MTAG: Multi-Trait Analysis of GWAS

Authors: All authors and their affiliations appear at the end of the paper*

*Correspondence to P.T. (paturley@broadinstitute.org), B.N. (bneale@broadinstitute.org), D.C. (dac12@nyu.edu), or D.J.B. (daniel.benjamin@gmail.com).

ABSTRACT

We introduce Multi-Trait Analysis of GWAS (MTAG), a method for the joint analysis of summary statistics from GWASs of different traits, possibly from overlapping samples. We demonstrate MTAG using data on depressive symptoms (N_{eff} = 354,862), neuroticism (N = 168,105), and subjective well-being (N = 388,538). Compared to 32, 9, and 13 genome-wide significant loci in the single-trait GWASs (most of which are novel), MTAG increases the number of loci to 74, 66, and 60, respectively. Moreover, the association statistics from MTAG yield more informative bioinformatics analyses and, consistent with theoretical calculations, improve prediction accuracy by approximately 25%.

INTRODUCTION

The standard approach in genetic association studies is to meta-analyze association statistics from cohort-level genome-wide association analyses of a single trait. Such single-trait analyses do not exploit information that may be available from genome-wide association studies (GWAS) of other, correlated traits. In this paper, we develop a method, Multi-Trait Analysis of GWAS (MTAG), which jointly analyzes GWAS results for several related traits, thereby boosting statistical power to detect genetic associations for *each* of the traits.

MTAG is a generalization of standard, inverse-variance-weighted meta-analysis. The estimator takes summary statistics from at least two single-trait GWASs and, after jointly analyzing these, outputs trait-specific association statistics. The set of results for each trait represents an optimal combination of the information from the single-trait summary statistics. The resulting *P* values can be interpreted like GWAS *P* values and used in the usual ways, for example, to prioritize SNPs for subsequent analyses such as biological annotation.

MTAG has several features that make it useful in many practical applications where existing multivariate methods may be infeasible. First, it can be applied without access to individual-level data, only requiring GWAS summary statistics from studies of quantitative or binary traits. Second, the summary statistics need not come from independent discovery samples: MTAG uses bivariate linkage disequilibrium (LD) score regression[1] to account for (possibly unknown) sample overlap between the GWAS results for different traits. Third, running MTAG does not require powerful computational resources.

We examine the performance of MTAG in a range of simulation scenarios, and then we apply it to GWAS meta-analysis results for three traits: depressive symptoms (DEP, N_{eff} = 354,862), neuroticism (NEUR, N = 168,105), and subjective well-being (SWB, N = 388,538). We compare the results from MTAG to the results from standard GWAS along three dimensions: the power to detect associations with individual single-nucleotide polymorphisms (SNPs), the predictive power of a polygenic score constructed from all available SNPs, and the findings from applying the bioinformatics tool DEPICT[2] to the results. We show theoretically that the gains in statistical power from these MTAG analyses are equivalent to increasing the sample size of the original, single-trait GWASs by 37% (DEP), 96% (NEUR), and 85% (SWB). The corresponding gains in predictive power of a polygenic score are consistent with these theoretical expectations.

The standard errors of MTAG are calculated under the assumption that the correlation across traits in the effect size of a SNP is the same across all SNPs. In our simulations, we examine some scenarios in which—in violation of this assumption—a fraction of SNPs are related to one trait but not another. In these scenarios, MTAG generates anti-conservative standard errors and thus inflated test statistics, especially for such SNPs. Therefore, when applying MTAG in settings where the correlation in effect sizes across traits is expected to be systematically

different for some SNPs than for others, we advise caution in interpreting associations with individual SNPs. Even in such settings, however, MTAG generates consistent effect-size estimates and hence is expected to produce gains in prediction accuracy.

The closest existing method we are aware of is a Bayesian analog of MTAG derived for the special case of no sample overlap[3]. Multivariate methods have also been developed to enable identification of SNPs associated with some but not all of the traits[4–6] and to discriminate between colocalization and different models of pleiotropy[7–10]. Yet another family of methods tests the joint hypothesis of a zero effect on every trait against the alternative of association with at least one of the traits, sometimes by combining multiple traits into a single variable prior to analysis[11–13]. A recent, independent paper combines DEP, NEUR, and SWB into GWAS analyses using a subset of the data we use here[14].

RESULTS

Overview of MTAG

The key idea underlying MTAG is that when GWAS estimates from two different traits are correlated, the effect estimates for each trait can be improved by appropriately incorporating information contained in the GWAS estimates for the *other* traits.

Correlation between GWAS estimates can arise for two reasons. First, the traits may be genetically correlated, in which case the *true effects* of the single-nucleotide polymorphisms (SNPs) are correlated across traits. We standardize all traits and genotypes to have mean zero and variance one. For SNP j, we denote the vector of marginal (i.e., not controlling for other SNPs), true effects on each of the T traits by $\boldsymbol{\beta}_j$. We assume that these true effects follow an infinitesimal model and treat them as random effects: $\boldsymbol{\beta}_j \sim N(\mathbf{0}, \boldsymbol{\Omega})$, where $\boldsymbol{\Omega}$ is the genetic variance-covariance matrix of the effect sizes across traits. Correlation in the true effects means that the off-diagonal elements of $\boldsymbol{\Omega}$ are non-zero. The matrix $\boldsymbol{\Omega}$ does not depend on the SNP j, meaning that it is the variance-covariance matrix calculated across all SNPs.

Second, the *estimation error* of the SNPs' effects may be correlated across traits. Such correlation will occur 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. We denote the vector of estimates of SNP j's effects on the traits by $\widehat{\boldsymbol{\beta}}_j$. As long as the GWAS of each trait was conducted in a large sample, its distribution can be well approximated as $\widehat{\boldsymbol{\beta}}_j | \boldsymbol{\beta}_j \sim N(\boldsymbol{\beta}_j, \boldsymbol{\Sigma}_j)$, where $\boldsymbol{\Sigma}_j$ is the variance-covariance matrix of the estimation error. The off-diagonal elements of $\boldsymbol{\Sigma}_j$ are non-zero whenever the estimation errors are correlated.

To derive the MTAG estimator of the effect of SNP j on trait t, we derive the marginal likelihood function[15] $f(\widehat{\boldsymbol{\beta}}_j|\beta_{j,t})$ and maximize it with respect to $\beta_{j,t}$. The solution (see **Online Methods**) is a weighted sum of the GWAS estimates:

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

where $\boldsymbol{\omega}_t$ is a vector equal to the tth column of $\boldsymbol{\Omega}$ and ω_{tt} is a scalar equal to the tth diagonal element of $\boldsymbol{\Omega}$. The MTAG estimator is consistent.

For intuition, here we describe how MTAG works for two traits with a positive phenotypic correlation (see **Supplementary Note**). First, if the genetic correlation is zero and there is no sample overlap—that is, neither of the two sources of possible correlation in the GWAS estimates is present—then the MTAG estimates are identical to the GWAS estimates. Second, if the genetic correlation is positive but there is no sample overlap, then the MTAG estimator for the SNP effect on trait 1 will incorporate evidence from the GWAS estimate for trait 2. Finally, if the genetic correlation is zero but the GWAS samples overlap—thereby making the estimation error positively correlated—then the MTAG estimator for the SNP effect of trait 1 will assign a *negative* weight to the GWAS estimate for trait 2. In this circumstance, the GWAS estimate for trait 2 contains information about the estimation error in the GWAS estimate for trait 1 (cf. Evans[16]). These are the two key forces at play with MTAG, but they can interact with each other in complex ways when both sources of correlation are present, when there are more than two traits, and when the GWASs for the traits are differently powered.

There are several useful special cases of MTAG (see **Online Methods**). For example, when all estimates are for the same trait (implying $\frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} = \boldsymbol{\Omega}$ and $\frac{\boldsymbol{\omega}_t}{\omega_{tt}} = \mathbf{1}$), equation (1) simplifies to: $\hat{\beta}_{\mathrm{MTAG},j,t} = \frac{\mathbf{1}'\boldsymbol{\Sigma}_j^{-1}}{\mathbf{1}'\boldsymbol{\Sigma}_j^{-1}\mathbf{1}}\widehat{\boldsymbol{\beta}}_j$. When the GWAS estimates are obtained from non-overlapping samples (i.e., $\boldsymbol{\Sigma}_j$ is diagonal), this case is the well-known formula for inverse-variance-weighted meta-analysis.

To make equation (1) operational, we use consistent estimates of $\boldsymbol{\Sigma}_j$ and $\boldsymbol{\Omega}$. In standard meta-analysis, the diagonal elements of $\widehat{\boldsymbol{\Sigma}}_j$ would be constructed using the squared standard errors from the GWAS results and the off-diagonal elements of $\widehat{\boldsymbol{\Sigma}}_j$ would be set to zero. In MTAG, however, we want to allow the estimation error to include bias (in addition to sampling variation) and to be correlated across the GWAS estimates.

Therefore, MTAG proceeds by running linkage disequilibrium (LD) score regressions[17] on the GWAS results and using the estimated intercepts to construct the diagonal elements of $\widehat{\boldsymbol{\Sigma}}_j$. Next, bivariate LD score regressions[1] are run using

each pair of GWAS results, and the estimated intercepts are used to construct the off-diagonal elements of $\widehat{\boldsymbol{\Sigma}}_j$. Under the assumptions of LD score regression, the resulting matrix $\widehat{\boldsymbol{\Sigma}}_j$ captures all relevant sources of estimation error, including population stratification, unknown sample overlap, and cryptic relatedness. Moreover, there is no need to adjust the results of MTAG using the LD score intercept (as is becoming standard for GWAS meta-analysis results) because the adjustment is already built into the MTAG estimates.

We estimate $\widehat{\boldsymbol{\Omega}}$ by maximum likelihood, using as the likelihood function the marginal distribution $\widehat{\boldsymbol{\beta}}_j \sim N(\mathbf{0}, \boldsymbol{\Sigma}_j + \boldsymbol{\Omega})$, with $\widehat{\boldsymbol{\Sigma}}_j$ substituted in place of $\boldsymbol{\Sigma}_j$. Details and explanations about how $\widehat{\boldsymbol{\Sigma}}_j$ and $\widehat{\boldsymbol{\Omega}}$ are estimated are in **Supplementary Note 1.2**.

In summary, the MTAG results for SNP *j* are obtained in three steps: (i) estimate the variance-covariance matrix of the GWAS estimation error, $\widehat{\boldsymbol{\Sigma}}_j$**,** by using a sequence of LD score regressions, (ii) estimate the variance-covariance matrix of the SNP effects, $\widehat{\boldsymbol{\Omega}}$, using maximum likelihood, and (iii) for each SNP, substitute these estimates into equation (1). We have made available for download a Python command line tool that runs our MTAG estimation procedure (see URLs). Because most of the above steps have closed-form solutions, genome-wide analyses using the MTAG software run quickly (see **Online Methods**).

Simulations

We conduct simulations to investigate the performance of MTAG across 3^5 = 243 different statistical scenarios. For simplicity, we analyze two traits. For each scenario, we generate GWAS results at 100,000 independent SNPs for each trait, and we perform 1,000 replications. The scenarios differ across five dimensions. First, the true effect sizes $\boldsymbol{\beta}_j$ are drawn from one the following distributions: multivariate normal (as assumed in the derivation of MTAG), multivariate *t*-distribution (which has fatter tails), or multivariate spike-and-slab (which has a positive fraction of truly null SNPs). Second and third, we set the expected χ^2-statistic of the GWAS results for each of the two traits to be equal to 1.1 (low power), 1.4 (medium power), or 2.0 (high power). (For comparison, the 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.) Fourth, we set the correlation in SNP effect sizes between the traits equal to 0, 0.35, or 0.7. Fifth, we set the correlation of the estimation error equal to 0, 0.35, or 0.7.

We assess the gain in performance of MTAG over GWAS in each scenario using the root mean squared error (RMSE) of the estimates over all SNPs for each trait. RMSE reflects both the bias and the variance of the effect estimates, but we find that the bias of the MTAG estimates is very close to zero in all scenarios (**Supplementary Tables 2.1**, **2.2**, and **2.3**). In **Supplementary Note** and **Supplementary Figures 2.1**, **2.2**, and **2.3**, we additionally assess the improvements from MTAG in terms of the mean χ^2-statistic and the number of top hits for each trait. The conclusions are the same from all three metrics.

The patterns we will describe hold across all the scenarios we examined, but we illustrate them using a small selection of scenarios (for complete results, see **Supplementary Note**). For each scenario, **Figure 1** shows the RMSE in the first trait's GWAS estimates and in its MTAG estimates.

There are four main results illustrated in **Figure 1**. First, as expected, MTAG generates gains (i.e., reduces RMSE) when the traits are genetically correlated. This can be seen in **Figure 1**'s Scenario A, which we call the "baseline" scenario: effect sizes are drawn from a multivariate normal distribution, effect sizes have a correlation of 0.7, the estimation error has zero correlation, and both GWAS summary statistics have low power.

Second, MTAG improves the estimates even when one or both GWASs are low powered. Relative to the baseline scenario, Scenario B shows results for the scenario in which the GWAS of the first trait has low power and that of the second trait has high power, and Scenario C shows the scenario in which both have high power. The gains are larger in Scenario B.

Third, we confirm the theoretical result that MTAG yields improvements when the traits have correlated estimation error even when they are not genetically correlated. Scenario D shows results for the scenario where, relative to the baseline, we set the correlation of effect sizes to zero and the estimation error correlation to 0.7. The gains from MTAG are roughly equivalent in this scenario as in the baseline.

Fourth, across the alternative effect-size distributions we examined, MTAG reduces the RMSE even when the true effect-size distribution deviates from the assumption of multivariate normality. Scenario E shows the results for the scenario that is the same as the baseline, except that effect sizes are drawn from a spike-and-slab distribution with 60% of true effects being null for the first trait and 80% being null for the second. In this case, the gains from MTAG are smaller than in the baseline scenario but still positive. In fact, in every scenario in which MTAG produces gains over GWAS when effect sizes are normally distributed, MTAG also produces gains in the corresponding scenario with spike-and-slab or *t*-distributed effect sizes.

The simulations also shed light on two possible ways that MTAG could produce anti-conservative standard errors and hence spuriously inflated χ^2-statistics. First, MTAG does not take into account the estimation error in $\widehat{\mathbf{\Sigma}}_j$ and $\widehat{\mathbf{\Omega}}$. Across the range of scenarios we examined, however, this bias was minimal. Second, the correlation of the effect sizes across traits may be systematically different for some SNPs than others, in which case the standard error does not take into account this source of variation across SNPs (see **Online Methods**). We study an extreme case: a spike-and-slab scenario where a fraction of SNPs is null for trait 1 and non-null for trait 2. In this case, MTAG is expected to produce an inflated Type I error rate for the effects of the null SNPs on trait 1. In our simulations, this bias can be substantial, particularly if the GWAS is lower-powered for the trait for which the SNP is null (**Supplementary Note** and **Supplemental Table 2.4**). We conclude that MTAG can generate anti-conservative standard errors when the correlation of effect sizes

across traits varies systematically across SNPs. However, as noted above, the MTAG point estimates are consistent and have lower RMSE (taken over all SNPs) in all the scenarios we considered. We may therefore expect improvements in prediction accuracy even in settings where single-SNP associations should be interpreted with caution.

GWAS Summary Statistics for Depression, Neuroticism, and Subjective Well-Being

For our empirical application of MTAG, we build on a recent study by the Social Science Genetic Association Consortium (SSGAC) of three traits that have been found to be highly polygenic and strongly genetically related: depressive symptoms (DEP), neuroticism (NEUR), and subjective well-being (SWB). Those analyses combined data from the largest previously published studies[18–21] with new genome-wide analyses from the genetic testing company 23andMe, Inc. and the first release of the UK Biobank (UKB) data. Relative to the previous SSGAC study, we reran the association analyses in UKB using a slightly revised analysis protocol, and much more importantly, we expanded the SSGAC meta-analyses for DEP and SWB. For DEP, we added the results from a recently published GWAS of depression in a large 23andMe cohort[21], and for SWB, we added new association analyses of SWB in a 23andMe cohort. As depicted in **Figure 2**, there is substantial overlap between the estimation samples for the three traits. For additional details, see **Online Methods** and **Supplementary Note**.

MTAG Results

To enable a fair comparison between the MTAG and GWAS results, we restrict all analyses only to those SNPs that are available in all three GWASs and that survive our recommended filters for MTAG. These filters include omitting SNPs with low minor allele frequency, SNPs with small sample size, and SNPs in genomic regions with uncharacteristically large effect sizes. Details and explanations for these restrictions are in **Online Methods**.

For each trait, we assess the gain in average power from MTAG relative to the GWAS results by the increase in the mean χ^2-statistic. We use this increase to calculate how much larger the GWAS sample size would have to be to attain an equivalent increase in expected χ^2 (**Online Methods**). We find that the joint analysis of DEP, NEUR, and SWB via MTAG yielded gains equivalent to augmenting the original samples sizes by 37%, 96%, and 85%, respectively. The resulting GWAS-equivalent sample sizes are thus 487,603 for DEP, 329,835 for NEUR, and 718,284 for SWB.

Figure 3 shows side-by-side Manhattan plots from the GWAS and MTAG analyses for each trait. Approximately independent genome-wide significant SNPs, hereafter “lead SNPs,” were defined by clumping (**Online Methods**). From GWAS to MTAG,

the number of lead SNPs increases from 32 to 74 for DEP, from 9 to 66 for NEUR, and from 13 to 60 for SWB. A list of all clumped SNPs with a *P* value less than 10^{-5} is in **Supplementary Table 4.1**.

Replication of MTAG-identified Loci

To test the lead SNPs for replication, we use the Health and Retirement Study (HRS) as our prediction sample because it contains high-quality measures of DEP ($N = 8{,}307$), NEUR ($N = 8{,}197$), and SWB ($N = 6{,}857$). Because HRS was included in the SSGAC discovery sample for SWB, we re-ran the GWAS and MTAG analyses for SWB after omitting it. Although replicating the lead SNPs individually would be underpowered, we are well powered to test them as a group: for the set of MTAG-identified lead SNPs for each trait, we regressed the effect sizes in HRS on the MTAG effect sizes, after correcting the MTAG effect-size estimates for the winner's curse (**Online Methods**). The regression slopes are 0.81 (s.e. = 0.24) for DEP, 0.79 (s.e. = 0.22) for NEUR, and 0.90 (s.e. = 0.42) for SWB. In all cases, the slope is statistically significantly greater than zero (one-sided $P = 6.28 \times 10^{-4}$, 3.22×10^{-4}, and 1.67×10^{-2}, respectively) and not statistically distinguishable from one, indicating that the lead SNPs replicate in an independent sample.

Polygenic Prediction

We next compare the predictive power of polygenic scores constructed from the GWAS and MTAG results. Each score is constructed as the weighted sum of SNP genotypes using all SNPs in the analysis. These weights are calculated by passing the GWAS or MTAG estimates into the software LDpred[22], which corrects for LD across SNPs. We again use the HRS as our prediction sample (and we obtain the SNP effect estimates for SWB from the analyses that omit it from the discovery sample). In addition to our filters above, the SNPs used in the polygenic score are further restricted to genotyped SNPs that meet a number of other strict quality control filters in the HRS data (see **Supplementary Note**).

We measure the predictive power of each polygenic score by its incremental R^2, defined as the increase in coefficient of determination (R^2) as we move from a regression of the trait only on a set of controls (year of birth, year of birth squared sex, their interactions, and 10 principal components of the genetic data) to a regression that additionally includes the polygenic score as an independent variable. We use the bootstrap to estimate confidence intervals for the incremental R^2's as well as for the differences in incremental R^2 across polygenic scores.

Figure 4 and **Table 1** summarize the results. The GWAS-based polygenic scores have incremental R^2's of 0.81% for DEP, 1.28% for NEUR, and 1.46% for SWB. The corresponding MTAG-based polygenic scores all have greater predictive power: 1.00% for DEP, 1.58% for NEUR, and 1.84% for SWB. The proportional

improvement in incremental R^2 is roughly 25% for each trait. In each case, the 95% confidence interval of the difference excludes a gain of zero. The improvements in R^2 are close to those we would expect theoretically based on the observed increases in mean χ^2-statistics (see **Online Methods**).

Biological Annotation

For a final comparison, we analyze both the GWAS and MTAG results using DEPICT[2]. We present the prioritized genes, enriched gene sets, and enriched tissues identified by DEPICT at the standard false discovery rate threshold of 5%.

Table 1 summarizes the results (see **Supplemental Tables 6.1 to 6.9** for the complete set of findings). In the GWAS-based analysis, very little enrichment is apparent. For DEP, 4 genes are identified, but no gene sets or tissues. For NEUR and SWB, no genes, gene sets, or tissues are identified. In contrast, the MTAG-based analysis is more informative. The strongest results are again for DEP, now with 99 genes, 351 gene sets, and 23 tissues. For NEUR, there are 7 genes, 1 gene set, and 13 tissues, and for SWB, 8 genes, 9 gene sets, and 14 tissues.

For brevity, we discuss the specific results only for DEP; the results for NEUR and SWB are similar but more limited. For the tissues tested by DEPICT, **Figure 5a** plots the *P* values based on both the GWAS and MTAG results. As expected, nearly all of the enrichment of signal is found in the nervous system. To facilitate interpretation of the enriched gene sets, we used a standard procedure[23] to group the 351 gene sets into 'clusters' defined by degree of gene overlap. Many of the resulting 41 clusters, shown in **Figure 5b**, implicate communication between neurons ('synapse,' 'synapse assembly,' 'regulation of synaptic transmission,' 'regulation of postsynaptic membrane potential'). This evidence is consistent with that from the DEPICT-prioritized genes, many of which encode proteins that are involved in synaptic communication. For example, *PCLO*, *BSN*, *SNAP25*, and *CACNA1E* all encode important parts of the machinery that releases neurotransmitter from the signaling neuron[24].

The results, however, also contain some intriguing findings. For example, while hypotheses regarding major depression and related traits have tended to focus on monoamine neurotransmitters, our results as a whole point much more strongly to glutamatergic neurotransmission. Moreover, the particular glutamate-receptor genes prioritized by DEPICT (*GRIK2*, *GRIK3*, *GRM1*, *GRM5*, and *GRM8*) suggest the importance of processes involving communication between neurons on an intermediate timescale[25,26], such as learning and memory. Such processes are also implicated by many of the enriched gene sets, which relate to altered reactions to stress and novelty in mice (e.g., 'decreased exploration in a new environment,' 'increased anxiety-related response,' 'behavioral fear response').

DISCUSSION

We have introduced MTAG, a method for conducting meta-analysis of GWAS summary statistics for different traits which is robust to sample overlap. Both our simulation and real-data results confirm that MTAG can increase the statistical power to identify genetic associations with *each* trait. MTAG can be especially useful when the GWAS for the trait of interest is underpowered but shows substantial genetic correlation with other traits. In our real-data application to the traits DEP, NEUR, and SWB, we found that relative to the separate GWASs for the traits, MTAG led to substantial improvements in number of loci identified, predictive power of polygenic scores, and informativeness of a bioinformatics analysis. **Table 1** summarizes the gains from MTAG across these analyses.

Although we have emphasized the use of MTAG to combine GWAS results across different traits, MTAG also has two features that can also make it a valuable tool for conventional meta-analyses of a single trait (see **Online Methods**). First, the method is robust to sample overlap and cryptic relatedness across cohorts. Such overlap can occur if closely related individuals are participants in different cohorts from the same country. The latter scenario may become increasingly common as data from large, national biobanks are incorporated into genetic-association studies together with data from other cohorts. Second, even in meta-analyses that are purportedly of a single trait, different cohorts often have phenotypic data from different measures, which may have different levels of heritability (due to having different amounts of measurement error). In such cases, using MTAG weights instead of sample-size or inverse-variance weights will improve statistical power.

URLs:

Social Science Genetic Association Consortium (SSGAC) website: http://www.thessgac.org/#!data/kuzq8.

ACKNOWLEDGMENTS: We thank Jonathan Beauchamp, Philipp Koellinger, Örjan Sandewall, Carl Shulman, and Ronald de Vlaming for helpful comments. This research was carried out under the auspices of the Social Science Genetic Association Consortium (SSGAC). The SSGAC seeks to facilitate studies that investigate the influence of genes on human behavior, well-being, and social-scientific outcomes using large genome-wide association study meta-analyses. The study was supported by the Ragnar Söderberg Foundation (E9/11 E42/15), the Swedish Research Council (421-2013-1061), The Jan Wallander and Tom Hedelius Foundation, an ERC Consolidator Grant (647648 EdGe), the Pershing Square Fund of the Foundations of Human Behavior, and the NIA/NIH through grants P01-AG005842, P01-AG005842-20S2, P30-AG012810, and T32-AG000186-23 to NBER and R01-AG042568-02 to the University of Southern California. This research has also been conducted using the UK Biobank Resource under Application Number 11425. We thank the research participants and employees of 23andMe for making this work possible. A full list of acknowledgments is provided in the **Supplementary Note**.

CONTRIBUTOR LIST FOR THE 23andMe RESEARCH TEAM: Michelle Agee, Babak Alipanahi, Adam Auton, Robert K. Bell, Katarzyna Bryc, Sarah L. Elson,

Pierre Fontanillas, Nicholas A. Furlotte, David A. Hinds, Bethann S. Hromatka, Karen E. Huber, Aaron Kleinman, Nadia K. Litterman, Matthew H. McIntyre, Joanna L. Mountain, Carrie A.M. Northover, J. Fah Sathirapongsasuti, Olga V. Sazonova, Janie F. Shelton, Suyash Shringarpure, Chao Tian, Joyce Y. Tung, Vladimir Vacic, Catherine H. Wilson, and Steven J. Pitts.

AUTHOR CONTRIBUTIONS: B.N., D.B., D.C. and P.T. oversaw the study. The theory underlying MTAG was conceived of and developed by P.T., with contributions from B.N., D.B., D.C., D.L., O.M. and R.W. O.M., P.T., and R.W. performed the simulations and developed the MTAG software. P.T. and P.V. designed the analyses comparing the observed MTAG gains to theoretical expectations. A.O., D.C., and O.M. played major roles in data analyses. J.J.L. designed and executed the bioinformatics analyses. D.B., D.C., and P.T. coordinated the writing of the manuscript. All authors provided input and revisions for the final manuscript.

COMPETING FINANCIAL INTERESTS: The authors declare no competing financial interests.

CORRESPONDING AUTHORS: P.T. (paturley@broadinstitute.org), B.N. (bneale@broadinstitute.org), D.C. (dac12@nyu.edu), or D.J.B. (daniel.benjamin@gmail.com).

REFERENCES

1. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47,** 1236–1241 (2015).
2. Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6,** 5890 (2015).
3. Maier, R. *et al.* Joint analysis of psychiatric disorders increases accuracy of risk prediction for schizophrenia, bipolar disorder, and major depressive disorder. *Am. J. Hum. Genet.* **96,** 283–94 (2015).
4. Majumdar, A., Haldar, T., Bhattacharya, S. & Witte, J. An efficient Bayesian meta-analysis approach for studying cross-phenotype genetic associations. *bioRxiv* (2017). at <http://biorxiv.org/content/early/2017/01/18/101543.abstract>
5. Zhu, X. *et al.* Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. *Am. J. Hum. Genet.* **96,** 21–36 (2015).
6. Bhattacharjee, S. *et al.* A subset-based approach improves power and interpretation for the combined analysis of genetic association studies of heterogeneous traits. *Am. J. Hum. Genet.* **90,** 821–835 (2012).
7. Shi, H., Mancuso, N., Spendlove, S. & Pasaniuc, B. Local genetic correlation gives insights into the shared genetic architecture of complex traits. *bioRxiv* (2016). at <http://biorxiv.org/content/early/2016/12/08/092668.abstract>
8. Pickrell, J. K. *et al.* Detection and interpretation of shared genetic influences on 42 human traits. *Nat. Genet.* **48,** 709–717 (2016).
9. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genet.* **10,** (2014).

10. Stephens, M. A Unified Framework for Association Analysis with Multiple Related Phenotypes. *PLoS One* **8,** (2013).

11. Ferreira, M. A. R. & Purcell, S. M. A multivariate test of association. *Bioinformatics* **25,** 132–133 (2009).

12. Shim, H. *et al.* A multivariate genome-wide association analysis of 10 LDL subfractions, and their response to statin treatment, in 1868 Caucasians. *PLoS One* **10,** (2015).

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

14. Baselmans, B. M. L. *et al.* Multivariate Genome-Wide and Integrated Transcriptome and Epigenome-Wide Analyses of the Well-being Spectrum. *bioRxiv* (2017). at <http://biorxiv.org/content/early/2017/03/11/115915.abstract>

15. Kass, R. & Raftery, A. Bayes Factors. *J. Am. Stat. Assoc.* **90,** 773–795 (1995).

16. Evans, D. M. The Power of Multivariate Quantitative-Trait Loci Linkage Analysis Is Influenced by the Correlation between Variables. *Am. J. Hum. Genet.* **70,** 1599–1602 (2002).

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

18. Ripke, S. *et al.* A mega-analysis of genome-wide association studies for major depressive disorder. *Mol. Psychiatry* **18,** 497–511 (2013).

19. 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).

20. Okbay, A. *et al.* Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* (2016). doi:10.1038/ng.3552

21. Hyde, C. L. *et al.* Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat. Genet.* **48,** 1031–6 (2016).

22. Vilhjálmsson, B. J. *et al.* Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am. J. Hum. Genet.* **97,** 576–592 (2015).

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

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

25. Pin, J.-P. & Bettler, B. Organization and functions of mGlu and GABAB receptor complexes. *Nature* **540,** 60–68 (2016).

26. Traynelis, S. F. *et al.* Glutamate receptor ion channels: structure, regulation, and function. *Pharmacol. Rev.* **62,** 405–96 (2010).

Authors:

Patrick Turley[1,2], Raymond K. Walters[1,2], Omeed Maghzian[3], Aysu Okbay[4], James J. Lee[5], Mark Alan Fontana[6], Tuan Anh Nguyen-Viet[7], Nicholas A. Furlotte[8], *23andMe Research Team*[8], *Social Science Genetic Association Consortium,* Patrik Magnusson[9], Sven Oskarsson[10], Magnus Johannesson[11], Peter M. Visscher[12,13], David Laibson[3,14], David Cesarini*[,14,15,16], Benjamin Neale*[1,2], and Daniel J. Benjamin*[,7,14,17]

[1] Broad Institute, 415 Main Street, Cambridge, Massachusetts 02142, United States

[2] Analytic and Translational Genetics Unit, Massachusetts General Hospital, 185 Cambridge Street, Cambridge, Massachusetts 02114, United States

[3] Department of Economics, Harvard University, 1805 Cambridge Street, Cambridge, Massachusetts 02138, United States

[4] Department of Complex Trait Genetics, Vrije Universiteit Amsterdam, De Boelelaan 1085, Amsterdam, 1081 HV, Netherlands

[5] Department of Psychology, University of Minnesota, 75 E River Road, Minneapolis, Minnesota 55455, United States

[6] Hospital for Special Surgery, 535 E 70th Street, New York, New York 10021, United States

[7] Center for Economic and Social Research, University of Southern California, 635 Downey Way, Los Angeles, California 90089, United States

[8] 23andMe, Inc., Mountain View, California 94043, United States

[9] Institutionen för Medicinsk Epidemiologi och Biostatistik, Karolinska Institutet, 171 77 Stockholm, Sweden

[10] Department of Government, Uppsala Universitet, 751 20 Uppsala, Sweden

[11] Department of Economics, Stockholm School of Economics, 113 83 Stockholm, Sweden

[12] Institute for Molecular Bioscience, The University of Queensland, Building 80 Services Road, QLD 4072, Brisbane, Queensland, Australia

[13] Queensland Brain Institute, The University of Queensland, Building 79 Services Road, QLD 4072, Brisbane, Queensland, Australia

[14] National Bureau of Economic Research, 1050 Massachusetts Avenue, Cambridge, Massachusetts 02138, United States

[15] Department of Economics and Center for Experimental Social Science, New York University, 19 W 4th Street, New York, New York 10012, United States

[16] Institutet för Näringslivsforskning, 114 53 Stockholm, Sweden

[17] Department of Economics, University of Southern California, 3620 South Vermont Avenue, Kaprielian Hall, 300 Los Angeles, CA 90089, United States

[*] These authors contributed equally

Fig. 1. Selected simulation results. The *y*-axis is the root mean squared error (RMSE) of the GWAS and MTAG estimates of Trait 1 in a two-trait simulation. Each pair of bars corresponds to a different scenario characterized by the distribution and parameter values listed under the bars. "Distr" reports the distribution of the effect sizes generated in the simulation: multivariate normal or multivariate spike-and-slab. χ_1^2 and χ_2^2 are the expected χ^2-statistic of the GWAS estimates for Trait 1 and Trait 2, respectively. r_β is the correlation of the effect sizes across traits. r_ε is the correlation of the estimation error across traits. For each scenario, we conduct 1,000 replications. In each replication, we generate GWAS estimates for 100,000 independent SNPs for each trait. The bars and the 95% confidence intervals show the mean across replications and the 2.5th and 97.5th percentiles, respectively. Complete results can be found in **Supplementary Note**.

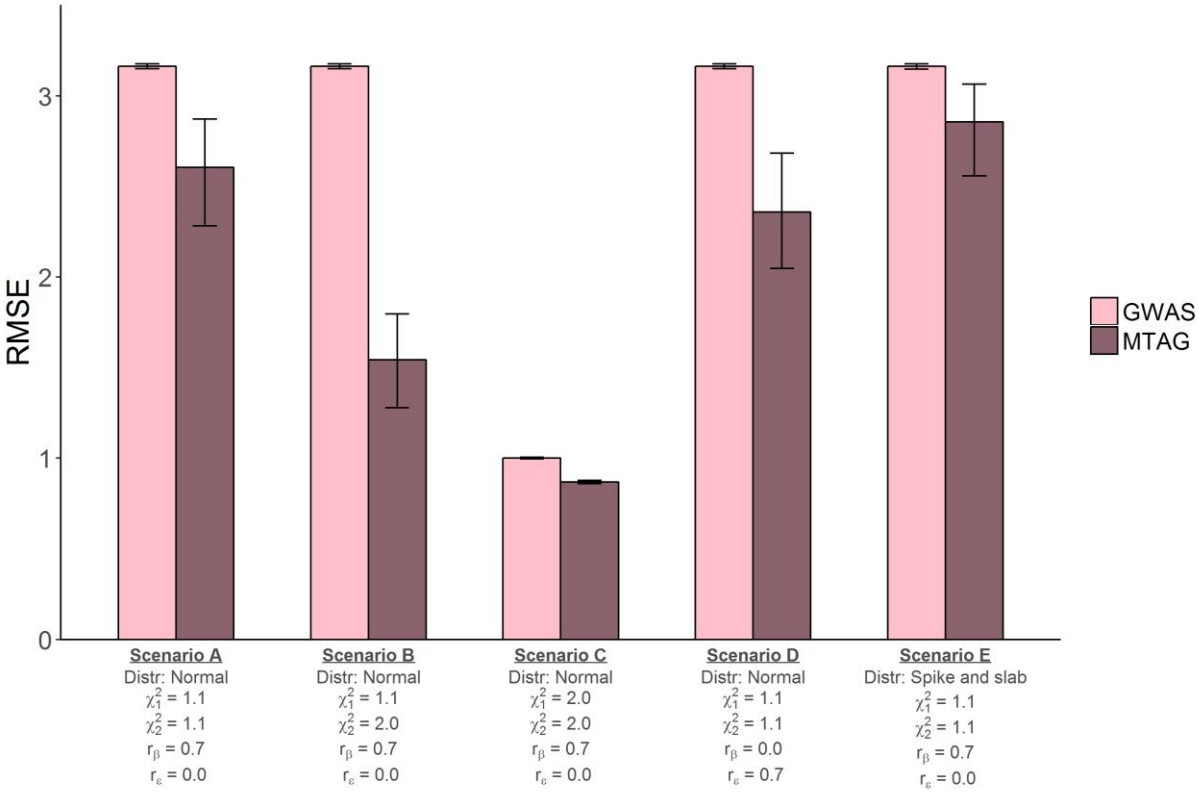


Fig. 2. Cohorts included in GWAS meta-analyses for DEP, NEUR, and SWB. In UKB, the sample overlap in the summary statistics across the traits is known, whereas in 23andMe, the sample overlap in the summary statistics is unknown. MTAG accounts for both sources of overlap. SSGAC results[20], GPC results[19], GERA results[18], and 23andMe results for DEP[21] all come from previously published work. The data from 23andMe for SWB are newly analyzed data for this paper. Data from the UKB for all three traits has been previously published[20], though we re-analyze it in this paper with slightly different protocols (see **Supplementary Note**). N_{eff} is used instead of N when the cohort has case-control data (**Supplementary Note 3.10**). The sample size listed for each cohort corresponds to the maximum sample size across all SNPs available for that cohort. The total sample size for each trait corresponds to the maximum sample size among the SNPs available after applying MTAG filters.

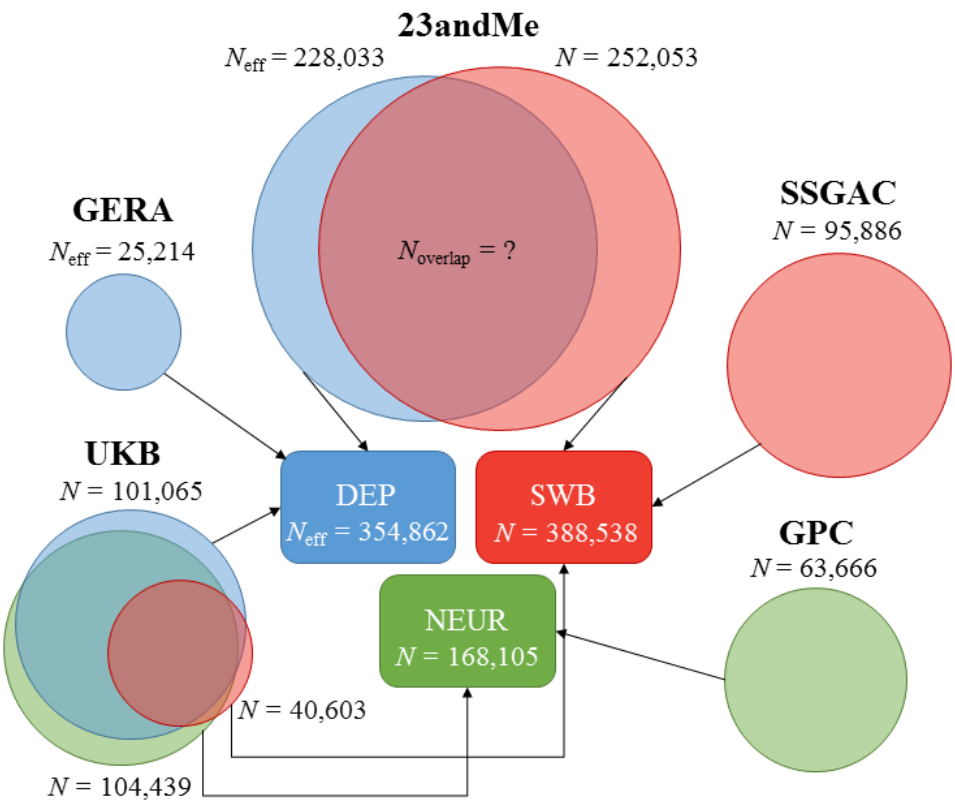
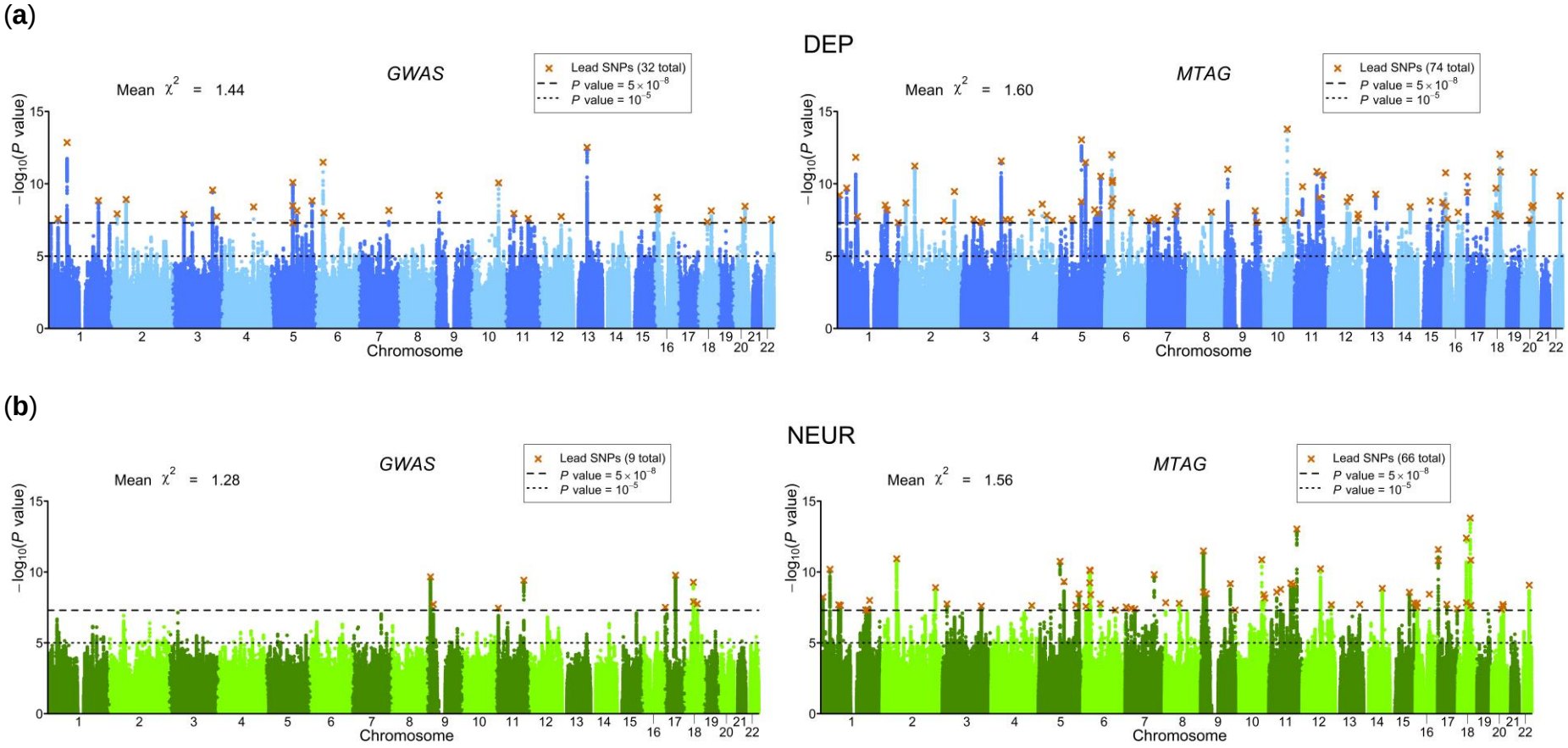


Fig. 3. Manhattan plots of GWAS and MTAG results. (**a**) DEP, (**b**) NEUR, (**c**) SWB. The left and right plots display the GWAS and MTAG results, respectively. All plots in all panels are based on the same set of SNPs. The *x*-axis is chromosomal position, and the *y*-axis is the significance on 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 ×. The mean χ^2-statistic across all SNPs included in the analysis is displayed in the top left corner of each plot.



(c)

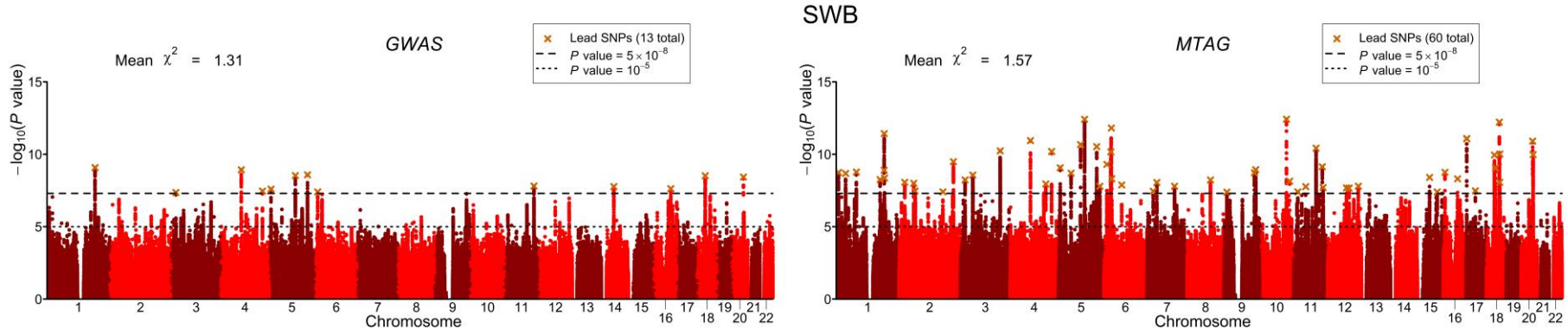
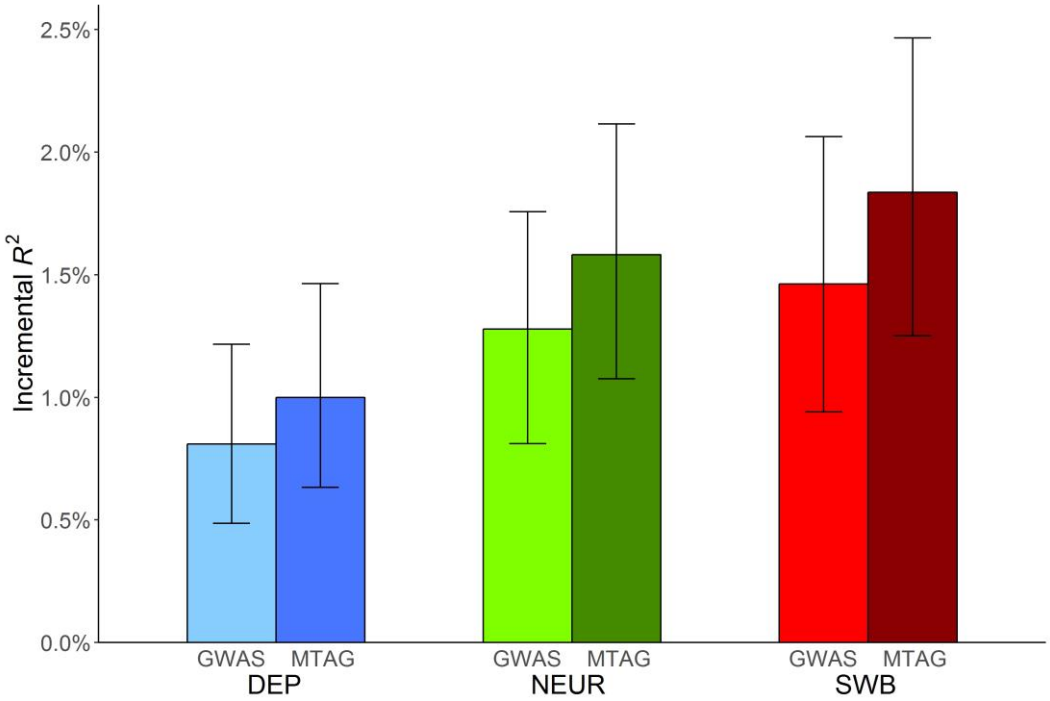

SWB
GWAS
Mean χ^2 = 1.31
Lead SNPs (13 total)
P value = 5×10^{-8}
P value = 10^{-5}
$-\log_{10}$(P value)
Chromosome
MTAG
Mean χ^2 = 1.57
Lead SNPs (60 total)
P value = 5×10^{-8}
P value = 10^{-5}
$-\log_{10}$(P value)
Chromosome

Fig. 4. Predictive power of GWAS- and MTAG-based polygenic scores. (**a**) The incremental R^2 in the Health and Retirement Study (HRS) cohort for scores based on GWAS and MTAG results that exclude the HRS. Incremental R^2 is a measure of 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 covariates are year of birth, year of birth squared, sex, their interactions, and 10 principal components of the variance-covariance matrix of the genotype data. (**b**) The difference in incremental R^2 between the GWAS- and the MTAG-based PGS for each trait. Bootstrapped 95% confidence intervals are shown in both plots.

(**a**)



(**b**)

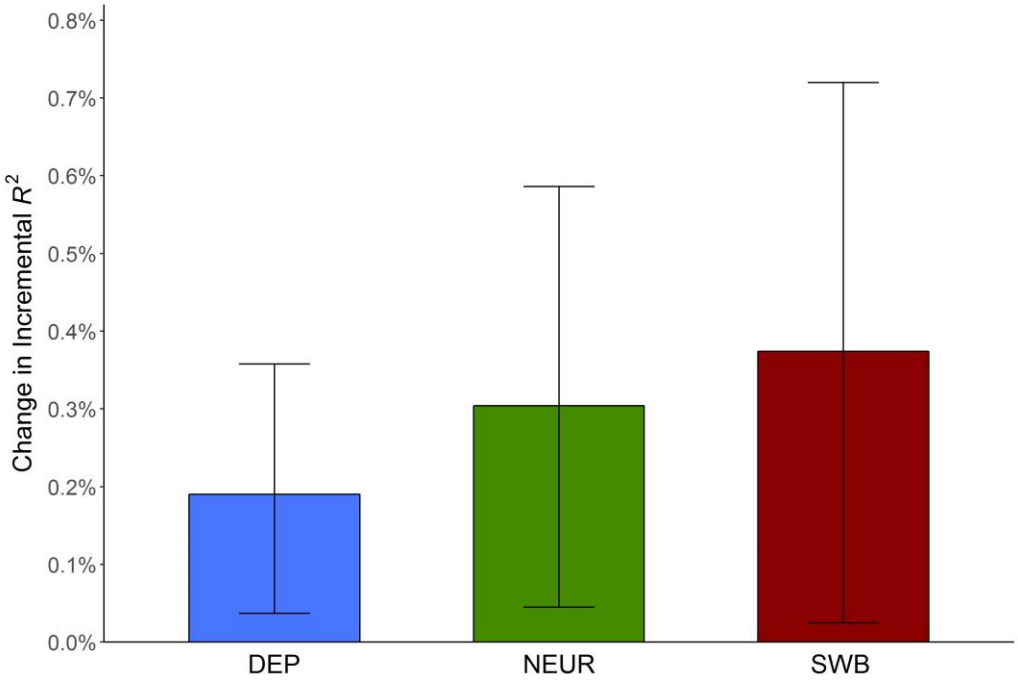
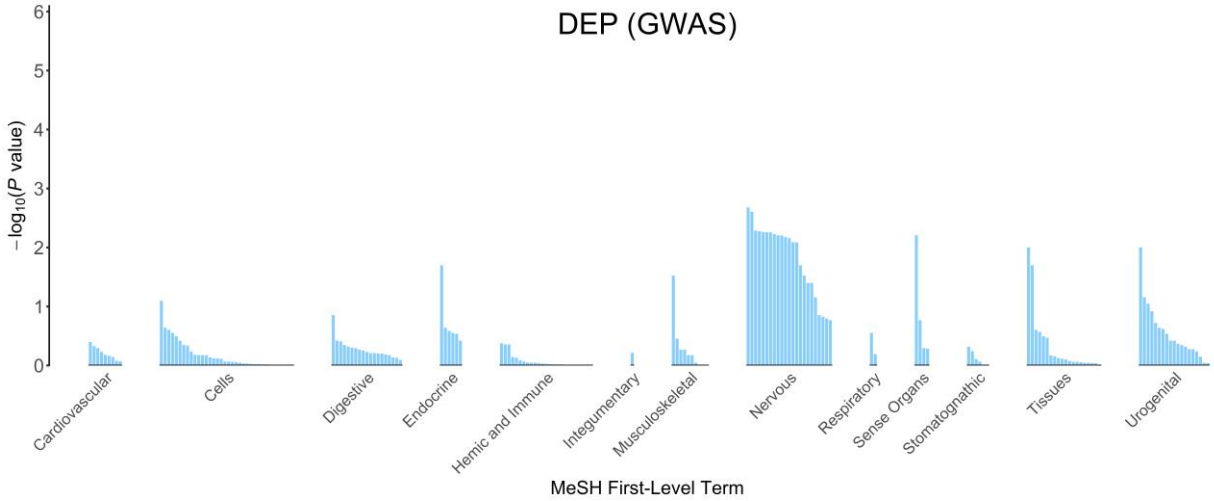
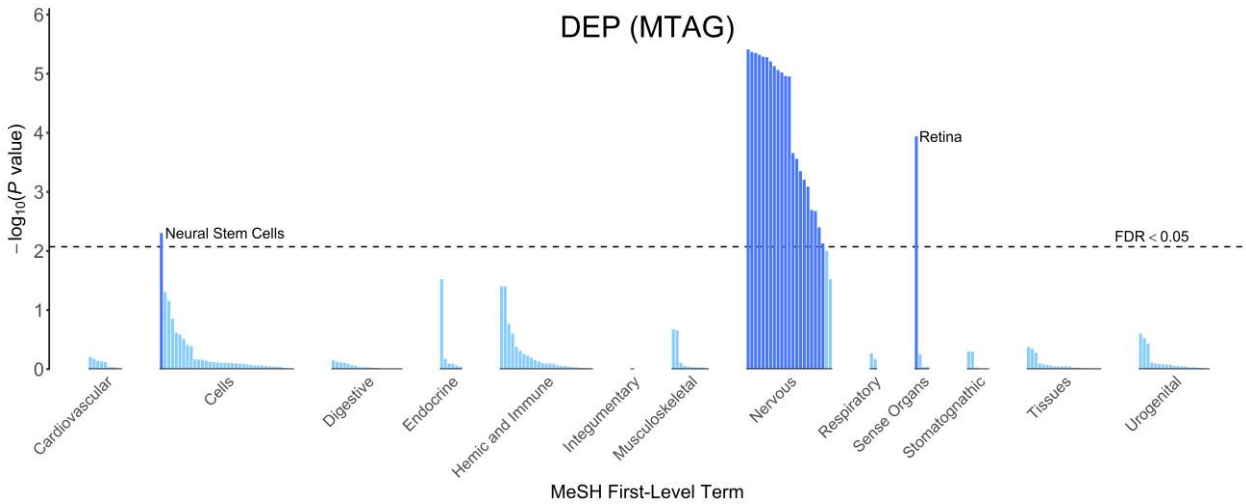


Fig. 5. Biological annotation for DEP using the bioinformatics tool DEPICT. (**a**) Results of the tissue enrichment analysis based on the GWAS and MTAG results. The *x*-axis lists the tissues tested for enrichment, grouped by the location of the tissue. The *y*-axis is statistical significance on a $-\log_{10}$ scale. The horizontal dashed line corresponds to a false discovery rate of 0.05, which is the threshold used to identify prioritized tissues. (**b**) Gene-set clusters as defined by the Affinity Propagation algorithm[23] over the gene sets from the MTAG results. The algorithm names clusters an exemplary member of the gene set. The color of the point signifies the *P* value of the most significant gene set in the cluster. The line thickness between the gene-set clusters corresponds to the correlation between the named gene sets for each pair of clusters.

(**a**)





(**b**)

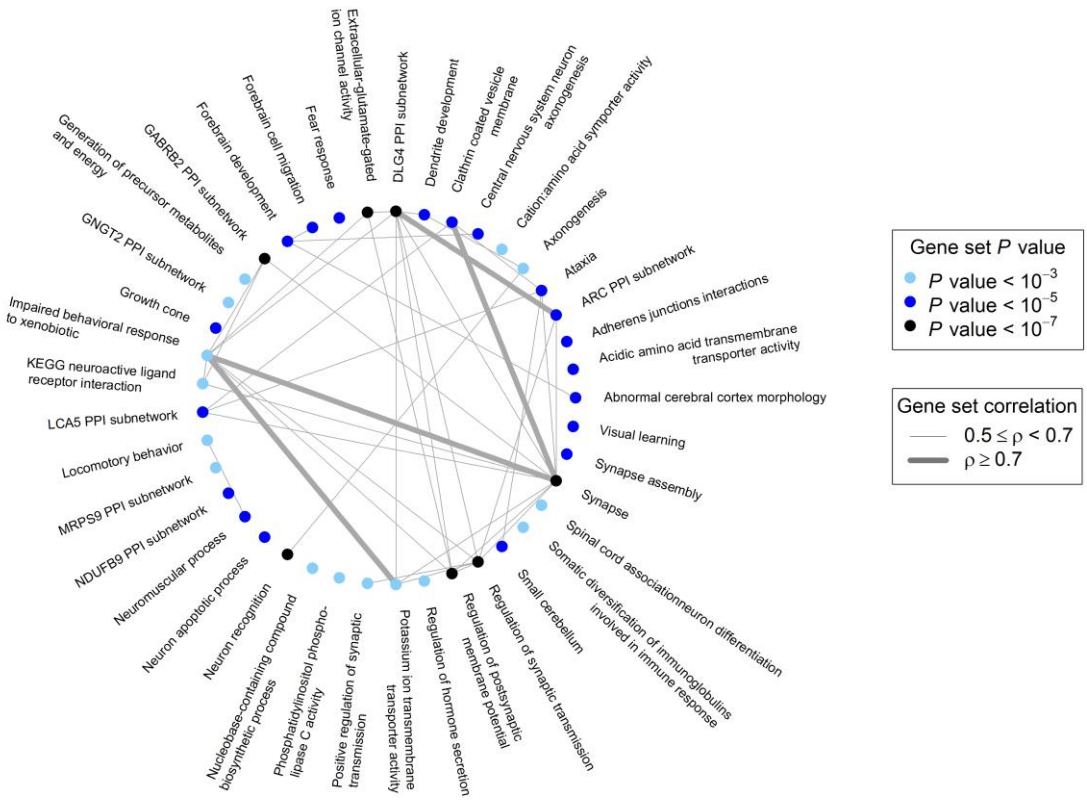

Extracellular-glutamate-gated ion channel activity
DLG4 PPI subnetwork
Dendrite development
Clathrin coated vesicle membrane
Central nervous system neuron axonogenesis
Cation:amino acid symporter activity
Axonogenesis
Ataxia
ARC PPI subnetwork
Adherens junctions interactions
Acidic amino acid transmembrane transporter activity
Abnormal cerebral cortex morphology
Visual learning
Synapse assembly
Synapse
Spinal cord associationneuron differentiation
Somatic diversification of immunoglobulins involved in immune response
Small cerebellum
Regulation of synaptic transmission
Regulation of postsynaptic membrane potential
Regulation of hormone secretion
Potassium ion transmembrane transporter activity
Positive regulation of synaptic transmission
Phosphatidylinositol phospho-lipase C activity
Nucleobase-containing compound biosynthetic process
Neuron recognition
Neuron apoptotic process
Neuromuscular process
NDUFB9 PPI subnetwork
MRPS9 PPI subnetwork
Locomotory behavior
LCA5 PPI subnetwork
KEGG neuroactive ligand receptor interaction
Impaired behavioral response to xenobiotic
Growth cone
GNGT2 PPI subnetwork
Generation of precursor metabolites and energy
GABRB2 PPI subnetwork
Forebrain development
Forebrain cell migration
Fear response
Gene set P value
P value < 10^−3
P value < 10^−5
P value < 10^−7
Gene set correlation
0.5 ≤ ρ < 0.7
ρ ≥ 0.7

Table 1. Summary of comparative analyses of GWAS and MTAG results

	DEP		NEUR		SWB	
	GWAS	MTAG	GWAS	MTAG	GWAS	MTAG
SNP-based comparisons						
Lead SNPs ($P < 5\times10^{-8}$)	32	74	9	66	13	60
Mean χ^2	1.44	1.60	1.28	1.56	1.31	1.57
N_{eff}	354,862	487,603	168,105	329,835	388,538	718,284
Polygenic score incremental R^2	0.81%	1.00%	1.28%	1.58%	1.46%	1.84%
Biological Annotation (DEPICT FDR < 0.05)						
# Prioritized Genes	4	99	0	7	0	8
# Gene Sets	0	351	0	1	0	9
# Tissues and cell types	0	23	0	13	0	14

ONLINE METHODS

This article is accompanied by a **Supplementary Note** with further details.

Theory. There are T traits, which may be binary or quantitative. We standardize each trait and the genotype for each single-nucleotide polymorphism (SNP) j so that they all have mean zero and variance one. The length-T vector of marginal (i.e., not controlling for other SNPs), true effects of SNP j on each of the traits is denoted $\boldsymbol{\beta}_j$. We assume that these are random effects drawn $\boldsymbol{\beta}_j \sim N(\mathbf{0}, \boldsymbol{\Omega})$. The mean is zero because we treat the choice of reference allele as arbitrary. We make the common assumption[17,22,27] that the $\boldsymbol{\beta}_j$'s are identically distributed across j. The assumption implies that each SNP's expected contribution to a trait's heritability is equal, regardless of SNP characteristics such as allele frequency.

The length-T vector of GWAS estimates is denoted $\widehat{\boldsymbol{\beta}}_j$. We assume that $\widehat{\boldsymbol{\beta}}_j | \boldsymbol{\beta}_j \sim N(\boldsymbol{\beta}_j, \boldsymbol{\Sigma}_j)$, where the matrix $\boldsymbol{\Sigma}_j$ is the variance-covariance matrix of estimation error, which includes sampling variation and biases (such as population stratification or technical artifacts). The matrix $\boldsymbol{\Sigma}_j$ may differ across SNPs due to differences in the SNPs' sample sizes per trait and the SNPs' sample overlap between traits.

By the properties of normal distributions, the conditional distribution of the estimated effect vector $\widehat{\boldsymbol{\beta}}_j$ conditional on the true effect of the SNP on *only one* of the traits is

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

where $\boldsymbol{\omega}_t$ is a vector equal to the tth column of $\boldsymbol{\Omega}$ and ω_{tt} is a scalar equal to the tth element of $\boldsymbol{\omega}_t$ (or equivalently, the tth diagonal element of $\boldsymbol{\Omega}$). The MTAG estimator is obtained by maximizing this marginal likelihood function[15] with respect to $\beta_{j,t}$. In the **Supplementary Note**, we show that the solution is:

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

Since $\mathrm{E}(\widehat{\boldsymbol{\beta}}_j | \beta_{j,t}) = \frac{\boldsymbol{\omega}_t}{\omega_{tt}}\beta_{j,t}$, the MTAG estimator with known $\boldsymbol{\Sigma}_j$ and $\boldsymbol{\Omega}$ is unbiased: $\mathrm{E}(\hat{\beta}_{\text{MTAG},j,t} | \beta_{j,t}) = \beta_{j,t}$. Therefore, when calculated with consistent estimates $\widehat{\boldsymbol{\Sigma}}_j$ and $\widehat{\boldsymbol{\Omega}}$, the MTAG estimator is consistent.

The sampling variance of the estimator is

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

For each trait t, the standard error of the estimate is the square root of this quantity. As is standard, we obtain a *P* value using the fact that in large samples, $\hat{\beta}_{\mathrm{MTAG},j,t}$ divided by its standard error follows a standard normal distribution under the null hypothesis.

The above formulas for the MTAG estimator and its standard error use the variance-covariance matrix of true SNP effects across all SNPs, $\boldsymbol{\Omega}$, to calculate the optimal MTAG weights for each SNP. If in fact there are different types of SNPs characterized by different variance-covariance matrices, then the standard error formula is anti-conservative because it does not account for this source of variation across SNPs. The MTAG estimator, however, remains consistent (but could be made more efficient if it took into account the different types of SNPs).

For expositional simplicity, our derivations above and in **Supplementary Note** are parameterized in terms of the parameter vector $\widehat{\boldsymbol{\beta}}_j$. We note, however, that the input to the MTAG software is the standard output from meta-analysis software: *z*-statistics and sample sizes (**Supplementary Note, Section 1**). Because MTAG is applied to *z*-statistics, the GWAS summary statistics do not need to have been estimated using traits and genotypes that were standardized.

Special Cases. There are four special cases of MTAG that may often be relevant in practice and for which the estimation procedure is made faster and more efficient. The MTAG software offers the option to specialize the analysis for these cases.

No sample overlap across traits. In this case, the off-diagonal elements of $\boldsymbol{\Sigma}_j$ can be set equal to zero, so LD score regression needs to be run only T rather than $T(T+1)/2$ times. Note that this version of MTAG does not take into account correlation in estimation error across traits that is due to bias. For this reason, LD score regression should be run on the MTAG results, and the resulting MTAG standard errors should be inflated by the square root of the estimated intercept.

Perfect genetic correlation but different heritabilities. This case arises when the "traits" are different measures of the same trait, some with more measurement error than others, or when the variance in the trait due to non-genetic factors differs. Here the $\boldsymbol{\Omega}$ matrix has only T rather than $T(T+1)/2$ unique parameters to be estimated.

Perfect genetic correlation and equal heritabilities. This special case corresponds to the "traits" being (the same measure of) a single trait; in other words, applying MTAG instead of inverse-variance-weighted meta-analysis to T GWAS results. Doing so

can be useful if there is sample overlap in the GWAS results. In this case, as noted in the main text, MTAG specializes to $\hat{\beta}_{\mathrm{MTAG},j,t} = \frac{\mathbf{1}'\boldsymbol{\Sigma}_j^{-1}}{\mathbf{1}'\boldsymbol{\Sigma}_j^{-1}\mathbf{1}}\widehat{\boldsymbol{\beta}}_j$ for all t, and it is no longer necessary to estimate $\boldsymbol{\Omega}$.

Equal sample sizes across SNPs. In this case, the maximum-likelihood estimation of the $\boldsymbol{\Omega}$ matrix has a closed-form solution, substantially reducing computation time. When applying MTAG to a large number of traits, numerical maximum-likelihood estimation becomes computationally intensive, so it may be preferable to instead restrict the set of SNPs to those with approximately equal sample sizes and use this special case.

Computational Run-time. The only step of MTAG that may be computationally intensive is estimating the $\boldsymbol{\Omega}$ matrix via numerical maximum likelihood (the other steps have closed-form solutions and therefore execute very quickly). In the real-data application described in this paper—with three traits and 6.1M SNPs—this step takes approximately 4.5 hours, and the MTAG estimation overall takes approximately 4.9 hours. For comparison, we apply MTAG with the option selected for equal sample sizes across SNPs to the same GWAS summary statistics except that, for each trait, we treat the sample size for each SNP as if it were equal to the mean sample size across SNPs for that trait. In that case, the run-times fall to 2 seconds and 31 minutes, respectively.

The reported run times are the medians of five identical runs using one core of a 2.20 GHz Intel(R) Xeon(R) CPU E5-2650 v4 processor. Of course, run time may vary as a function of the computing environment.

Simulations. To speed computations, instead of simulating data and then estimating effect sizes, we directly generated effect-size estimates by adding multivariate normally distributed noise to the simulated effect sizes. The variance of the noise for each trait was pinned down by the scenario's expected χ^2-statistic, and the covariance of the noise between the traits was pinned down by the scenario's value for the correlation of the effect sizes.

In our simulation, we cannot estimate $\boldsymbol{\Sigma}_j$ using LD Score regression because there is no LD between SNPs. Nonetheless, we would like to verify that MTAG is robust to the sampling variation in the estimate $\widehat{\boldsymbol{\Sigma}}_j$. To accomplish this, in each replication we directly generated an estimate $\widehat{\boldsymbol{\Sigma}}_j$ by adding noise to the true value of $\boldsymbol{\Sigma}_j$. The variance of the noise was set to be roughly equal to the sampling variance observed in the empirical application considered in this paper.

GWAS Meta-analyses of DEP, NEUR, and SWB. Details on the cohorts, phenotype measures, genotyping, quality-control filters, and association models are provided in **Supplementary Note** and **Supplementary Table 1.1**. As shown in **Figure 2**, there is substantial overlap in samples across the three GWAS meta-analyses.

All analyses were based on autosomal SNPs from cohorts with genotypes imputed against the 1000 Genomes reference panel. The input files in each meta-analysis

were subject to a uniform set of quality-control and diagnostic procedures. These are described in the previous SSGAC study[20] and **Supplementary Note**.

As expected under polygenicity[28], we observe inflation of the median test statistic in each GWAS ($\lambda_{GC,DEP}$ = 1.36, $\lambda_{GC,NEUR}$ = 1.24, $\lambda_{GC,SWB}$ = 1.28; see **Supplementary Fig 3.5**). The intercept estimates from LD score regression are all below 1.02, however, suggesting that nearly all of the observed inflation is due to polygenic signal[17]. When we report GWAS results, as in the SSGAC study[20] we account for the potential bias due to this small amount of stratification by inflating the standard errors of our GWAS estimates by the square root of the LD score regression intercept.

Manhattan plots from each of the GWAS meta-analyses are shown in **Supplementary Figs 3a**, **3b** and **3c**. Our NEUR meta-analysis was based on the same cohort-level data as the SSGAC study[20] and unsurprisingly yielded substantively identical results: 12 lead SNPs. Consistent with what studies have reported for other complex traits, the increased discovery samples for DEP and SWB relative to the SSGAC study increased the number of lead SNPs: from 2 to 32 for DEP (N_{eff} = 149,707 to 354,862) and from 3 to 13 for SWB (N = 298,420 to 388,538). Applying bivariate LD score regression[1] to the GWAS results, we estimate the genetic correlations to be 0.72 (s.e. = 0.025) between DEP and NEUR, -0.67 (s.e. = 0.027) between NEUR and SWB, and -0.687 (s.e. = 0.024) between DEP and SWB.

Clumping Algorithm. We applied the same clumping algorithm to the GWAS and MTAG results to identify each set of "lead SNPs." Our clumping algorithm is the same as in the previous SSGAC study[20]. First, the SNP with the smallest P value was identified in the meta-analysis results. This SNP was designated the index SNP of clump 1. Second, we identified all SNPs on the same chromosome whose LD with the index SNP exceeds R^2 = 0.1 and assigned them to clump 1. To generate the second clump, we removed the SNPs in clump 1 and then iterated the process to identify further index SNPs and their corresponding clumps until no SNPs remain.

MTAG SNP Filters. Since the derivation of MTAG relies on some assumptions regarding the distributions of the effect sizes and estimation error, its performance may be sensitive to violations of those assumptions. To reduce the risk of extreme violations, when we apply MTAG, we impose three additional SNP filters beyond the standard filters used in a GWAS.

First, we restrict the set of SNPs to those with a minor allele frequency greater than 1%. This filter is motivated by the assumption that each SNP contributes equally to heritability in expectation. Rare variants may follow a different effect-size distribution both in terms of the variance and covariance of their effect sizes, which could lead to biased MTAG estimates.

Second, for each trait, we restrict variation in sample size by calculating the 90th percentile of the SNP sample-size distribution and removing SNPs with a sample size smaller than 75% of this value. This filter is similar to, though slightly more strict than, the sample-size filter recommended for LD Score regression[17]. If a SNP's

effect is estimated in a relatively small sample, then the sample overlap across traits will likely be different for that SNP than for other SNPs in the sample. In that case, the covariance of the estimation error across traits as estimated by LD Score regression may not be a good approximation of the covariance of the estimation error for that SNP.

Third, we drop SNPs in genomic regions containing SNPs that are outliers with respect to their effect-size estimates. Because the effect size of these SNPs appear to have a different variance-covariance matrix than the rest of the genome, including these regions would likely lead to the biases and inefficiencies that can occur when $\mathbf{\Omega}$ is not constant across SNPs. In our real-data application, in GWAS of neuroticism, the effect sizes of SNPs in a region of chromosome 8 that tag an inversion polymorphism have been found to be strongly inflated relative to the effects estimated for SNPs in all other regions of the genome[20,29]. Therefore, we omit SNPs in chromosome 8 between base-pair positions 7,962,590 and 11,962,591.

GWAS-Equivalent Sample Size for MTAG. The increase in the mean χ^2-statistic for each trait from the GWAS results to the MTAG results can be used to calculate a "GWAS-equivalent sample size" for MTAG. Under the assumptions of LD score regression[17], the expected χ^2-statistic for some SNP with LD score ℓ_j is

$$\mathrm{E}\left(\chi_j^2 \mid \ell_j\right) = \frac{N_j h^2 \ell_j}{M} + N_j a + 1,$$

where N_j is the sample size for SNP j; h^2 is the SNP-heritability of the trait; M is the number of SNPs for which we define the SNP-heritability; and a is the variance due to biases (e.g., due to population stratification). Note that $\mathrm{E}\left(\chi_j^2 \mid \ell_j\right)$-1 scales linearly with N_j. Thus, we can use the mean χ^2-statistic from the GWAS and the MTAG results to calculate how much larger the GWAS sample size would have to be to give a mean χ^2-statistic equal to that attained by MTAG:

$$N_{\text{GWAS-equiv},j} = N_{\text{GWAS},j} \frac{\overline{\chi^2_{\text{MTAG}}} - 1}{\overline{\chi^2_{\text{GWAS}}} - 1},$$

where $\overline{\chi^2_{\text{GWAS}}}$ is the mean χ^2-statistic in the GWAS results, and $\overline{\chi^2_{\text{MTAG}}}$ is the mean χ^2-statistic in the MTAG results. Put another way, conducting MTAG gives the same power (as measured by mean χ^2-statistic) as conducting GWAS in sample size that is larger by a factor of $\frac{\overline{\chi^2_{\text{MTAG}}}-1}{\overline{\chi^2_{\text{GWAS}}}-1}$.

For DEP, going from GWAS to MTAG, the mean χ^2-statistic increases from 1.44 to 1.60, implying an increase in the GWAS-equivalent sample size by a factor of

$$\frac{1.600 - 1}{1.437 - 1} = 1.37.$$

Thus, the MTAG analysis has statistical power equivalent to a GWAS of DEP conducted in $354{,}862 \times 137\% = 487{,}603$ individuals. For NEUR, the mean χ^2-statistic rises from 1.284 to 1.557, implying a GWAS-equivalent sample size for MTAG that is 96% larger than the GWAS sample size: the effective sample size is $168{,}105 \times 196\% = 329{,}835$ individuals. For SWB, the mean χ^2-statistics rises from 1.308 to 1.570, implying a GWAS-equivalent sample size 85% larger than the GWAS: $388{,}538 \times 185\% = 718{,}284$ individuals.

Expected Increase in Mean χ^2-statistic from MTAG. The expected χ^2-statistic of the GWAS summary statistics for trait t is

$$\mathrm{E}\left(\chi^2_{\mathrm{GWAS},t}\right) = \frac{\omega_{tt} + \sigma_t^2}{\sigma_t^2}$$

where σ_t^2 is the tth diagonal element of $\boldsymbol{\Sigma}_j$. The expected χ^2-statistic of the MTAG summary statistics for trait t is

$$\mathrm{E}\left(\chi^2_{\mathrm{MTAG},t}\right) = \frac{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1} \mathrm{E}\left(\widehat{\boldsymbol{\beta}}_j\widehat{\boldsymbol{\beta}}_j'\right)\boldsymbol{\Sigma}_j\left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}}{\frac{\boldsymbol{\omega}_t'}{\omega_{tt}}\left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1}\boldsymbol{\Sigma}_j\left(\boldsymbol{\Omega} - \frac{\boldsymbol{\omega}_t\boldsymbol{\omega}_t'}{\omega_{tt}} + \boldsymbol{\Sigma}_j\right)^{-1}\frac{\boldsymbol{\omega}_t}{\omega_{tt}}}, =$$

where we can substitute $\mathrm{E}\left(\widehat{\boldsymbol{\beta}}_j\widehat{\boldsymbol{\beta}}_j'\right) = \left(\boldsymbol{\Omega} + \boldsymbol{\Sigma}_j\right)$.

Plugging our estimates $\widehat{\boldsymbol{\Omega}}$ and $\widehat{\boldsymbol{\Sigma}}_j$ into the above equations, we can use the expected χ^2-statistics to calculate the theoretically expected gain in GWAS-equivalent sample size from applying MTAG (as derived previously):

$$\frac{\mathrm{E}\left(\chi^2_{\mathrm{MTAG},t}\right) - 1}{\mathrm{E}\left(\chi^2_{\mathrm{GWAS},t}\right) - 1}.$$

Note that $\widehat{\boldsymbol{\Sigma}}_j$ is a function of the sample sizes used to generate the GWAS summary statistics, and we use the $\widehat{\boldsymbol{\Sigma}}_j$ corresponding to the maximum sample size in the SNPs used in MTAG. Based on this formula, the theoretically expected increases in the GWAS-equivalent sample sizes are 37%, 96%, and 85% for DEP, NEUR, and SWB, respectively. These are close to the observed increases of 39%, 89%, and 92%.

Winner's Curse Correction for Replication Analysis. MTAG estimates are corrected for winner's curse following the same procedure found in Okbay et al.[20] Briefly, for each trait, we use maximum likelihood to fit the MTAG results to a spike-and-slab distribution such that

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

For DEP, NEUR, and SWB, we estimate $\hat{\pi}$ to be 0.565, 0.563, and 0.549 and $\hat{\tau}^2$ to be 3.04×10^{-6}, 4.72×10^{-6}, and 2.02×10^{-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 $\hat{\beta}_{\mathrm{MTAG},j}$ for each SNP as

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

where $\pi_{\mathrm{post},j}$ is the posterior probability that $\beta_j = 0$ and $\hat{s}_j^2$ is the squared standard error of the MTAG estimate. For further details about the estimation procedure, see Section 8 of the Supplementary Note of Okbay et al.[20]

Polygenic Prediction. We used the Health and Retirement Study[30] (HRS) as our holdout cohort for polygenic prediction. We applied the same SNP filters as in the main MTAG analyses. We also imposed several additional, HRS-specific filters in order to ensure that the polygenic scores are based only on SNPs that were reliably genotyped in the HRS. Specifically, SNPs were omitted unless they (i) were genotyped in the HRS cohort with a call rate greater than 98%, (ii) have a Hardy-Weinberg Equilibrium (HWE) test *P* value greater than 10^{-4} in the HRS, (iii) have a minor allele frequency (MAF) greater than 1% in the HRS, and (iv) are not strand ambiguous.

Bootstrapped confidence intervals were calculated by drawing, with replacement, a sample of equal size to the prediction sample, and calculating the incremental R^2 for the GWAS-based polygenic score, the MTAG-based polygenic score, and the difference between them. As the bounds of the 95%-confidence intervals, we use the 2.5th- and 97.5th-percentile values of the incremental R^2's across 500 bootstrap draws.

Expected Increase in Polygenic-score Predictive Power from MTAG. The phenotypic value of a trait in individual i, denoted y_i, can be decomposed into the sum of the additive genetic variance component and a residual:

$$y_i = g_i + \varepsilon_i.$$

We denote the GWAS- and MTAG-based polygenic scores for the trait by $\hat{g}_{\mathrm{GWAS},i}$ and $\hat{g}_{\mathrm{MTAG},i}$, respectively. Note that GWAS and MTAG produce unbiased estimates of the SNP effect sizes, and LDpred produces a consistent estimate of the additive genetic variance component. Therefore, each polygenic score $k \in \{\mathrm{GWAS}, \mathrm{MTAG}\}$ is approximately equal to g_i plus estimation error:

$$\hat{g}_{k,i} = g_i + e_{k,i}.$$

By the central limit theorem, the estimation error is approximately normally distributed,

$$e_{k,i} \sim N(0, V_k),$$

with variance V_k that is proportional to the GWAS-equivalent sample size of the estimator (which we can estimate for the MTAG polygenic score using the approach described in a previous section of the Online Methods).

The expected predictive power of a polygenic score is

$$\mathrm{E}(R_k^2) = \frac{(h^2)^2}{h^2 + V_k},$$

where h^2 is the SNP heritability of the trait[31]. Using the GWAS results, we obtain an estimate of h^2 using LD score regression[17] and an estimate of $\mathrm{E}(R_k^2)$ from the predictive power of the GWAS-based polygenic score. Plugging these estimates into the above formula, we solve for an estimate of V_{GWAS}. We then multiply this value by $\frac{\overline{\chi^2_{\mathrm{GWAS}}} - 1}{\overline{\chi^2_{\mathrm{MTAG}}} - 1}$ (which we showed previously is equal to the ratio of the GWAS sample size to the MTAG's GWAS-equivalent sample size) to obtain an estimate of V_{MTAG}. Substituting this back into the above formula along with our estimate of h^2 gives us the expected predictive power of the MTAG-based PGS.

For DEP, we anticipated an increase in predictive power of 0.25 percentage points, which is within the estimated confidence interval [0.04, 0.36]. For NEUR, we expected an increase of 0.95 percentage points, which is greater than the confidence interval [0.05, 0.59]. For SWB, we expected an increase of 0.62 percentage points, which is again within the confidence interval [0.03, 0.72]. While not all of these predictions fall within the estimated confidence intervals, some of this discrepancy may be due to differences in heritability or imperfect genetic correlation between the discovery and prediction sample.[32] As a whole, we find that the observed predictive power of the MTAG-based polygenic score is largely consistent with this theoretical expectation.

Biological Annotation. For the identification of genes, we supplemented the DEPICT inventory with protein-coding genes that have a status of 'known' in GENCODE (downloaded February 26, 2015). Specifically, we assigned such a gene to a lead SNP in **Supplementary Tables 6.1**, **6.4**, and **6.7** if it either encompasses the SNP in a DEPICT-defined locus or has the transcription start site closest to such a SNP.

METHODS-ONLY REFERENCES

27. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88,** 76–82 (2011).

28. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* **19,** 807–812 (2011).

29. Smith, D. J. *et al.* Genome-wide analysis of over 106 000 individuals identifies 9 neuroticism-associated loci. *Mol. Psychiatry* **21,** 749–57 (2016).

30. Sonnega, A. *et al.* Cohort Profile: the Health and Retirement Study (HRS). *Int.*

J. Epidemiol. **43,** 576–585 (2014).

31. Daetwyler, H. D., Villanueva, B. & Woolliams, J. a. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One* **3,** e3395 (2008).

32. de Vlaming, R. *et al.* Meta-GWAS Accuracy and Power (MetaGAP) Calculator Shows that Hiding Heritability Is Partially Due to Imperfect Genetic Correlations across Studies. *PLOS Genet.* (2017). doi:10.1371/journal.pgen.1006495