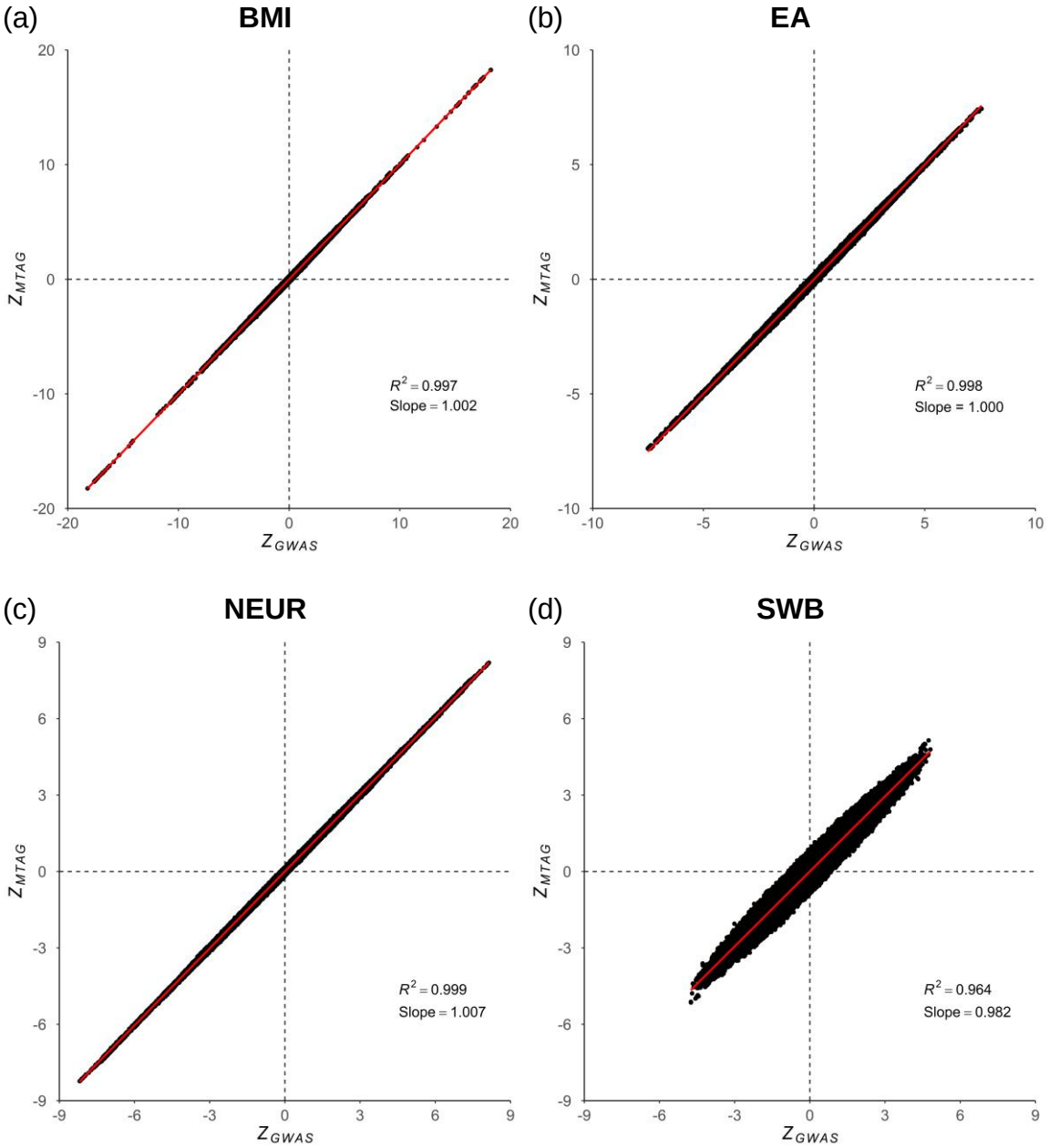
Supplementary Figure 1. Two-trait illustration of MTAG's theoretical mean squared error (MSE) and power. (**a**) The *y*-axis is the ratio of the MSE of the MTAG results to the MSE of the GWAS results of Trait 1. We set $r_\varepsilon = 0.3$ and expected $\chi_1^2 = 1.1$. (**b**) The *y*-axis is MTAG's power to detect a SNP association with trait 1 under an infinitesimal model with normally distributed effect sizes. We set $r_\varepsilon = 0.3$ and expected $\chi_1^2 = 1.1$.



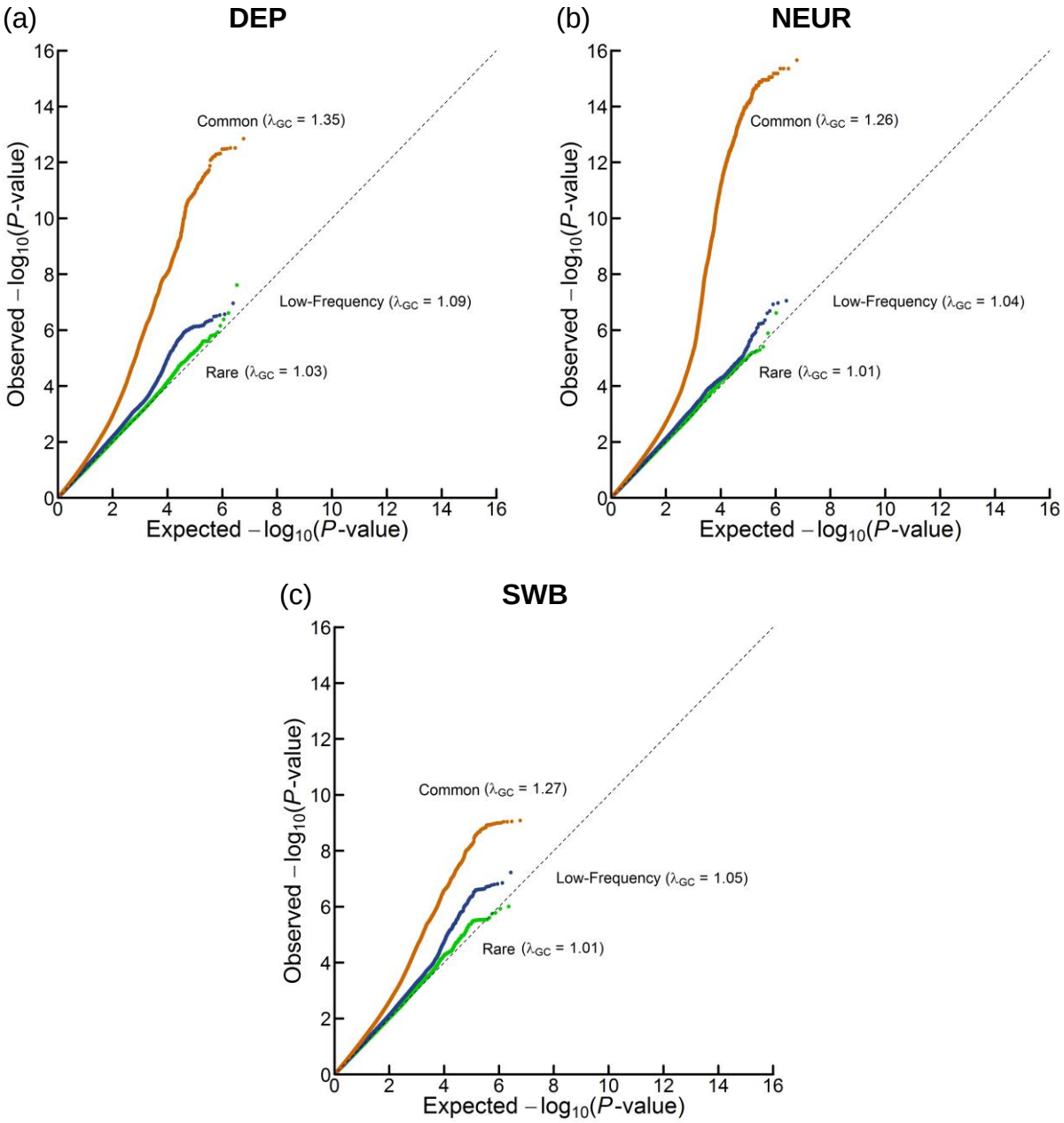
Supplementary Figure 2. Two-trait illustration of MTAG's maxFDR. (**a**) The *y*-axis is the maxFDR attainable under certain assumptions, including the assumption that the fraction of SNPs that are non-null for any individual trait is at least 10%, holding fixed r_β, r_ε, expected χ_1^2, and expected χ_2^2 (see **Online Methods**). In this panel, $r_\varepsilon = 0$ and expected $\chi_1^2 = 1.1$. (**b**) The *y*-axis is maxFDR as in panel (a), but the expected χ^2-statistic is constrained to be equal for the GWASs of the two traits.



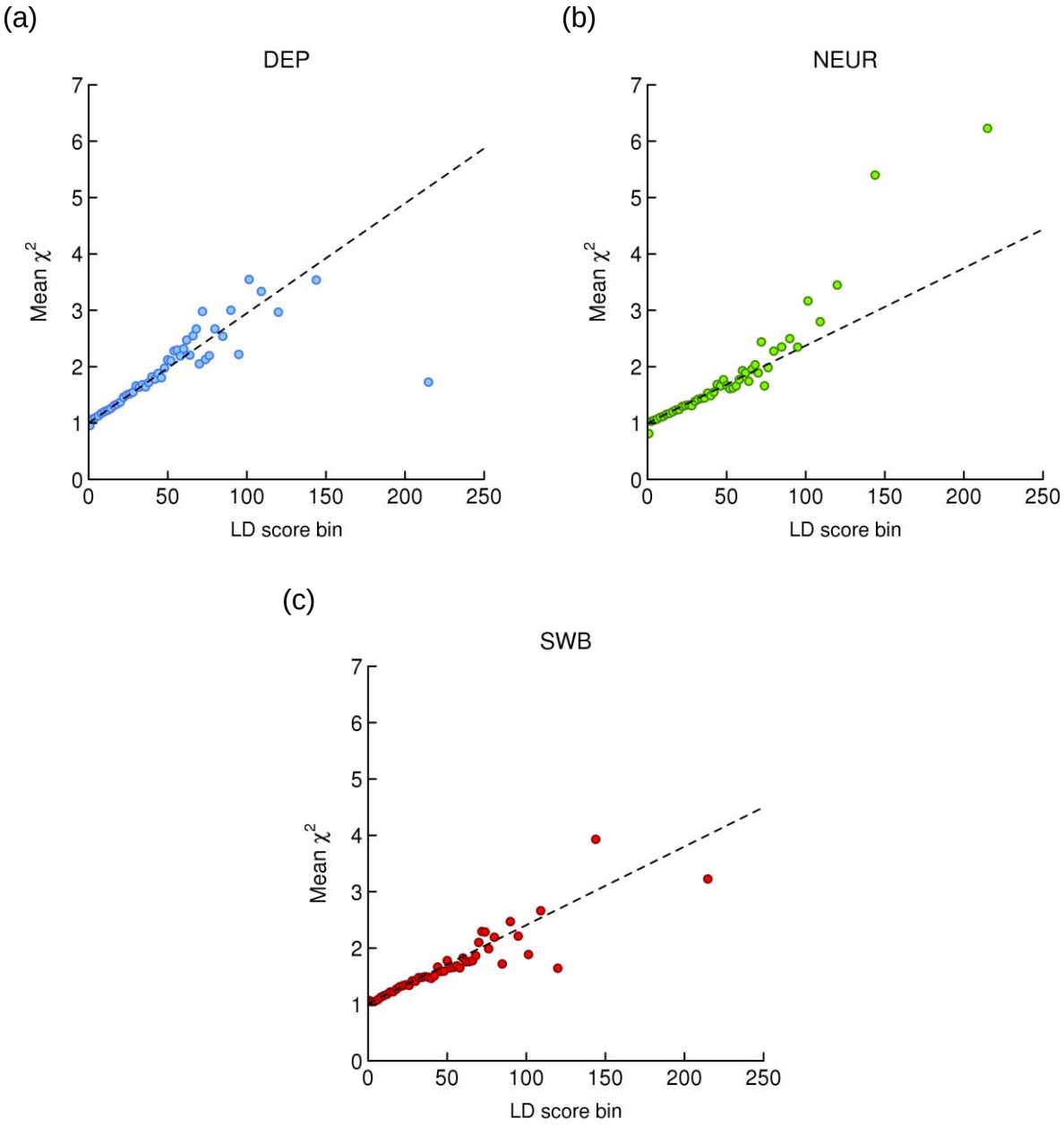
Supplementary Figure 3. Evaluation of MTAG's standard errors when there is sample overlap for additional phenotypes. The *x*-axis is a SNP's *z*-statistic from a baseline GWAS conducted in UK Biobank. The *y*-axis is a SNP's *z*-statistic from applying MTAG to three GWASs of each trait conducted on equally sized subsamples of the baseline sample, in which every pair of samples has 50% overlap. The figure illustrates near-perfect alignment across all phenotypes. See **Supplementary Note** for details. **(a)** body-mass index (BMI), **(b)** educational attainment (EA), **(c)** neuroticism (NEUR), **(d)** subjective well-being (SWB).



Supplementary Figure 4. Quantile-quantile (QQ) plots of GWAS results by allele frequency. (**a**) DEP, (**b**) NEUR, (**c**) SWB. Common variants ($MAF \geq 0.05$), Low-frequency variants ($0.01 \leq MAF < 0.05$), Rare Variants ($MAF < 0.01$). The estimated LD score regression intercept used to adjust the GWAS standard errors is 1.0129 for DEP and 1.0157 for SWB. The standard errors for NEUR were not adjusted because the estimated LD score regression intercept is less than one (0.9896).

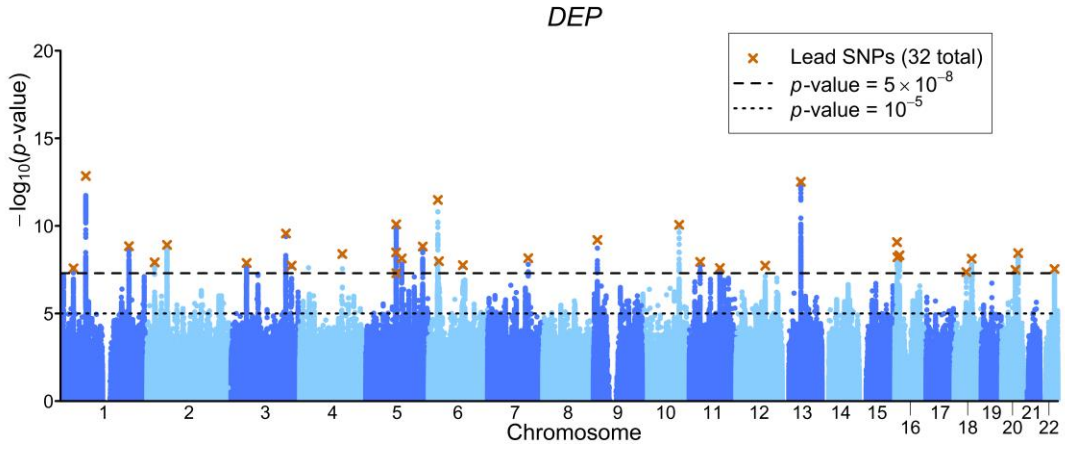


Supplementary Figure 5. LD score regression plots of GWAS results. (**a**) DEP, (**b**) NEUR, (**c**) SWB. Each point represents an LD score quantile. The *x* and *y* coordinates of the point are the mean LD score and the mean χ^2-statistic of SNPs in that quantile, respectively. The facts that the intercepts are close to 1 and that the χ^2-statistics increase linearly with the LD scores for all three traits suggest that the bulk of the inflation in the χ^2-statistics for the three traits is due to true polygenic signal and not to population stratification.

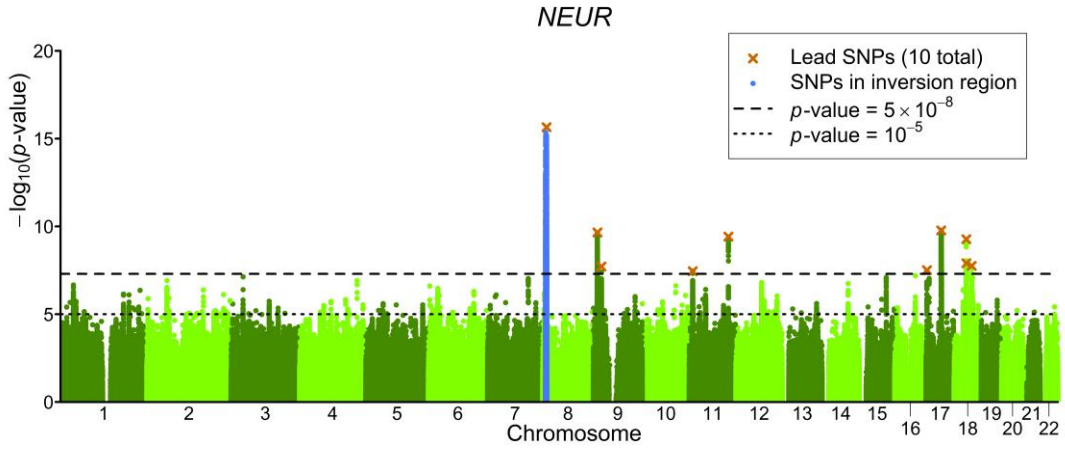


Supplementary Figure 6. Manhattan plots of GWAS results. (**a**) DEP, (**b**) NEUR, (**c**) SWB. The *x*-axis is chromosomal position, and the *y*-axis is the significance on a $-\log_{10}$ scale. The upper dashed line marks the threshold for genome-wide significance ($P = 5 \times 10^{-8}$), and the lower line marks the threshold for nominal significance ($P = 10^{-5}$). Each approximately independent genome-wide significant association ("lead SNP") is marked by ×. In (**b**), the blue spike on chromosome 8 corresponds to SNPs in LD with an inversion polymorphism. These SNPs are dropped in the MTAG analyses. For these plots, we do not impose the MTAG filters described in **Online Methods**.

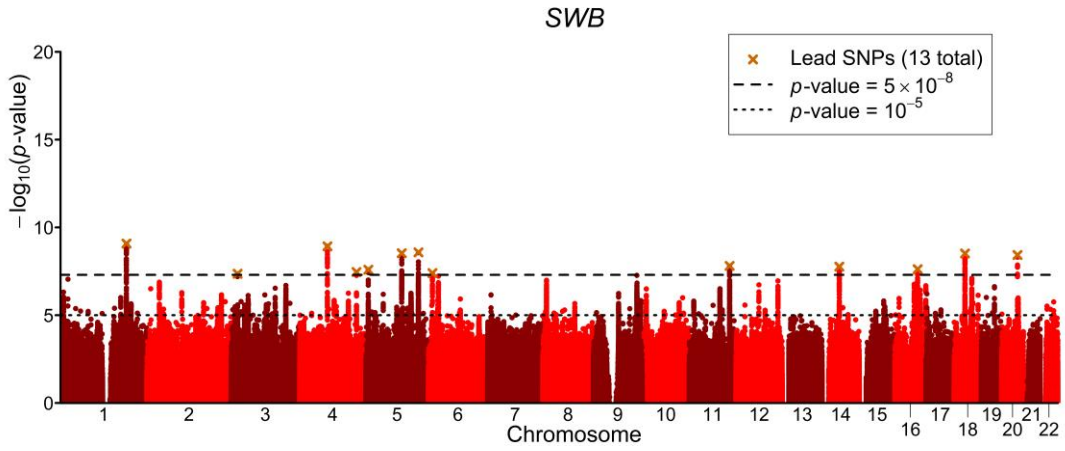
(a)



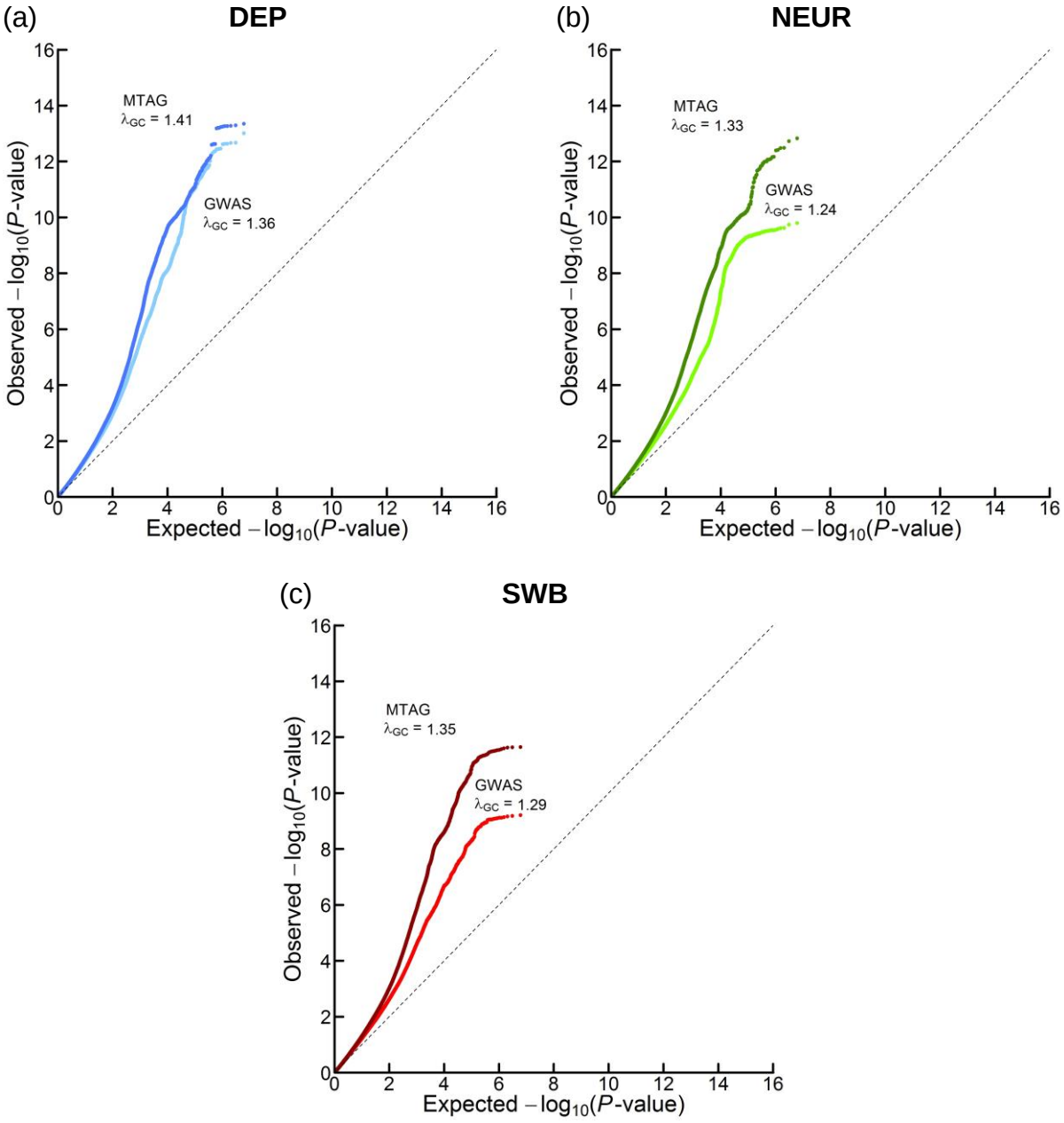
(b)



(c)

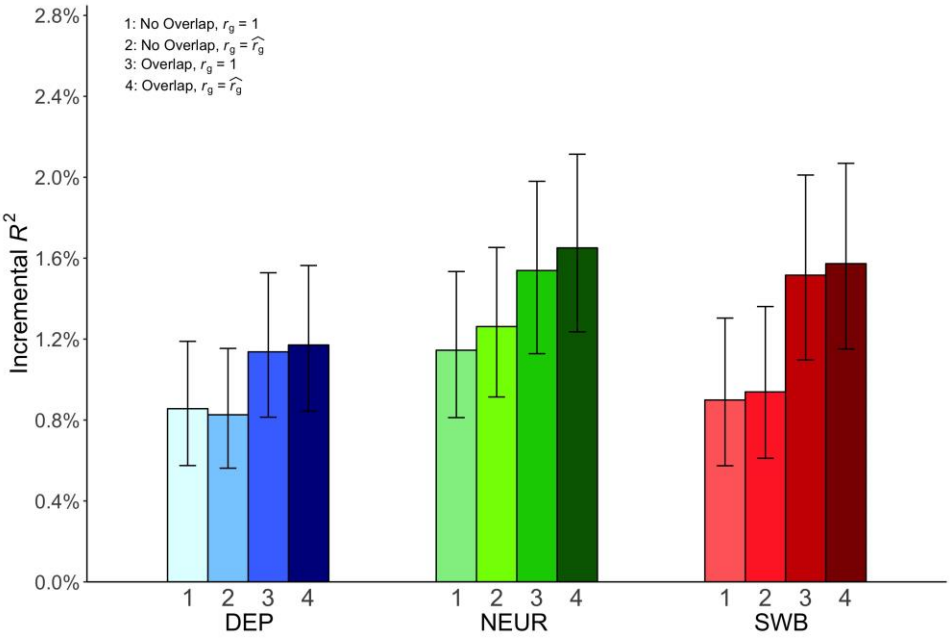

SWB
Lead SNPs (13 total)
p-value = 5 × 10^-8
p-value = 10^-5
−log10(p-value)
20
15
10
5
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Chromosome

Supplementary Figure 7. Quantile-quantile (QQ) Plots for GWAS and MTAG results. (**a**) DEP, (**b**) NEUR, (**c**) SWB. To facilitate comparisons, the GWAS and MTAG results are shown for an identical set of SNPs. GWAS results are adjusted using the estimated LD score regression intercept. The standard errors of the MTAG results were not adjusted since MTAG adjusts using LD score regression as part of its estimation procedure.

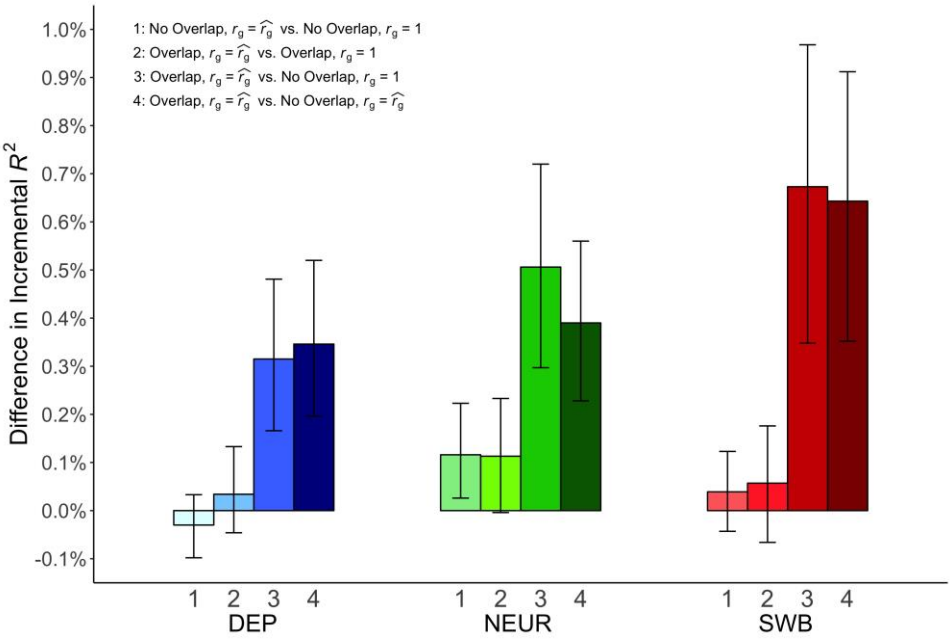


Supplementary Figure 8. Comparison of polygenic scores based on restricted analyses. Incremental R^2 is the increase in R^2 from a linear regression of the trait on the polygenic score and covariates, relative a linear regression of the trait on only covariates. The plotted incremental R^2's are the sample-size-weighted means across the replication cohorts (HRS and Add Health, combined *N* = 12,641), with 95% confidence intervals. See **Supplementary Note** for details and cohort-level results. Scores labeled "No Overlap" correspond to MTAG analyses where the GWAS summary statistics are restricted such that there is no known overlap between estimation samples for each trait. Scores labeled "Overlap" use the full GWAS summary statistics. Scores labeled "$r_\beta = 1$" correspond to MTAG analyses where $\mathbf{\Omega}$ is constrained such that the effect-size correlation is set to one rather than estimated. Scores labeled "$r_\beta = \hat{r}_\beta$" correspond to MTAG analyses where $\mathbf{\Omega}$ is estimated, which is the default in MTAG. (**a**) Incremental R^2 of MTAG-based polygenic scores. (**b**) Differences in incremental R^2 of MTAG-based polygenic scores across analyses.

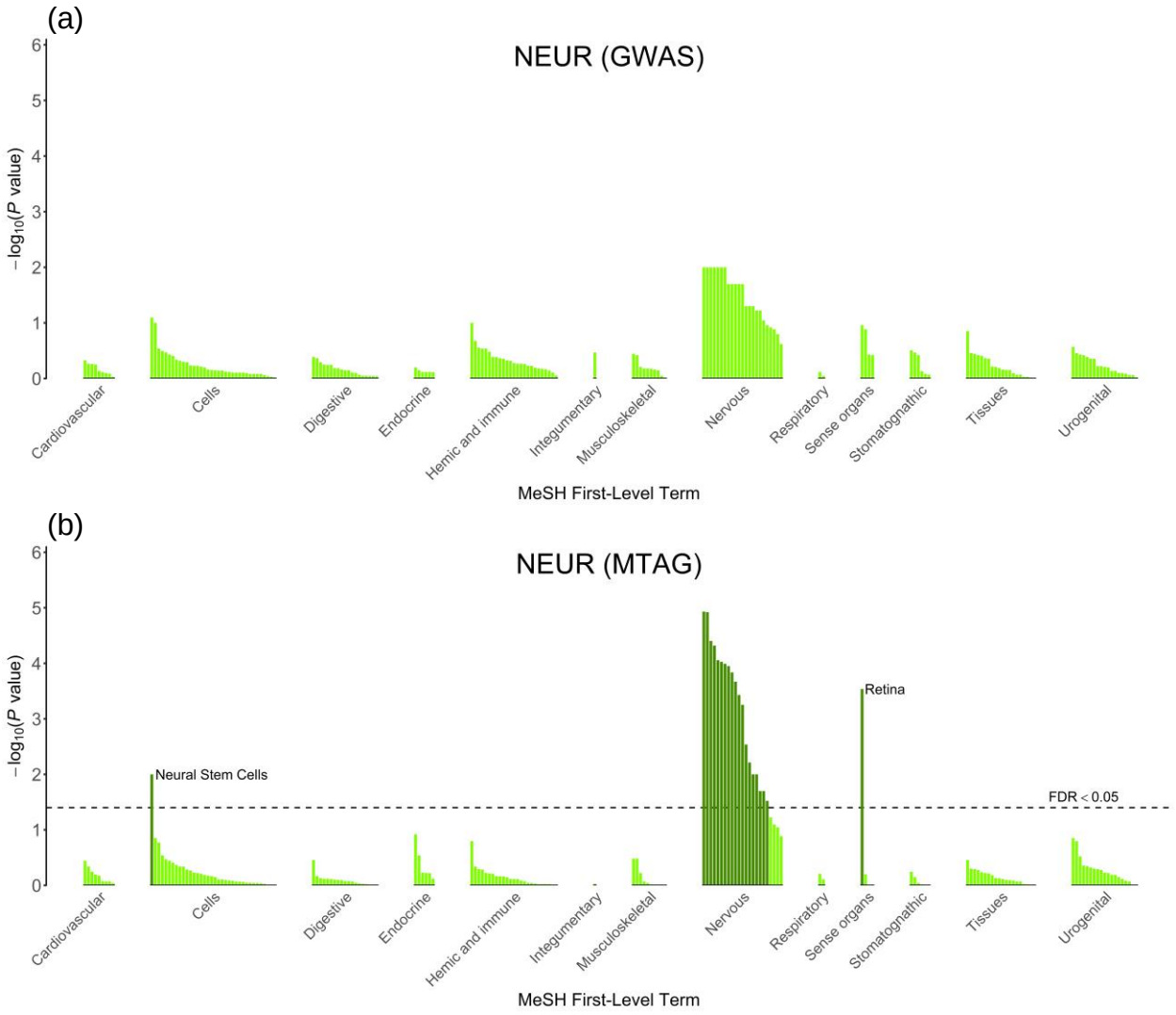
(a)



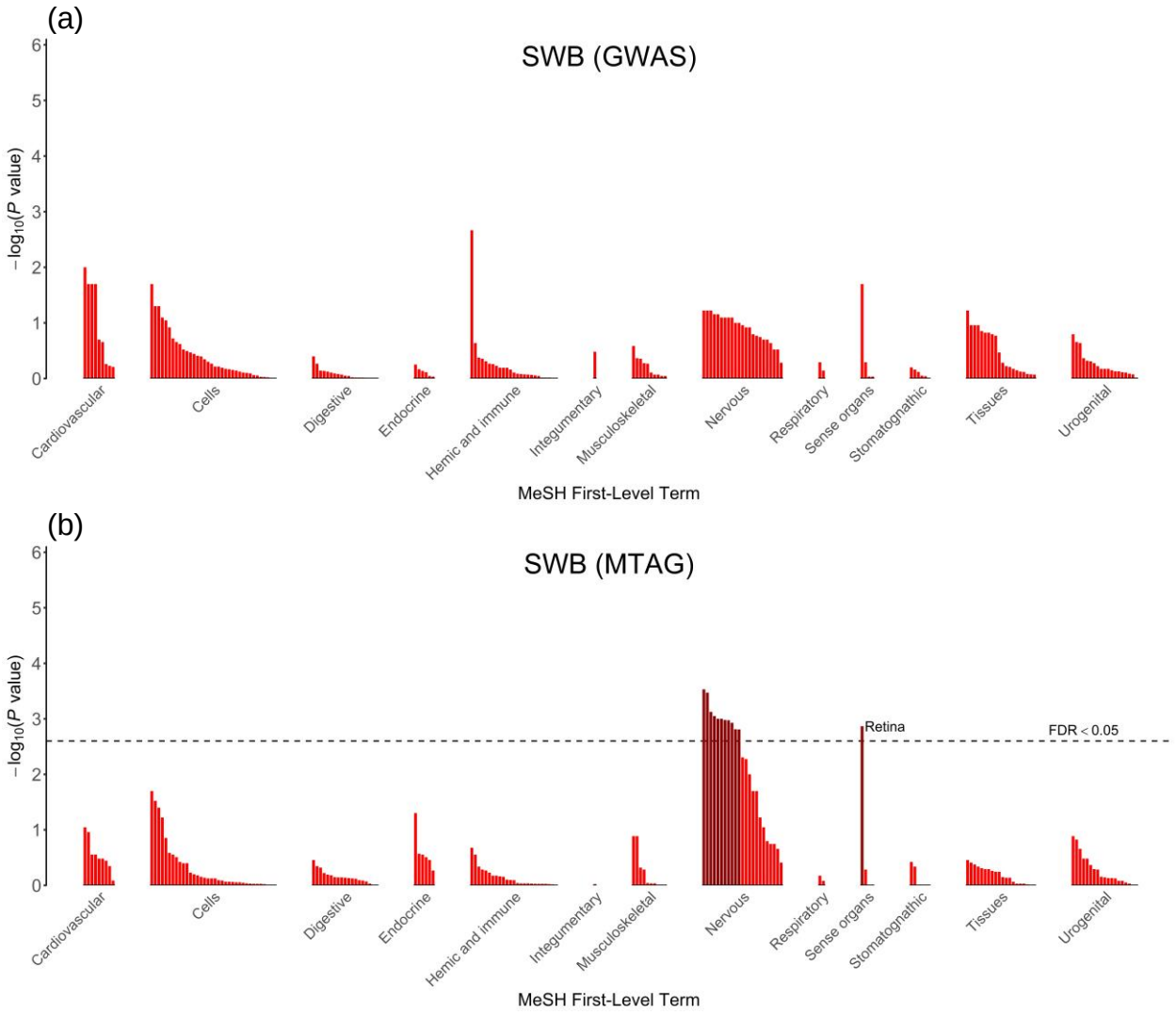
(b)


1: No Overlap, $r_g = \widehat{r_g}$ vs. No Overlap, $r_g = 1$
2: Overlap, $r_g = \widehat{r_g}$ vs. Overlap, $r_g = 1$
3: Overlap, $r_g = \widehat{r_g}$ vs. No Overlap, $r_g = 1$
4: Overlap, $r_g = \widehat{r_g}$ vs. No Overlap, $r_g = \widehat{r_g}$
Difference in Incremental R^2
1.0%
0.9%
0.8%
0.7%
0.6%
0.5%
0.4%
0.3%
0.2%
0.1%
0.0%
-0.1%
1 2 3 4
DEP
1 2 3 4
NEUR
1 2 3 4
SWB

Supplementary Figure 9. Tissue enrichment estimated using the bioinformatics tool DEPICT for NEUR. (**a**) Results based on the GWAS summary statistics, (**b**) results based on the MTAG summary statistics. The *x*-axis lists the tissues tested for enrichment, grouped by the location of the tissue. The *y*-axis is statistical significance on a $-\log_{10}$ scale. The horizontal dashed line corresponds to a false discovery rate of 0.05, which is the threshold used to identify prioritized tissues. There is no line in (**a**) since no tissue surpasses this threshold.

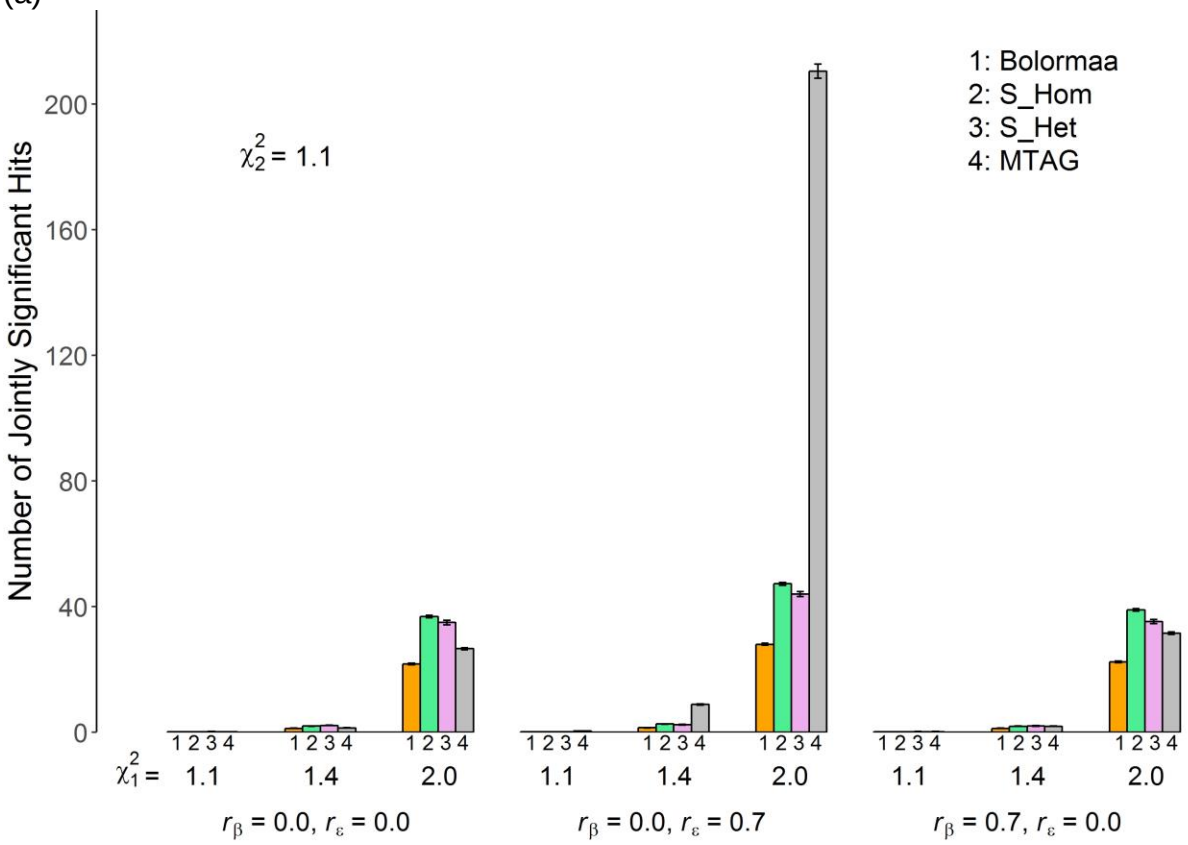


Supplementary Figure 10. Tissue enrichment estimated using the bioinformatics tool DEPICT for SWB. (**a**) Results based on the GWAS summary statistics, (**b**) results based on the MTAG summary statistics. The *x*-axis lists the tissues tested for enrichment, grouped by the location of the tissue. The *y*-axis is statistical significance on a $-\log_{10}$ scale. The horizontal dashed line corresponds to a false discovery rate of 0.05, which is the threshold used to identify prioritized tissues. There is no line in (**a**) since no tissue surpasses this threshold.

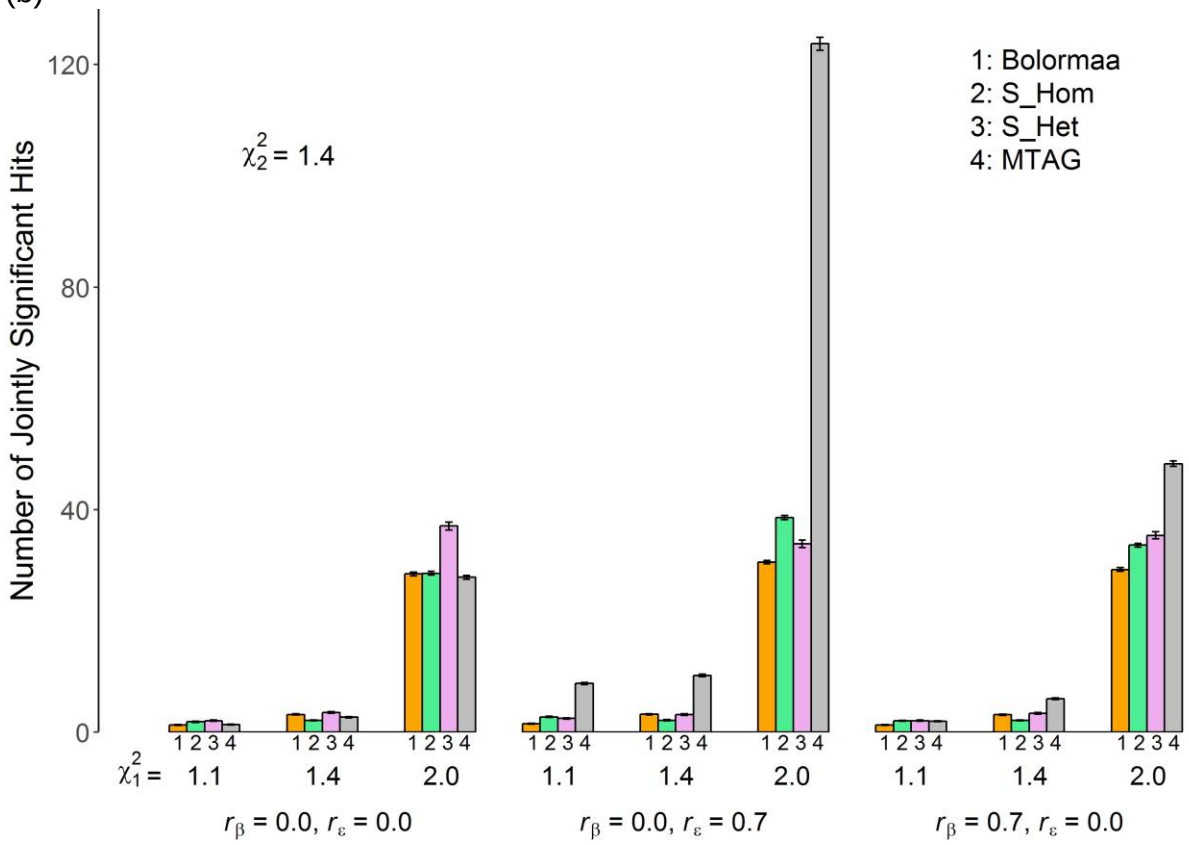


Supplementary Figure 11. Results of comparison of MTAG to Bolormaa, S_{Hom}, and S_{Het} in simulated data. We simulate effect sizes for 100,000 SNPs and two traits, with different values for the expected mean χ^2-statistics for each trait, r_β, and r_ε. *P* values for MTAG are the Bonferroni-adjusted minimum *P* value for the SNP across all traits. The *y*-axis is the mean number of SNPs with a *P* value less than 5×10^{-7} (the Bonferroni-adjusted *P* value for 100,000 SNPs) across 1000 replications. The GWAS for Trait 2 is (**a**) low powered (expected χ^2-statistic of 1.1), (**b**) medium powered (1.4), (**c**) high powered (2.0).

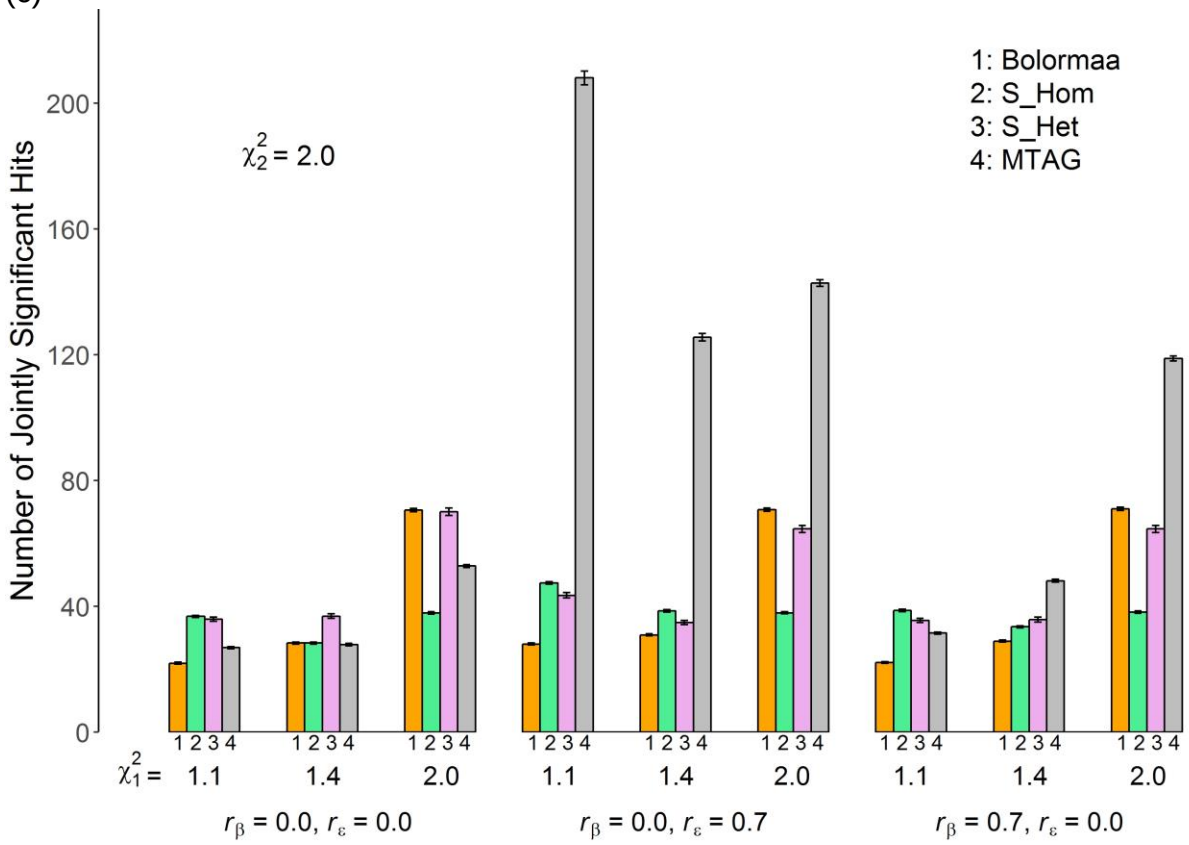
(a)



(b)

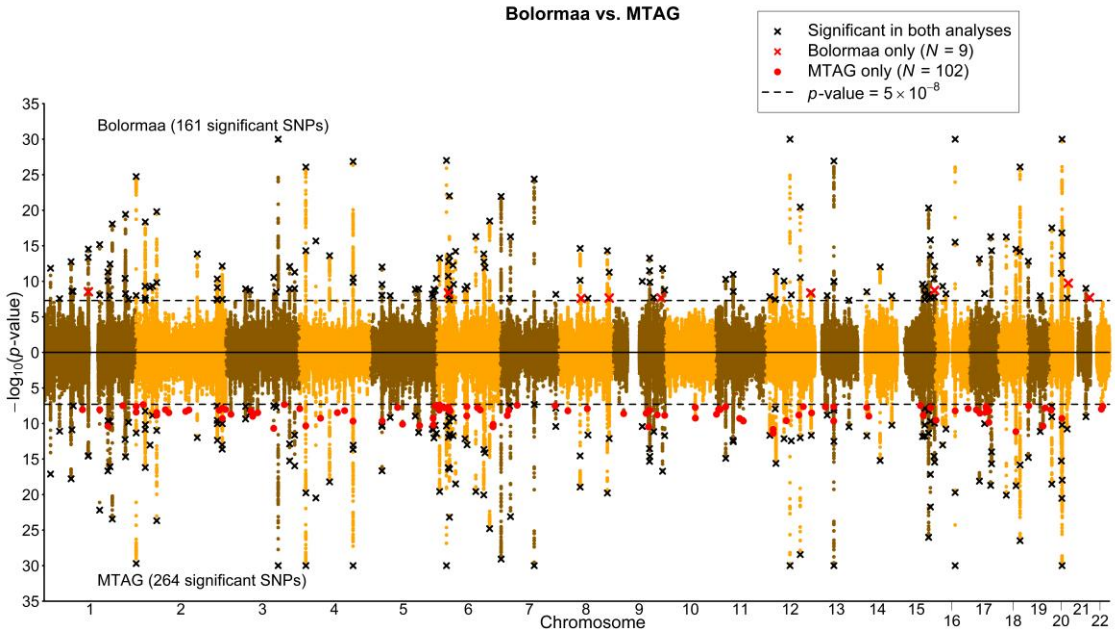

Number of Jointly Significant Hits
120
80
40
0
χ_2^2 = 1.4
1: Bolormaa
2: S_Hom
3: S_Het
4: MTAG
1 2 3 4
χ_1^2 = 1.1 1.4 2.0
r_β = 0.0, r_ε = 0.0
r_β = 0.0, r_ε = 0.7
r_β = 0.7, r_ε = 0.0

(c)

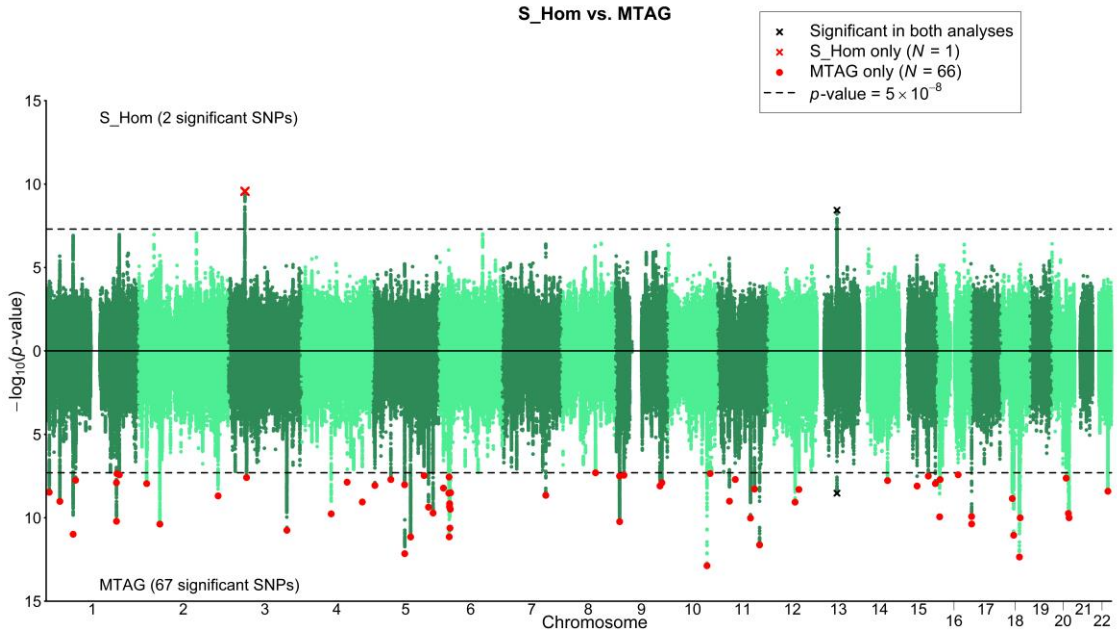

Number of Jointly Significant Hits
0
40
80
120
160
200
χ^2_2 = 2.0
1: Bolormaa
2: S_Hom
3: S_Het
4: MTAG
1 2 3 4
χ^2_1 = 1.1
1.4
2.0
r_β = 0.0, r_ε = 0.0
r_β = 0.0, r_ε = 0.7
r_β = 0.7, r_ε = 0.0

Supplementary Figure 12. Results of comparison of MTAG to Bolormaa, S_{hom}, and S_{het} in GIANT data. (**a**) MTAG vs. Bolormaa, (**b**) MTAG vs. S_{hom}, (**c**) MTAG vs. S_{Het}. For all multi-trait methods, the *P* value for each SNP tests the null hypothesis that the SNP is null for all six traits (height, BMI, and waist-hip ratio each in men and women separately). *P* values for MTAG are the Bonferroni-adjusted minimum *P* value for the SNP across all six traits. Below the *x*-axis in each figure is a Manhattan plot for the MTAG results; above the *x*-axis is a Manhattan plot for the alternative method. MTAG hits that are in LD with a hit identified by the alternative method (or hits from the alternative method that are in LD with a hit found by MTAG) are marked with black x's. Significant SNPs identified by the alternative analysis that are not in LD with any significant MTAG SNPs are marked with red x's. Significant MTAG SNPs not in LD with any significant SNP in the alternative analysis are marked with red dots. The GIANT summary statistics are from Randall et al. (2013).

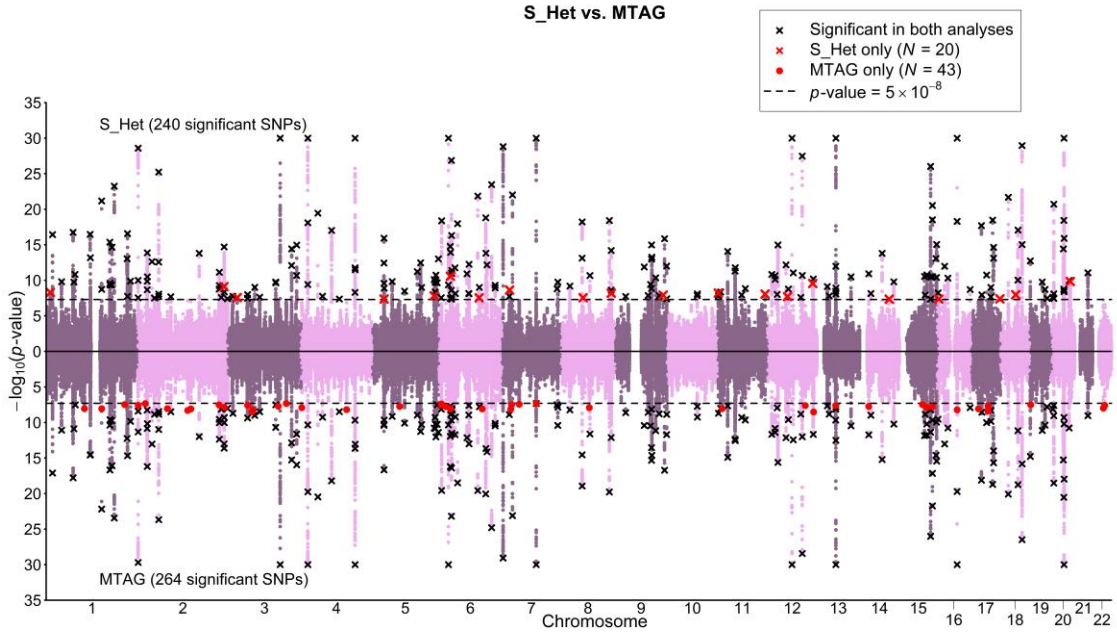
(a)



(b)

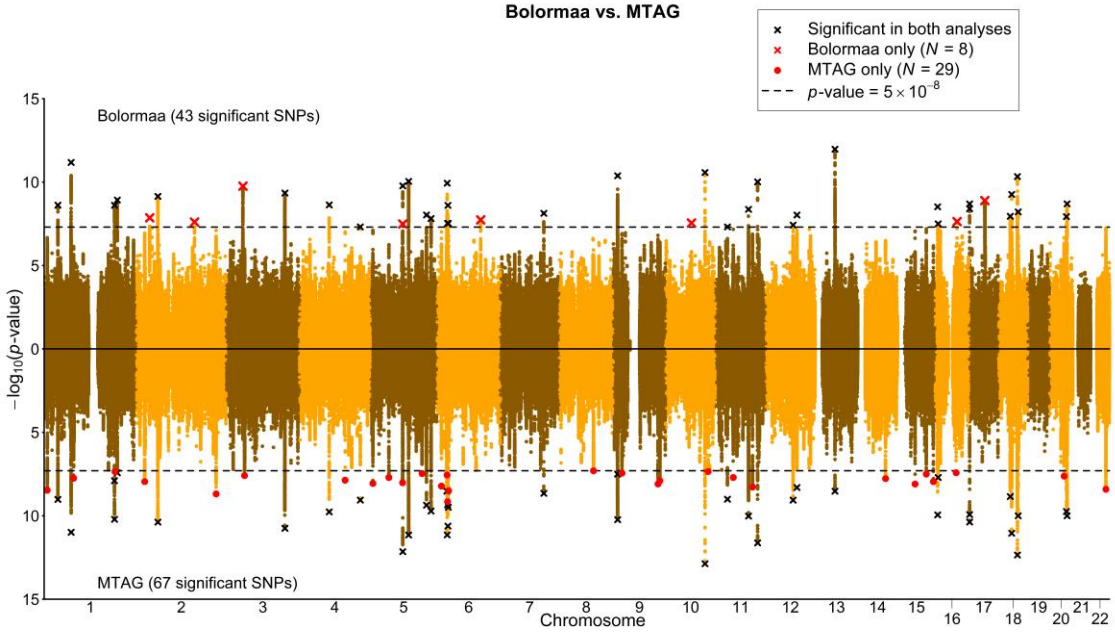

S_Hom vs. MTAG
Significant in both analyses
S_Hom only (N = 1)
MTAG only (N = 66)
p-value = 5 × 10^-8
S_Hom (2 significant SNPs)
MTAG (67 significant SNPs)
−log10(p-value)
Chromosome

(c)

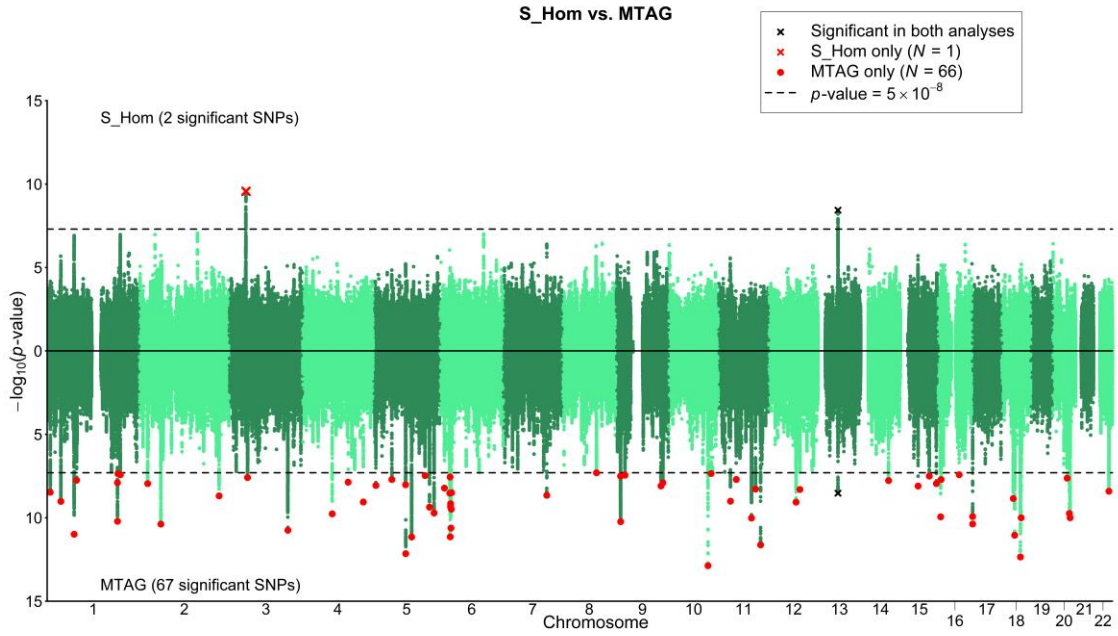

S_Het vs. MTAG
Significant in both analyses
S_Het only (N = 20)
MTAG only (N = 43)
p-value = 5 × 10^-8
S_Het (240 significant SNPs)
MTAG (264 significant SNPs)
−log10(p-value)
Chromosome

Supplementary Figure 13. Results of comparison of MTAG to Bolormaa, S_{hom}, and S_{het} in DEP, NEUR and SWB data. (**a**) MTAG vs. Bolormaa, (**b**) MTAG vs. S_{hom}, (**c**) MTAG vs. S_{Het}. For all multi-trait methods, the *P* value for each SNP tests the null hypothesis that the SNP is null for all three traits. *P* values for MTAG are the Bonferroni-adjusted minimum *P* value for the SNP across all six traits. Below the *x*-axis in each figure is a Manhattan plot for the MTAG results; above the *x*-axis is a Manhattan plot for the alternative method. MTAG hits that are in LD with a hit identified by the alternative method (or hits from the alternative method that are in LD with a hit found by MTAG) are marked with black x's. Significant SNPs identified by the alternative analysis that are not in LD with any significant MTAG SNPs are marked with red x's. Significant MTAG SNPs not in LD with any significant SNP in the alternative analysis are marked with red dots.

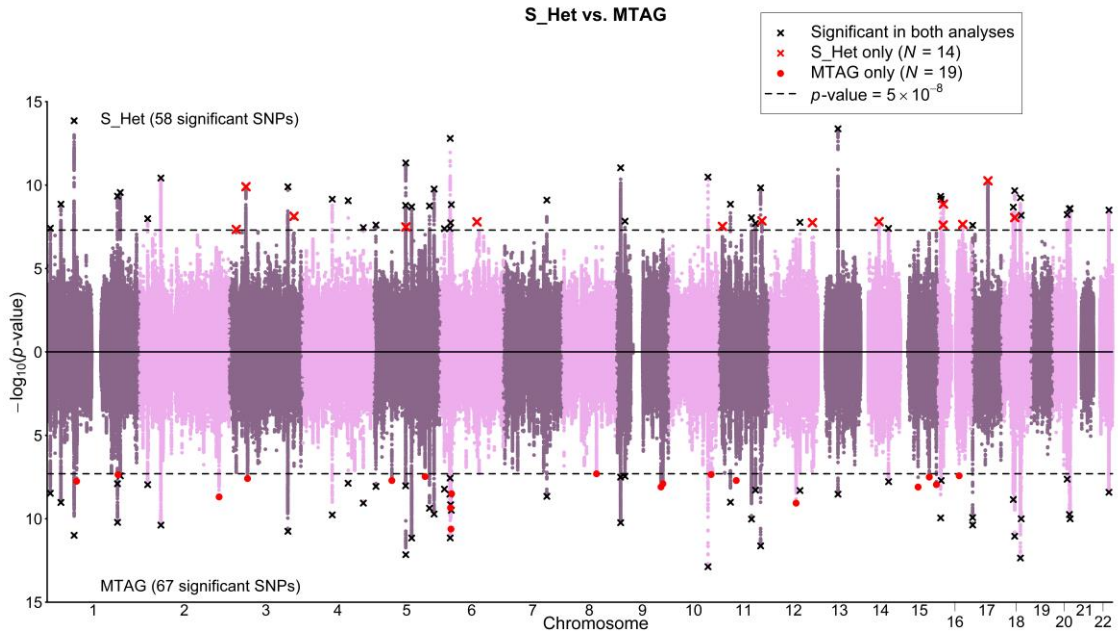
(a)



(b)


S_Hom vs. MTAG
Significant in both analyses
S_Hom only (N = 1)
MTAG only (N = 66)
p-value = 5 × 10^-8
S_Hom (2 significant SNPs)
MTAG (67 significant SNPs)
−log10(p-value)
Chromosome

(c)


S_Het vs. MTAG
Significant in both analyses
S_Het only (N = 14)
MTAG only (N = 19)
p-value = 5 × 10^-8
S_Het (58 significant SNPs)
MTAG (67 significant SNPs)
−log10(p-value)
Chromosome

Supplementary Table 4. Quality Control Filters

		Imputation Quality				
	MAC	MaCH	IMPUTE	PLINK	HWE *P*-value	Call rate
$N < 1000$	30	0.6	0.7	0.8	10^{-3}	95%
$1000 < N < 2000$	30	0.6	0.7	0.8	10^{-4}	95%
$2000 < N > 10000$	30	0.6	0.7	0.8	10^{-5}	95%
$N > 10000$	30	0.6	0.7	0.8	None	95%

Note: N is the sample size per SNP. HWE is Hardy Weinberg equilibrium. HWE p-value filter is applied only to genotyped SNPs. Call rate is the minimum fraction of subjects for which the association results for a SNP must be available in order for the SNP to be included. MAC is minor allele count.

Supplementary Table 6. LDSC Estimates of SNP-based Heritability

Phenotype	*N*	*SNPs (M)*	h^2	SE	λ_{GC}	Mean $\chi 2$	Intercept	SE	Ratio	SE
DEP	354,862	1,143,276	0.0639	0.0031	1.3581	1.4515	1.0129	0.0096	0.0285	0.0213
NEUR	168,105	1,130,133	0.0973	0.0080	1.2431	1.3171	0.9896	0.0117	-	-
SWB	388,535	1,138,093	0.0412	0.0020	1.2798	1.3276	1.0157	0.0082	0.0480	0.0251

Note: Output of various statistics from LD score regressions with the summary statistics from GWAS's. The statistics were computed using the LDSC python software package [1] and the "eur_w_ld_chr/" files of LD Scores calculated by Finucane et al. [2]. For all phenotypes, the reported heritabilities are on the observed scale (not liability scale). In the LD Score regressions, we include only HapMap3 SNPs with MAF > 0.01, and the standard errors of the LD Score regressions are estimated using a block jackknife over SNPs (by the LDSC software). The ratio estimate is not applicable when the intercept is < 1.

References

[1] Bulik-Sullivan, B. K. *et al* ., LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

[2] H. K. Finucane *et al.* , Partitioning heritability by functional category using GWAS summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).

Supplementary Table 7. LDSC Estimates of Genetic Correlations

	DEP	NEUR	SWB	SWB_excl_HRS	SWB_HRS
DEP	**1.000**	0.7166 (0.0264)	-0.6850 (0.0239)	-0.6922 (0.0242)	-0.5247 (0.2927)
NEUR	0.7166 (0.0264)	**1.000**	-0.6715 (0.0272)	-0.6737 (0.0277)	-0.6974 (0.4662)
SWB	-0.6850 (0.0239)	-0.6715 (0.0272)	**1.000**	1.0025 (0.0004)	1.1482 (0.5730)
SWB_excl_HRS	-0.6922 (0.0242)	-0.6737 (0.0277)	1.0025 (0.0004)	**1.000**	1.1842 (0.6059)
SWB_HRS	-0.5247 (0.2927)	-0.6974 (0.4662)	1.1482 (0.5730)	1.1842 (0.6059)	**1.000**

Note : Genetic correlation estimates from LD score regression [1] based on the meta-analyses used in this study. The "SWB_excl_HRS" meta-analysis corresponds to a meta-analysis of the SWB cohorts, omitting the Health and Retirement Study. "SWB_HRS" corresponds to the results from the Health and Retirement Study. Standard errors are in parentheses.

References

[1] Bulik-Sullivan, B. *et al* . An atlas of genetic correlations across human diseases and traits. *Nat. Genet* . 47, 1236–1241

Supplementary Table 8. LDSC Estimates of Intercepts

	DEP	NEUR	SWB	SWB_excl_HRS	SWB_HRS
DEP	**1.0129 (0.0213)**	0.2890 (0.0074)	-0.1082 (0.0057)	-0.1083 (0.0058)	-0.0031 (0.0047)
NEUR	0.2890 (0.0074)	**0.9896 (0.0117)**	-0.0905 (0.0067)	-0.0903 (0.0066)	-0.0050 (0.0050)
SWB	-0.1082 (0.0057)	-0.0905 (0.0067)	**1.0157 (0.0082)**	1.0026 (0.0082)	0.0876 (0.0049)
SWB_excl_HRS	-0.1083 (0.0058)	-0.0903 (0.0066)	1.0026 (0.0082)	**1.0173 (0.0082)**	-0.0063 (0.0048)
SWB_HRS	-0.0031 (0.0047)	-0.0050 (0.0050)	0.0876 (0.0049)	-0.0063 (0.0048)	**1.0109 (0.0061)**

Note : Intercepts estimates from bivariate LD score regression [1] based on the meta-analyses used in this study. Diagonal elements correspond to intercept estimated from univariate LD score regression. The "SWB_excl_HRS" meta-analysis corresponds to a meta-analysis of the SWB cohorts, omitting the Health and Retirement Study. "SWB_HRS" corresponds to the results from the Health and Retirement Study. Standard errors are in parentheses.

References

[1] Bulik-Sullivan, B. *et al* . An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* 47, 1236–1241

Supplementary Table 10. MTAG Filters Applied to Summary Statistics

					A: Main	B: Replication	C: Prediction	D: Prediction Pruned
	GWAS filters	Excluding inversion	MAF > 0.01	N filter	Intersection of SNPs		Hapmap3	Pruned Hapmap3
DEP	11,764,401		8,356,688	7,601,067	6,107,952	-	-	-
NEUR	9,428,194	9,409,403	8,348,642	6,238,079	6,107,952	6,107,941	1,080,892	54,068
SWB	10,849,064		8,585,637	7,606,986	6,107,952	6,107,941	1,080,892	54,068
SWB (no HRS)	10,765,546		8,587,763	7,606,149	-	6,107,941	1,080,892	54,068
				$\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma_{LD}}$ estimation:	5,168,052	5,168,050	1,080,892	54,068

Note: Numbers indicate the number of SNPs (*M*) remaining after applying the relevant MTAG filter to summary statistics. GWAS filters corresponds to the standard QC filters described in Supplementary Note. The inversion region excluded corresponds to the region on chromosome 8 between base-pair locations 7,962,590 and 11,962,591. The N filter removes SNPs that are estimated in a sample of less than 75% of the 90 percentile of the sample size distribution. Strand-ambiguous SNPs are additionally excluded for the $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma_{LD}}$ estimation steps. MAF is minor allele frequency.

Supplementary Table 11. Summary of GWAS and MTAG Results

Panel A. Main Analysis					
Phenotype	*N*	*SNPs (M)*	GWAS mean χ^2	MTAG mean χ^2	GWAS equiv. *N*
DEP	354,861	6,107,952	1.434	1.550	449,649
NEUR	168,105	6,107,952	1.287	1.446	260,897
SWB	388,538	6,107,952	1.302	1.466	600,834

Panel B. Replication Analysis (HRS omitted)					
Phenotype	*N*	*SNPs (M)*	GWAS mean χ^2	MTAG mean χ^2	GWAS equiv. *N*
DEP	354,861	6,107,941	1.434	1.549	449,250
NEUR	168,105	6,107,941	1.287	1.443	259,327
SWB_excl_HRS	378,596	6,107,941	1.291	1.460	598,352

Panel C. Prediction Analysis					
Phenotype	*N*	*SNPs (M)*	GWAS mean χ^2	MTAG mean χ^2	GWAS equiv. *N*
DEP	354,861	1,080,893	1.456	1.572	445,460
NEUR	168,105	1,080,893	1.301	1.464	259,341
SWB_excl_HRS	378,596	1,080,893	1.299	1.475	601,670

Panel D. Prediction Analysis (Pruned)					
Phenotype	*N*	*SNPs (M)*	GWAS mean χ^2	MTAG mean χ^2	GWAS equiv. *N*
DEP	354,861	54,069	1.219	1.285	462,038
NEUR	168,105	54,069	1.173	1.247	240,150
SWB_excl_HRS	378,596	54,069	1.870	1.268	542,054

Note: Comparison of mean $\chi 2$ statistics from GWAS and MTAG analyses. Panel A reports the results based on the main analysis. Panel B reports results based on GWAS summary statistics that omit the HRS. Panel C reports results based on GWAS summary statistics that omit the HRS and are restricted to SNPs that are genotyped in the HRS. Panel D reports results based on a pruned set of SNPs that are genotyped in the HRS. GWAS-equivalent sample size (GWAS equiv. *N*) is computed as described in the **Online Methods**.

Supplementary Table 12. Estimates of Σ_{LD}

Panel A. Main Analysis			
	DEP	NEUR	SWB
DEP	**1.016**	0.291	-0.11
NEUR	0.291	**0.998**	-0.096
SWB	-0.11	-0.096	**1.021**

Panel B. Replication Analysis			
	DEP	NEUR	SWB
DEP	**1.016**	0.291	-0.109
NEUR	0.291	**0.998**	-0.096
SWB	-0.109	-0.096	**1.021**

Panel C. Prediction Analysis			
	DEP	NEUR	SWB
DEP	**1.016**	0.291	-0.109
NEUR	0.291	**0.998**	-0.096
SWB	-0.109	-0.096	**1.021**

Panel D. Prediction Analysis (Pruned)			
	DEP	NEUR	SWB
DEP	**1.033**	0.304	-0.112
NEUR	0.304	**0.988**	-0.096
SWB	-0.112	-0.096	**1.005**

Note: The estimate of Σ_{LD} used in the MTAG analysis for different samples. These matrices were estimated using bivariate LD score regression[1]. Panel A reports the results based on the main analysis. Panel B reports results based on GWAS summary statistics that omit the HRS. Panel C reports results based on GWAS summary statistics that omit the HRS and are restricted to SNPs that are genotyped in the HRS. Panel D reports results based on a pruned set of SNPs that are genotyped in the HRS.

References

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

Supplementary Table 13. Estimates of Ω

Panel A. Main analysis

	DEP	NEUR	SWB
DEP	**1.26E-06**	1.09E-06	-7.02E-07
NEUR	1.09E-06	**1.76E-06**	-7.84E-07
SWB	-7.02E-07	-7.84E-07	**7.99E-07**

Panel B. Replication Analysis

	DEP	NEUR	SWB
DEP	**1.26E-06**	1.09E-06	-7.06E-07
NEUR	1.09E-06	**1.76E-06**	-7.80E-07
SWB	-7.06E-07	-7.80E-07	**7.92E-07**

Panel C. Prediction Analysis

	DEP	NEUR	SWB
DEP	**1.32E-06**	1.15E-06	-7.31E-07
NEUR	1.15E-06	**1.84E-06**	-8.20E-07
SWB	-7.31E-07	-8.20E-07	**8.14E-07**

Panel D. Prediction Analysis (Pruned)

	DEP	NEUR	SWB
DEP	**6.54E-07**	5.65E-07	-3.82E-07
NEUR	5.65E-07	**1.06E-06**	-4.55E-07
SWB	-3.82E-07	-4.55E-07	**5.07E-07**

Note: The estimate of **Ω** from each MTAG analysis. Panel A reports the results based on the main analysis. Panel B reports results based on GWAS summary statistics that omit the HRS. Panel C reports results based on GWAS summary statistics that omit the HRS and are restricted to SNPs that are genotyped in the HRS. Panel D reports results based on a pruned set of SNPs that are genotyped in the HRS.

Supplementary Table 15. LDSC Intercept Estimates between Discovery and Replication Summary Statistics

		Discovery		
		DEP	NEUR	SWB
Replication	DEP	-0.0002 (0.0057)	0.0019 (0.0051)	-0.0023 (0.0048)
	NEUR	0.0009 (0.0048)	0.0028 (0.0053)	-0.0066 (0.0049)
	SWB	0.0057 (0.0054)	0.0034 (0.0049)	-0.005 (0.0047)

Note : Intercepts estimates from bivariate LD score regression [1] based on the meta-analyses used in discovery and replication (or prediction) analyses in this study. A non-zero intercept implies some potential sample overlap. The repliction sample includes results from the Health and Retirement Study and Add Health cohorts. Discovery sample estimates for SWB omit HRS, which is the only GWAS that includes it in the main MTAG results reported. Standard errors are in parentheses.

References

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

Supplementary Table 16. Replication of Genome-Wide Significant MTAG Loci

Panel A: DEP			
	HRS and Add Health	HRS	Add Health
Beta	0.884	0.787	1.070
SE	(0.216)	(0.278)	(0.352)
SNPs	59	59	59

Panel B: NEUR			
	HRS and Add Health	HRS	Add Health
Beta	0.756	0.913	0.466
SE	(0.213)	(0.266)	(0.409)
SNPs	37	37	37

Panel C: SWB			
	HRS and Add Health	HRS	Add Health
Beta	0.991	1.263	0.503
SE	(0.335)	(0.460)	(0.460)
SNPs	43	43	43

Note : Slope (Beta) of regressing the replication effect size on winner's curse adjusted MTAG effect size estimates, constraining the intercept to 0. SNPs is the number of significant SNPs included in the regression. For the winner's curse adjustment, the effect sizes of each trait are fit to a spike-and-slab distribution through maximum likelihood estimation as described in the **Online Methods**. For DEP, NEUR, SWB we estimate the fraction of SNPs with null effects (pi_null) to be 0.598, 0.652, and 0.633 and the variance of standardized effect sizes (tau^2) to be 3.12e-06, 5.05e-6, and 2.15e-6. Standard errors (SE) are in parentheses.

Supplementary Table 17. Incremental R^2 and Change in Incremental R^2

Panel A. Prediction in HRS

	DEP (N=8,307)	NEUR (N=8,197)	SWB (N=6,857)
Incremental R^2			
GWAS	0.873% (0.510% to 1.313%)	1.330% (0.868% to 1.871%)	1.482% (0.954% to 2.095%)
MTAG	1.036% (0.653% to 1.486%)	1.603% (1.103% to 2.168%)	1.788% (1.226% to 2.396%)
Change in Incremental R^2			
MTAG vs. GWAS	0.163% (0.022% to 0.315%)	0.273% (0.017% to 0.552%)	0.306% (-0.052% to 0.639%)
Expected Change	0.192%	0.581%	0.418%

Panel B. Prediction in Add Health

	DEP (N=4,334)	NEUR (N=4,332)	SWB (N=3,673)
Incremental R^2			
GWAS	1.233% (0.688% to 1.889%)	1.168% (0.586% to 1.887%)	0.658% (0.228% to 1.218%)
MTAG	1.430% (0.833% to 2.178%)	1.742% (1.006% to 2.566%)	1.170% (0.577% to 1.923%)
Change in Incremental R^2			
MTAG vs. GWAS	0.197% (-0.069% to 0.465%)	0.575% (0.223% to 0.999%)	0.512% (0.177% to 0.946%)
Expected Change	0.251%	0.524%	0.267%

Panel C. Meta-analyzed Prediction Results in HRS and Add Health

	DEP (N=12,641)	NEUR (N=12,529)	SWB (N=10,530)
Incremental R^2			
GWAS	0.997% (0.693% to 1.359%)	1.274% (0.897% to 1.676%)	1.195% (0.812% to 1.636%)
MTAG	1.171% (0.843% to 1.564%)	1.651% (1.236% to 2.114%)	1.573% (1.151% to 2.069%)
Change in Incremental R^2			
MTAG vs. GWAS	0.175% (0.043% to 0.313%)	0.377% (0.156% to 0.610%)	0.378% (0.123% to 0.648%)
Expected Change	0.214%	0.562%	0.385%

Note: Incremental R^2 is the increase in R^2 from adding the polygenic score as a covariate to a regression of the phenotype which includes covariates for year of birth, year of birth squared, sex, their interaction, and ten principal components. 95% confidence intervals for the incremental R^2, computed from 1,000 bootstrap replicates as described in the **Online Methods**, are reported in parentheses. Expected Change is defined by the anticipated increase in incremental R^2 given the observed increase in mean χ^2-statistic and estimated heritability. All scores estimated in meta-analyses excluding HRS.

Supplementary Table 18. Incremental R^2 and Change in Incremental R^2 (Pruned SNPs)

Panel A. Prediction in HRS

	DEP (N=8,307)	NEUR (N=8,197)	SWB (N=6,857)
Incremental R^2			
GWAS	0.299% (0.109% to 0.523%)	0.410% (0.171% to 0.694%)	0.410% (0.152% to 0.712%)
MTAG	0.336% (0.138% to 0.570%)	0.451% (0.197% to 0.735%)	0.539% (0.256% to 0.893%)
Change in Incremental R^2			
MTAG vs. GWAS	0.037% (-0.068% to 0.142%)	0.041% (-0.108% to 0.204%)	0.130% (-0.050% to 0.319%)

Panel B. Prediction in Add Health

	DEP (N=4,334)	NEUR (N=4,332)	SWB (N=3,673)
Incremental R^2			
GWAS	0.544% (0.192% to 1.005%)	0.625% (0.229% to 1.094%)	0.358% (0.072% to 0.819%)
MTAG	0.663% (0.260% to 1.168%)	0.963% (0.459% to 1.572%)	0.512% (0.145% to 1.048%)
Change in Incremental R^2			
MTAG vs. GWAS	0.119% (-0.102% to 0.350%)	0.338% (0.069% to 0.638%)	0.153% (-0.060% to 0.415%)

Panel C. Meta-analyzed Prediction Results in HRS and Add Health

	DEP (N=12,641)	NEUR (N=12,529)	SWB (N=10,530)
Incremental R^2			
GWAS	0.383% (0.188% to 0.605%)	0.484% (0.271% to 0.738%)	0.392% (0.189% to 0.655%)
MTAG	0.448% (0.233% to 0.697%)	0.628% (0.391% to 0.902%)	0.530% (0.268% to 0.846%)
Change in Incremental R^2			
MTAG vs. GWAS	0.065% (-0.036% to 0.172%)	0.143% (-0.005% to 0.293%)	0.138% (-0.003% to 0.300%)

Note: Incremental R^2 is the increase in R^2 from adding the polygenic score as a covariate to a regression of the phenotype which includes covariates for year of birth, year of birth squared, sex, their interaction, and ten principal components. 95% confidence intervals for the incremental R^2, computed from 1,000 bootstrap replicates as described in the **Online Methods**, are reported in parentheses. Expected Change is defined by the anticipated increase in incremental R^2 given the observed increase in mean χ^2-statistic and estimated heritability. All scores estimated in meta-analyses excluding HRS.

Supplementary Table 30. Summary of DEPICT Results

	DEP			NEUR			SWB		
	GWAS	MTAG	GWAS (Constant SNPs)	GWAS	MTAG	GWAS (Constant SNPs)	GWAS	MTAG	GWAS (Constant SNPs)
Prioritized genes	3	72	51	0	51	12	0	0	0
Enriched gene sets	0	347	127	0	1	1	0	7	0
Enriched tissues	10	22	19	0	21	14	0	12	0

Note : Comparison of the number of statistically significant prioritized genes, enriched gene sets, and significant tissues/cell types yielded by DEPICT. We used DEPICT FDR < 0.05 as the criterion of output statistical significance. We compare the number of significant results yielded by: the standard GWAS approach with the DEPICT default P < 1E-5 used to define lead SNPs; MTAG with the DEPICT default P < 1E-5 used to define lead SNPs; and the standard GWAS approach with the SNP significance threshold relaxed to yield approximately as many lead SNPs as MTAG.

Supplementary Table 31. Replication of Loci Identified in Multitrait Analyses of GIANT Phenotypes

Panel 1: Replication rate with GIANT phenotypes across all hits

		N loci replicated		
	N loci	p < .01	p < .001	p < 5E-8
MTAG	264	234	206	108
		88.64%	78.03%	40.91%
Bolormaa	161	151	137	88
		93.79%	85.09%	54.66%
S_Het	240	218	194	102
		90.83%	80.83%	42.50%
S_Hom	86	81	66	54
		94.19%	76.74%	62.79%

Panel 2: Replication rate with GIANT phenotypes across distinct hits

		N loci replicated		
	N loci	p < .01	p < .001	p < 5E-8
MTAG not Bolormaa	102	81	64	17
		79.41%	62.75%	16.67%
Bolormaa not MTAG	9	6	5	0
		66.67%	55.56%	0.00%
MTAG not S_Het	43	31	22	5
		72.09%	51.16%	11.63%
S_Het not MTAG	20	13	8	0
		65.00%	40.00%	0.00%
MTAG not S_Hom	181	154	128	52
		85.08%	70.72%	28.73%
S_Hom not MTAG	12	8	4	0
		66.67%	33.33%	0.00%

Notes : GIANT phenotypes include body mass index (BMI), height, and waist to hip ratio adjusted for BMI (WHRadjBMI). Discovery summary statistics are from Randall et al. 2013. Replication summary statistics are calculated by comparing earlier summary statistics (Randall et al. 2013) to later summary statistics (Locke et al. 2015 [BMI]; Wood et al. 2014 [height]; Shungin et al. 2015 [WHRadjBMI]). Replication *P* values for MTAG are defined as 3*minimum(p[BMI], p[height], p[WHRadjBMI]).