Trait selection strategy in multi-trait GWAS: Boosting SNPs discoverability

Yuka Suzuki, Hervé Ménager, Bryan Brancotte, Raphaël Vernet, Cyril Nerin, Christophe Boetto, Antoine Auvergne, Christophe Linhard, Rachel Torchet, Pierre Lechat, et al.

HGG Advances

ARTICLE

Trait selection strategy in multi-trait GWAS: Boosting SNP discoverability

Yuka Suzuki,[1,*] Hervé Ménager,[2] Bryan Brancotte,[2] Raphaël Vernet,[3] Cyril Nerin,[1] Christophe Boetto,[1] Antoine Auvergne,[1] Christophe Linhard,[3] Rachel Torchet,[2] Pierre Lechat,[2] Lucie Troubat,[1] Michael H. Cho,[4,5] Emmanuelle Bouzigon,[3] Hugues Aschard,[1,*] and Hanna Julienne[1,2,6,*]

Summary

Since the first genome-wide association studies (GWASs), thousands of variant-trait associations have been discovered. However, comprehensively mapping the genetic determinant of complex traits through univariate testing can require prohibitive sample sizes. Multi-trait GWAS can circumvent this issue and improve statistical power by leveraging the joint genetic architecture of human phenotypes. Although many methodological hurdles of multi-trait testing have been solved, the strategy to select traits has been overlooked. In this study, we conducted multi-trait GWAS on approximately 20,000 combinations of 72 traits using an omnibus test as implemented in the Joint Analysis of Summary Statistics. We assessed which genetic features of the sets of traits analyzed were associated with an increased detection of variants compared with univariate screening. Several features of the set of traits, including the heritability, the number of traits, and the genetic correlation, drive the multi-trait test gain. Using these features jointly in predictive models captures a large fraction of the power gain of the multi-trait test (Pearson's r between the observed and predicted gain equals 0.43, $p < 1.6 \times 10^{-60}$). Applying an alternative multi-trait approach (Multi-Trait Analysis of GWAS), we identified similar features of interest, but with an overall 70% lower number of new associations. Finally, selecting sets based on our data-driven models systematically outperformed the common strategy of selecting clinically similar traits. This work provides a unique picture of the determinant of multi-trait GWAS statistical power and outlines practical strategies for multi-trait testing.

Introduction

Despite the increasing sample size of genome-wide association studies (GWASs), many genetic variants underlying human complex traits and diseases remain undetected. To increase statistical power and detection of associations at low cost, investigators have developed various multi-trait approaches based on GWAS summary statistics.[1–6] Few studies investigated how the choice of phenotypes impacts the gain in the power of multi-trait approaches. In the standard univariate GWAS, statistical power mostly depends on minor allele frequency, sample size, the size of genetic effect, and polygenicity (the number of causal variants).[7] In the multi-trait GWASs, it additionally depends on complex characteristics of the set of traits, including their shared etiology, their genetic correlation, and the number of traits in the set. As previously described, the increased power of multi-trait test partly comes from adjusting for the correlation across GWASs due to sample overlaps and genetic relationships across phenotypes.[4,8,9] Previous works explored this question using simulated data.[8,9] However, simulations are limited in their scope and can overlook existing constraints in real data. Large-scale studies on real data are needed to better characterize scenarios increasing the gain of multi-trait tests.

Here we empirically examined how trait characteristics impact the statistical power of multi-trait GWAS. We performed our analyses on 72 curated GWAS summary statistics and analyzed the impact of 11 genetic features describing both individual and collective characteristics of sets of GWAS. Among available multi-trait methods, we conducted our primary analysis using a standard k-degree of freedom joint test (the omnibus test) as implemented in the Joint Analysis of Summary Statistics (JASS).[2] The JASS package and its associated tools solve all practical issues commonly encountered in multi-trait GWAS analyses, including missing data and computational efficiency, allowing for a large-scale power analysis on real data. We inspected the association between a range of features (heritability, polygenicity, genetic correlation, number of traits analyzed, etc), and the gain in association detection in multi-trait GWASs over univariate GWASs. We then assessed how well these features predict the gain. We further compared JASS and Multi-Trait Analysis of GWAS (MTAG),[4] an alternative multi-trait approach, in terms of absolute power gain and features that impact the gain.

[1]Institut Pasteur, Université Paris Cité, Department of Computational Biology, 75015 Paris, France; [2]Institut Pasteur, Université Paris Cité, Bioinformatics of Biostatistics Hub, 75015 Paris, France; [3]Université Paris Cité, Institut National de la Santé et de la Recherche Médicale (INSERM), UMR-1124, Group of Genomic Epidemiology of Multifactorial Diseases, Paris, France; [4]Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, 181 Longwood Avenue, Boston, MA 02115, USA; [5]Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
[6]Lead contact
*Correspondence: yuka.en25@gmail.com (Y.S.), hugues.aschard@pasteur.fr (H.A.), hanna.julienne@pasteur.fr (H.J.)
https://doi.org/10.1016/j.xhgg.2024.100319.

Material and methods

Ethics approval and consent to participate

Not applicable.

Database of curated summary statistics

We assembled a database of 72 GWAS summary statistics of quantitative traits and diseases conducted in European ancestry population pulled from the GWAS catalog[10] and a variety of publicly available meta-analyses. We cleaned, harmonized, and imputed each study using our previously developed pipeline.[2] In brief, the process includes the following steps. (1) Alignment of each GWAS to the 1000G GRCh37 reference panel, removal of strand ambiguous SNPs, and SNPs with low sample size.[11] (2) Imputation of missing summary statistics using robust and accurate imputation from summary statistics.[12] (3) Computation of the heritability, genetic, and residual covariance matrices, referred further as h^2_{GWAS}, $\Sigma_{\boldsymbol{g}}$ and $\Sigma_{\boldsymbol{r}}$, using linkage disequilibrium (LD)-score regression.[13] (4) Aggregation of curated GWAS in a unique entry file used as input for JASS. We filtered all GWASs with negative heritability, resulting in a total of 72 traits (Table S1). Curated GWAS summary statistics used in the analysis are available at https://zenodo.org/records/10876029 and can be interactively explored on the JASS webserver https://jass.pasteur.fr/.

Joint test and association gain

Multi-trait analyses were conducted using the omnibus test implemented in the JASS package.[2,8] For a set of k GWAS, the omnibus statistics is defined as $T_{omni} = \mathbf{z}^t \Sigma_{\boldsymbol{r}} \mathbf{z}$ where $\mathbf{z}$ is the vector of Z-scores across traits $\mathbf{z} = (z_1 \ldots z_k)$ and $\Sigma_{\boldsymbol{r}}$ is the residual Z score covariance derived using the LD-score regression.[13] Under the null hypothesis of no association with any of the k phenotypes, T_{omni} follows a χ^2 distribution with k degree of freedom. To maximize data use, the default setting of JASS uses all variants, even those with missing association statistics. In this case, JASS returns association p values based on the subset of Z-scores available. In contrast, the *–remove-nans* option removes variants with incomplete data. Here, we used the *–remove-nans* option as the primary analysis for a better characterization of trait sets and for a fair comparison with the default setting of MTAGs[4] that do not allow for missing statistics, whereas we also provide some results from the default setting of JASS as additional information.

All power comparisons were conducted at a locus level. The entire genome was split into a total of 1,703 quasi-independent loci defined based on LD-independent blocks, as proposed by Berisa and Pickrell.[14] For both multi-trait and univariate analyses, we obtained the minimum p-value across variants in each locus. The gain of the multi-trait test was derived as the fraction of loci whose p-values were smaller than corresponding p-values in univariate GWASs corrected for the number of traits jointly analyzed:

$$gain = f_{multi<univ} = \sum_i (P_{multi.i} < P_{uni.i} \cdot k / R), \qquad \text{(Equation 1)}$$

where R is the number of loci (R = 1,703), k is the number of traits (the number of GWASs) jointly analyzed, $P_{multi.i}$ is the minimum p value of the multi-trait test in region i, and $P_{univ.i}$ is the minimum p value of the univariate tests across all GWASs analyzed in locus i.

Estimated and derived genetic features

We investigated the effect of both single and multi-trait features. Single GWAS features include the effective sample size (N_{eff}), linear additive common variants heritability (h^2_{GWAS}), polygenicity, mean genetic effect sizes (MESs), and the proportion of uncaptured linear additive common variants heritability ($\%h^2_u$). Multi-trait GWAS features include the number of traits in a trait set (k), the average of the off-diagonal of genetic covariance, and the residual covariance ($\overline{\Sigma}_g$ and $\overline{\Sigma}_r$, respectively), condition numbers of genetic covariance matrix and residual covariance matrix (κ_g and κ_r, respectively) across traits, and the average distance between the genetic and residual correlation matrices (Δ_Σ). All parameters were aggregated to form a vector of 11 features per trait set. For the single GWAS parameters, MES, polygenicity, N_{eff}, h^2_{GWAS}, and $\%h^2_u$, we calculated mean values across each set of traits.

Polygenicity and heritability (h^2_{GWAS}) were estimated using MiXeR,[15,16] with the 1000 Genomes Phase3 reference panel provided along the MiXeR package containing approximately 10 million common variants.[11] Following the authors recommendation, we defined the parameter for effective sample size as $N_{eff} = 1/(1/N_{case} + 1/N_{controls})$. For comparison purposes, we also estimated h^2_{GWAS} using the LD-score regression,[13] and the two metrics were consistent (Pearson $r = 0.86$) (Figure S1). The estimated polygenicity by MiXeR showed a dependency on the GWAS sample size, with approximately a 10-fold increase of the polygenicity for an increase of 500,000 of the sample size (Figure S2). We, therefore, adjusted polygenicity by taking the residuals of linear regression between $\log_{10}$ polygenicity and N_{eff}: $\log_{10}\ polygenicity\ (adjusted) = \log_{10}\ polygenicity\ (mixer) - \alpha N_{eff}$, where α was estimated by a linear regression $\log_{10}\ polygenicity\ (mixer) \sim \alpha N_{eff} + \varepsilon$. We obtained the "adjusted polygenicity" as $10^{[\log_{10}\ polygenicity\ (adjusted)]}$. We computed MES as $\frac{h^2_{GWAS}}{polygenicity\ (adjusted)}$.

The proportion of uncaptured linear additive common variants heritability (% h^2_u) was derived as $(h^2_{GWAS} - h^2_{GWAS-hits})/h^2_{GWAS}$, where $h^2_{GWAS-hits}$ denotes the heritability accounted by the univariate GWAS association loci. It was derived using the lead variants from each locus reaching genome-wide significance ($p < 5 \times 10^{-8}$): $h^2_{GWAS-hits} = \sum_{i \in I} \beta_i^2$, where $\beta_i = z_i/\sqrt{N_{eff}}$. We excluded loci with lead variant whose $abs(\beta) > 0.19$, because $h^2_{GWAS-hits}$ tended to become larger than h^2_{GWAS}.

The mean genetic covariance and mean residual covariance were defined as the mean of the absolute value of the upper off-diagonal elements of the genetic and residual covariance matrices ($\Sigma_{\boldsymbol{g}}$ and $\Sigma_{\boldsymbol{r}}$), i.e., $\overline{\sigma}_g = \sum_{i,j;i<j} \sigma_{gij} / \sum_{i,j;i<j} 1$, and $\overline{\sigma}_r = \sum_{i,j;i<j} \sigma_{rij} / \sum_{i,j;i<j} 1$, respectively, where σ_{gij} and σ_{rij} are ij elements of $\Sigma_{\boldsymbol{g}}$ and $\Sigma_{\boldsymbol{r}}$, and k is the number of traits. The condition numbers of genetic and residual covariance matrices were computed as: $\kappa_g = \sqrt{\max(\lambda_{g,i})/\min(\lambda_{g,i})}$ and $\kappa_r = \sqrt{\max(\lambda_{r,i})/\min(\lambda_{r,i})}$, where $\lambda_{g,i}$ and $\lambda_{r,i}$ are the eigenvalues of the genetic and residual covariance matrices. We used *numpy.linalg.eig*[17] for the eigen decomposition and assigned an infinite value to κ when the minimum eigenvalue was negative or close to zero. The average distance Δ_Σ was defined as the mean over the absolute values of pairwise difference between the corresponding upper off-diagonal elements in genetic and residual correlation matrices $\Delta_\Sigma = \sum_{i,j;i<j} |\rho_{gij} - \rho_{rij}| / \sum_{i,j;i<j} 1$, where ρ_{gij} and ρ_{rij} indicate the ij elements in the genetic and residual correlation matrices.

Assessment of features associated with multi-trait association gain

The contribution of features to multi-trait gain was estimated using a 5-fold cross-validation (CV). For each round of CV, the 72

GWASs were randomly split into a training and validation data, each including 36 GWAS. Within each CV, we generated 1,980 unique random trait sets, containing 2–12 traits, for each of the training and validation data: random sampling out of the 36 traits (660 sets generated), random sampling out of traits with common SNP heritability below the median (330 sets) and above the median (330 sets); random sampling out of traits with MES below the median (330 sets) and above the median (330 sets). For this stratification of traits by the median of MES, we used polygenicity/h^2_{GWAS} estimated by MiXeR without adjusting for the effective sample size. For each set, we ran JASS and derived the 11 features of interest (F_i, $i = 1, \ldots, 11$) and the multi-trait gain. We used 19,793 trait sets out of the total 19,800 (= 1,980 × 2 × 5) sampled, for which the whole analysis process completed without error. Errors include cases where there was no association detected by either the univariate or the joint tests.

Moving to the multivariate regression analysis, we selected 6 of the 11 features based on collinearity analysis (as described below). The six features and the multi-trait gain were standardized into a range between 0 and 1 using *MinMaxScaler* in scikit-learn.[18] This standardization was applied at once on the entire dataset including training and validation data across the 5-fold CV sets.

We used the training data to estimate the joint effect $\hat{\delta}_i$ of each feature i from a multiple regression: $gain_{train} \sim \sum_i \delta_i F_{train,i}$. This was conducted using the *OLS* function in statsmodels in Python.[19] We report Pearson's correlation coefficients as a metric of predictive power.

Collinearity and selection of features

We observed collinearity among some of the 11 features of traits across 19,266 unique trait sets. $\text{Log}_{10}\ \kappa_r$, $\log_{10}\ \kappa_g$, $\log_{10}\ \overline{\Sigma}_r$, and the number of traits were highly correlated (Pearson $r > 0.65$). Likewise, the mean $\%h^2_u$, mean $\log_{10}$MES, and mean $\log_{10}$polygenicity were highly correlated (abs(r) > 0.8). These correlated features capture redundant characteristics of traits. Thus, we selected one of each of the correlated features: the number of traits and mean $\%h^2_u$. We chose the number of traits because it had the smallest p value in a multivariate linear regression with all the features included, where inf values in condition number were replaced with their non-inf max value. We chose the mean($\%h^2_u$) because it captures both mean($\log_{10}$MES) ($r = -0.90$) and mean ($\log_{10}$polygenicity) ($r = 0.84$). The estimation of the mean($\%h^2_u$) is also more straightforward than the mean($\log_{10}$MES) and mean ($\log_{10}$polygenicity). After this pre-selection, we built models with the remaining six features as described in the previous section.

Non-linear models

We considered two alternative non-linear models for prediction purposes: support vector regression (SVR), and random forest regression (RFR). SVR and RFR are regression approaches that allow for non-linear relationships. The goal of SVR is to find a hyperplane (or line, in the case of two-dimensional data) that best fits the data. It is effective at handling non-linear and complex data by using the kernel trick—mapping data with a kernel function into a higher-dimensional space where it is easier to find the best-fit hyperplane. SVR also penalizes the complexity of the model and gives the flexibility in how much error is acceptable. RFR performs a regression using decision trees. It generates multiple trees, fits each to a random subset of training data, and averages the predictions across trees as the final prediction. While the random sampling and averaging supposedly makes the model robust to outliers, RFR's performance relies on a high quality of training data; the training data need to cover a wide range, as RFR does not work well for extrapolation, and RFR leads to biased predictions when the training data are sampled in a biased way.[20] In contrast, SVR is suggested to be capable of extrapolation.[21] We used the scikit-learn[18] python implementation of RFR and of the SVR. We fitted SVR and RFR models to the $gain_{train}$ including hyperparameters. Hyperparameters were tuned using *RandomizedSearchCV* in scikit-lean across the following range: SVR's kernel = [linear, rbf, sigmoid, poly], C = [1,10,50,100], epsilon = [10^{-3},10^{-2},10^{-1},1], degree = [2,3,4], and RFR's n_estimators = [5,20,50,100], max_features = ['auto','sqrt'], max_depth = [12 values ranging from 2 to 100], min_samples_split = [2,5,10], min_samples_leaf = [1,2,4], bootstrap = [True, False].

MTAG analyses

For comparison purposes, we repeated the multi-trait test and the prediction analyses with the MTAG approach.[4] MTAG uses a weighted sum of Z score (Note S1). Trait weights are derived using the generalized method of moments, as $(\hat{\beta}_j \hat{\beta}_j' - \Omega - \Sigma_j) = 0$, where Ω is the genetic covariance matrix to be estimated for the weights, and Σ_j is the genetic covariance matrix estimated using the LD score regressions by Bulik-Sullivan et al.[13] The model assumes that the genetic covariance matrix is homogeneous across variants. We used MTAG with its default setting, which considered only complete cases. We ran MTAG for the same trait sets used for the analysis with JASS (all data used in the 5-fold CV). MTAG outputs p values for each trait in each trait set, whereas JASS gives a single p value for a trait set. To account for the number of tests run by MTAG for one set, we obtained minimum p values by MTAG across traits and variants in each locus and multiplied the minimum p values by the number of traits in the set. For the comparison of the association gain between MTAG and JASS, we used both the minimum p values with and without multiple test correction (Figures 5 and S3).

Simulation studies on the calibration and statistical power of the MTAG and JASS tests

To assess the statistical power and type 1 error of the JASS and MTAG tests, we designed the following simulation scenarios. We simulated N = 50K individuals with five outcomes $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_5)$:

$$\boldsymbol{Y} = \boldsymbol{M}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{M}$ is the genotype matrix, β is the matrix of genetic effects and $\boldsymbol{\varepsilon}$ is the matrix of residuals. $\boldsymbol{M}$ is a matrix 1,000 × 50k representing 1,000 independent SNPs drawn from a binomial distribution (minor-allele frequency uniformly distributed between 0.05 and 0.95). The β matrix was set according to two scenarios: "null hypothesis," $\boldsymbol{\beta} = \boldsymbol{0}$ and $h^2 = 0$; "alternative hypothesis," the β matrix was drawn from a centered multivariate with a covariance matrix set to $\begin{pmatrix} 1 & \ldots & 0.3 \\ \vdots & \ddots & \vdots \\ 0.3 & \ldots & 1 \end{pmatrix} \times 0.25/1000$. The vectors of residuals $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon_1}, \boldsymbol{\varepsilon_2}, \boldsymbol{\varepsilon_3}, \boldsymbol{\varepsilon_4}, \boldsymbol{\varepsilon_5})$ were drawn from a multivariate normal with means of 0, a covariance matrix with diagonal elements equal to $1 - h^2$, and non-diagonal elements r_e varying between 0 and 0.5 across simulations. For each simulation (scenario × the number of environmental covariances considered), we derived the Z score for each of the m variants as $\mathbf{Z} = (\hat{\boldsymbol{\beta}}_{\mathbf{1}}\sqrt{N},\ \hat{\boldsymbol{\beta}}_{\mathbf{2}}\sqrt{N},\ \hat{\boldsymbol{\beta}}_{\mathbf{3}}\sqrt{N},\ \hat{\boldsymbol{\beta}}_{\mathbf{4}}\sqrt{N},$

$\widehat{\beta}_5\sqrt{N}$). Then, the JASS and MTAG tests were applied to the Z score matrices.

Comparison of strategies for trait selection

To compare the performances of trait selection strategies, we classified the 19,266 unique sets into clinically homogeneous, low heterogeneity, high heterogeneity, or high predicted gain according to our predictive model. To assess clinical homogeneity, we first classified the 72 traits into clinical groups using the broadest categories in the MeSH Tree Structures.[22] The grouping was further refined based on clinician's insights (Table S1). We labeled each set of traits as "homogeneous" if all the traits are in the same clinical group, "low heterogeneity" if trait belonged to two to four clinical groups, and "high heterogeneity" if traits spanned five or more clinical groups. For the data-driven method, we selected 100 sets of traits by CV fold that had the largest gains predicted by the linear model with aggregated coefficients across CV folds (Table 1). We used the Welch's t test to evaluate the differences in gain and the number of new associated loci. We used a two-sided Welch's t tests to determine whether the data-driven method achieves a greater association gain than other methods, and whether jointly analyzing clinically heterogeneous traits achieves a greater association gain than jointly analyzing clinically homogeneous traits. For this test, we used a pair of training data and validation data that are mutually exclusive to ensure independence.

Evaluation of the relevance of new association

To evaluate the relevance of new associations detected by JASS (i.e., if most of them were true positive), we attempted to predict loci discovered in a recent large meta-analysis on body mass index (BMI) (sample size of 683,365 on average across ~2.3 million variants[23]) from the results of multi-trait GWAS applied on 1,776 trait sets containing a smaller study of BMI (sample size of 339,224) (Table S1). First, we compared loci detected by JASS (after a Bonferroni correction to account for the number of sets) and in the larger GWAS using the standard genome wide significance threshold of 5×10^{-8}. Second, we fitted a logistic regression to predict associated loci in the larger GWAS by combining JASS p value and the number of sets where the loci was considered associated with JASS. We use odd number chromosomes to fit the logistic regression and evaluated its performances on even number chromosomes.

Results

Study overview

We conducted a series of analyses to identify features that influence the gain in association detection of multi-trait GWAS relative to univariate GWAS. The key principles and main steps of the study are depicted in Figure 1. We used 72 curated GWAS summary statistics spanning a range of clinical domains (Table S1) and considered 11 features describing the univariate and multivariate characteristics of the GWASs analyzed. Single GWAS features are MES, polygenicity, effective sample size (N_{eff}), linear additive heritability of common variants (h^2_{GWAS}), and the proportion of uncaptured linear additive heritability of common variants ($\%h^2_u$). The multi-trait GWAS features are the number of GWAS analyzed jointly and five metrics related to the genetic (Σ_g) (Table S2 and Figure S4) and residual (Σ_r) (Table S3 and Figure S5) covariance matrices, where the latter represents the covariance between the Z score statistics of two GWAS in the absence of genetic effect at the loci (due to sample overlap and phenotypic correlation). Those five features are the mean of the off-diagonal terms denoted as $\overline{\Sigma}_g$ and $\overline{\Sigma}_r$; the conditional numbers of Σ_g and Σ_r denoted as κ_g and κ_r, which we used as a measure of multi-collinearity; and the average difference between the two matrices Δ_Σ. Previous works highlighted Δ_Σ as a key driver of the power of multi-trait test[8,9] (Notes S1, S2, and S6).

Given the 72 traits, there are more than 4.7×10^{21} possible sets of 2–72 traits. In this study, we used 19,266 unique random sets of 2–12 GWAS. The maximum number of traits included in sets was constrained by two requirements: (1) dividing the data in two to form a training and a validation set, and (2) avoiding drawing sets that are nearly identical combinations (e.g., 2 sets of 35 traits selected from an ensemble of 36 traits share 34 traits). For these reasons, we fixed the maximum number of traits (12 traits) to a relatively small number compared with the total number of traits in the study (72 traits). Sets were drawn using a stratified sampling conditional on mean effect size, heritability, and the number of traits, in order to maximize the range of the genetic features studied (see Material and methods and Note S3). For each set, we derived the average of the single GWAS feature (polygenicity, MES, N_{eff}, $h^2{}_{GWAS}$, and $\%h^2{}_u$) and the six set features ($\overline{\Sigma}_g$, $\overline{\Sigma}_r$, κ_g, κ_r, Δ_Σ, and the number of GWAS selected). Note that we used the log of polygenicity, MES, $\overline{\Sigma}_g$, $\overline{\Sigma}_r$, κ_g, κ_r, and Δ_Σ when investigating their association with other variables due to their skewed distributions (Figure S7). We compared multi-trait and univariate GWAS results based on the minimum p value of the multi-trait tests across all variants within LD-independent loci (P_{JASS}) and the minimum p value of univariate tests across all variants and all GWASs within the same loci (P_{uni}). We used two metrics: (1) the proportion of significant loci found associated at genome-wide level ($p < 5 \times 10^{-8}$) by JASS and missed by the univariate GWAS, and (2) the fraction of loci where P_{JASS} was smaller than P_{uni} ($f_{JASS<univ}$).

We first describe the distribution of the genetic features and their correlation. Second, we applied JASS to each set and compared multi-trait and univariate GWAS association results (Figure 1B). Third, we build a predictive model of the association gain using a 5-fold CV approach (Figure 1C). For each CV, we pulled two independent datasets of 1,980 sets out of the 19,266 unique sets, one for training and one for validation. We conducted a regression analysis to estimate the joint contribution of the features in the training dataset, and measured the correlation between observed and predicted gain in the validation dataset. Fourth, we reran the analyses using MTAG,[4] a popular multi-trait approach that leverages genetic correlation

Table 1. Coefficients of the multivariate linear regression models from 5-fold CVs

	CV1	CV2	CV3	CV4	CV5	Mean ± SD
No. of traits k	0.096 ($p = 3.447 \times 10^{-44}$)	0.077 ($p = 1.592 \times 10^{-17}$)	0.082 ($p = 1.638 \times 10^{-44}$)	0.049 ($p = 6.148 \times 10^{-8}$)	0.083 ($p = 3.071 \times 10^{-19}$)	0.077 ± 0.017
Mean $\log_{10}\Delta_\Sigma$	−0.787 ($p = 8.338 \times 10^{-29}$)	−0.536 ($p = 8.338 \times 10^{-29}$)	−0.393 ($p = 2.242 \times 10^{-10}$)	−1.174 ($p = 1.086 \times 10^{-26}$)	−0.610 ($p = 1.947 \times 10^{-9}$)	−0.700 ± 0.301
Mean $\log_{10}\overline{\Sigma}_g$	0.767 ($p = 3.813 \times 10^{-45}$)	0.789 ($p = 2.055 \times 10^{-20}$)	0.355 ($p = 5.990 \times 10^{-15}$)	1.149 ($p = 1.604 \times 10^{-56}$)	0.675 ($p = 5.636 \times 10^{-21}$)	0.747 ± 0.284
Mean N_{eff}	0.206 ($p = 4.342 \times 10^{-15}$)	−0.016 ($p = 0.554$)	−0.071 ($p = 0.003$)	0.260 ($p = 1.099 \times 10^{-10}$)	−0.015 ($p = 0.598$)	0.073 ± 0.149
Mean h^2_{GWAS}	−0.509 ($p = 2.937 \times 10^{-96}$)	−0.437 ($p = 1.455 \times 10^{-34}$)	−0.518 ($p = 2.216 \times 10^{-111}$)	−0.690 ($p = 6.670 \times 10^{-79}$)	−0.429 ($p = 2.225 \times 10^{-26}$)	−0.516 ± 0.105
Mean $\%h^2_u$	0.112 ($p = 5.158 \times 10^{-6}$)	0.370 ($p = 4.353 \times 10^{-35}$)	0.091 ($p = 3.877 \times 10^{-10}$)	0.277 ($p = 3.022 \times 10^{-33}$)	−0.064 ($p = 0.002$)	0.157 ± 0.169

Mean genetic residual distance and mean genetic covariance were $\log_{10}$ transformed. All the features were scaled using *MinMaxScaler* in *scikit-learn*.[18]

across traits to boost statistical power, and compared results with JASS. Finally, we compared common trait selection strategies—e.g., choosing clinically homogeneous or heterogeneous traits—in their impacts on the association gain to evaluate our model prediction and provide a practical strategy in multi-trait test (Figure 1D).

Distribution of the genetic features across GWAS trait sets

The individual genetic features of the 72 studied traits were distributed as follows. The sample sizes ranged from 5,318 to 697,828 with a median of 85,559. Heritability (h^2_{GWAS}) ranged from 1% to 48% with a median of 10% (Figure 2A) and was consistent across the software used for its estimation (Figure S1). Polygenicity and MES were highly variable: polygenicity (i.e., the estimated number of causal variants) ranged from 6.9 (estimated glomerular filtration rate from cystatin C) to 570,102 (variability sleep duration) with a median of 1,110, and MES ranged from 4.48×10^{-8} (variability sleep duration) to 3.3×10^{-3} (fasting proinsulin) with a median of 1.0×10^{-4}. The distribution of the average of these parameters across the 19,266 sets is presented in Figure 2B along with the GWAS set features. The latter metrics ranged (in 25–75 percentiles) as −2.04 to −1.79 for $\log_{10}\overline{\Sigma}_g$, −1.96 to −1.44 for $\log_{10}\overline{\Sigma}_r$, 0.36 to 0.78 for $\log_{10}\kappa_g$, 0.05 to 0.32 for $\log_{10}\kappa_r$, and −1.11 to −0.89 for $\log_{10}\Delta_\Sigma$. In particular, the variability in $\log_{10}\Delta_\Sigma$ was limited (given the theoretical upper bound of $\log_{10}\Delta_\Sigma$ is $\log_{10} 2 \cong 0.3$).

We measured the correlation across features at the trait and set levels (Figures 2C and 2D). The proportion of uncaptured linear additive heritability of common variants ($\%h^2_u$) was strongly associated with $\log_{10}$polygenicity ($\rho = 0.76$, $p = 8.8 \times 10^{-15}$) and with $\log_{10}$MES ($\rho = -0.8$, $p = 3.5 \times 10^{-17}$) (Figures 2C and S8), in agreement with previous reports showing that univariate GWAS performs better for traits with a larger MES and a smaller polygenicity.[7] As expected, we observed similar correlation at the single trait and sets levels (e.g., mean $\%h^2_u$ and mean polygenicity are highly correlated as are $\%h^2_u$ and polygenicity) (Figure 2D). Metrics related to genetic and residual covariance matrices were positively correlated with each other ($\rho = 0.47$). The correlation between Σ_g and Σ_r and associated parameters is expected as the latter is a function of sample overlap and phenotypic correlation, itself depending on genetic correlation.

Multi-trait versus univariate GWAS across 19,266 random sets

We applied JASS to all 19,266 sets and quantified the gain in association detection of the multi-trait against univariate test. On average, univariate tests identified 285 loci per set, and JASS detected an additional 26 loci, corresponding with a 1.1-fold increase in power (Figure S9A). JASS gain was maximal for sets with a relatively low number (<300 loci) of previously detected loci (Figure S9B). JASS detected at least one new association in 98% of the trait sets (18,787 sets) (Figure S9C) and 508,829 new associations in total (note that these can be overlapping loci). These numbers are obtained when applying JASS on variants with beta coefficients available for all traits in the set. When analyzing all variants including those with missing values as allowed by JASS (see Material and methods), approximately 1.4 times more new associations were detected (693,382 new associations in total by JASS, including variants with missing values) (Figure S9D).

We assessed the marginal relationship between genetic features and the number of loci detected by the univariate and multi-trait GWAS tests, and the association gain of multi-trait test. Figure 3 presents the correlation between each feature and the number of univariate and multi-trait GWAS-associated loci. As expected, the number of univariate associated loci was positively correlated with the number of traits, mean N_{eff}, the mean h^2_{GWAS}, and mean MES, and negatively correlated with mean uncaptured linear additive heritability (mean $\%h^2_u$). Multi-trait GWAS gain over univariate GWAS was positively associated with mean polygenicity and mean $\%h^2_u$, and negatively associated with mean MES.

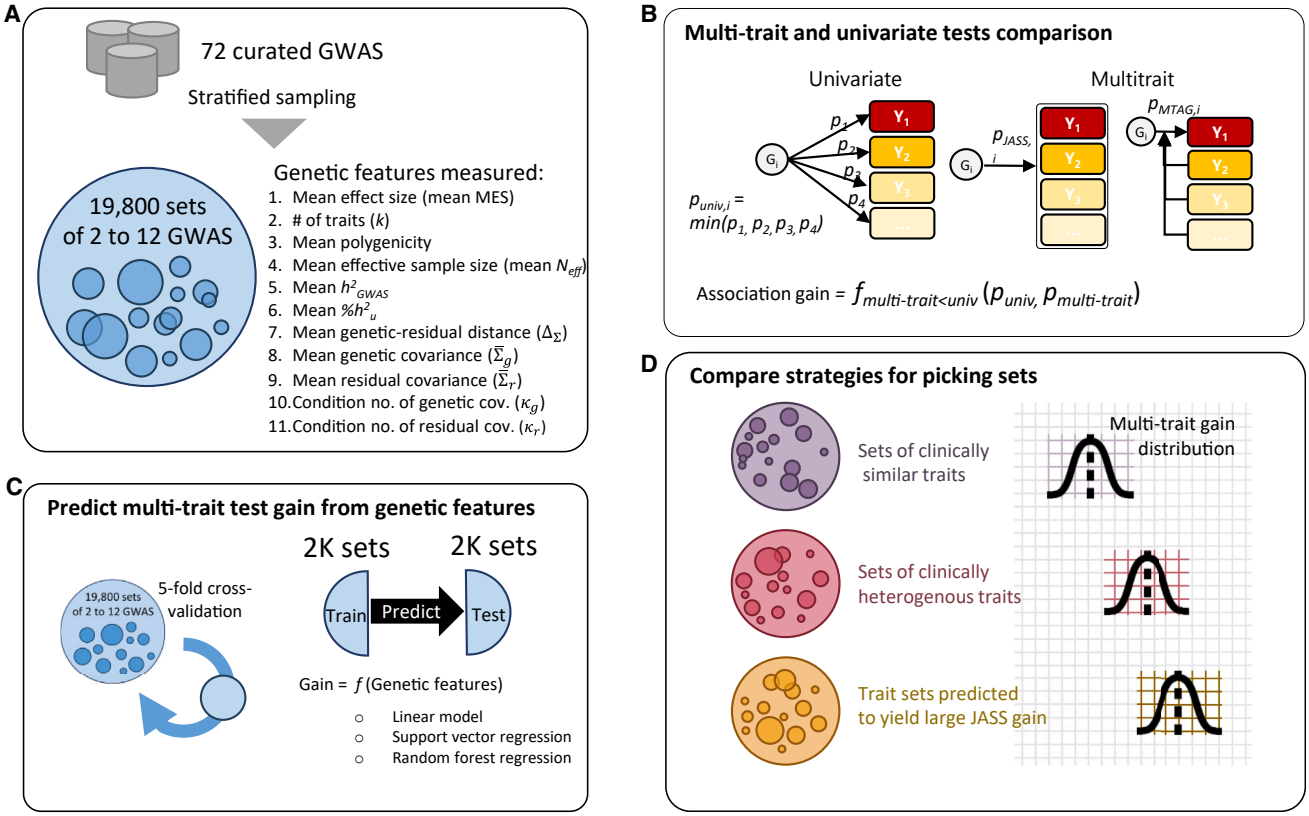

Figure 1. Study overview

We conducted a power analysis on real data to understand in which setting a standard multi-trait test—the omnibus test—outperforms univariate GWAS.

(A) To assemble our real data, we curated 72 GWAS summary statistics and formed approximately 20k sets of traits by random sampling. Each set of traits was characterized by assessing key genetic features such as polygenicity, MES, and heritability.

(B) For each set of traits, we ran omnibus test using JASS and computed the association gain compared with the univariate test. We defined this association gain as the number of LD-independent loci where the omnibus yields a smaller p value than the univariate test. We repeated the analysis using MTAG, a popular multi-trait approach.

(C) To investigate which genetic features (highlighted in A) explain the JASS (omnibus) association gain, we applied statistical models to predict the gain as a function of genetic features. Several models were benchmarked to optimize prediction performances.

(D) To suggest a practical strategy for selecting traits that yield a large association gain, we compared the performance of JASS on trait sets that are clinically similar, clinically heterogeneous, and that were predicted to have a large gain by the predictive model highlighted in (C).

The mean h^2_{GWAS} showed an opposite effect, being positively associated with the number of univariate associated loci, but negatively associated with gain. Overall, this suggests that the multi-trait test can be highly complementary to the univariate test, performing better in situations where the univariate test displays low power. We noted in a recent study[24] that a high multicollinearity of the matrix underlying the null hypothesis (Σ_r) can lead to a lack of robustness of the omnibus test.[24] We checked how the condition number κ_r was related to JASS gain (Figures 3B and S10). The condition number stayed in a reasonable range for 99% of sets (min = 1, max = 11). Furthermore, note that the trait sets contained overlapping traits and are therefore not fully independent from each other.

Predicting multi-trait test gain from genetic features

To robustly assess the impacts of genetic features on the multi-trait gain, we conducted regression analyses with a CV scheme establishing independence between training and validation data (Figure S11). We investigated the predictive power of the association gain ($f_{JASS<univ}$) from a joint modeling of the genetic features (Figure 4). Note that some features are almost linear combinations of other ones (e.g., Δ_Σ is proportional to the difference between $\overline{\Sigma}_g$ and $\overline{\Sigma}_r$). To avoid extreme collinearity and ensure parsimony, we selected six moderately correlated features: the number of traits, $\log_{10}\Delta_\Sigma$, $\log_{10}\overline{\Sigma}_g$, mean N_{eff}, mean h^2_{GWAS}, and mean $\%h^2_u$ (see Material and methods). We favored mean $\%h^2_u$ over mean MES and mean polygenicity, since mean $\%h^2_u$ captures both $\log_{10}$(MES) and $\log_{10}$(polygenicity) (Figure 2). We assessed performances using a 5-fold CV. For each CV, the model parameters were derived on a training dataset, and the prediction accuracy was derived in an independent validation dataset (Material and methods) (Figure S11). All six features were highly associated with the multi-trait gain (Table 1). The mean

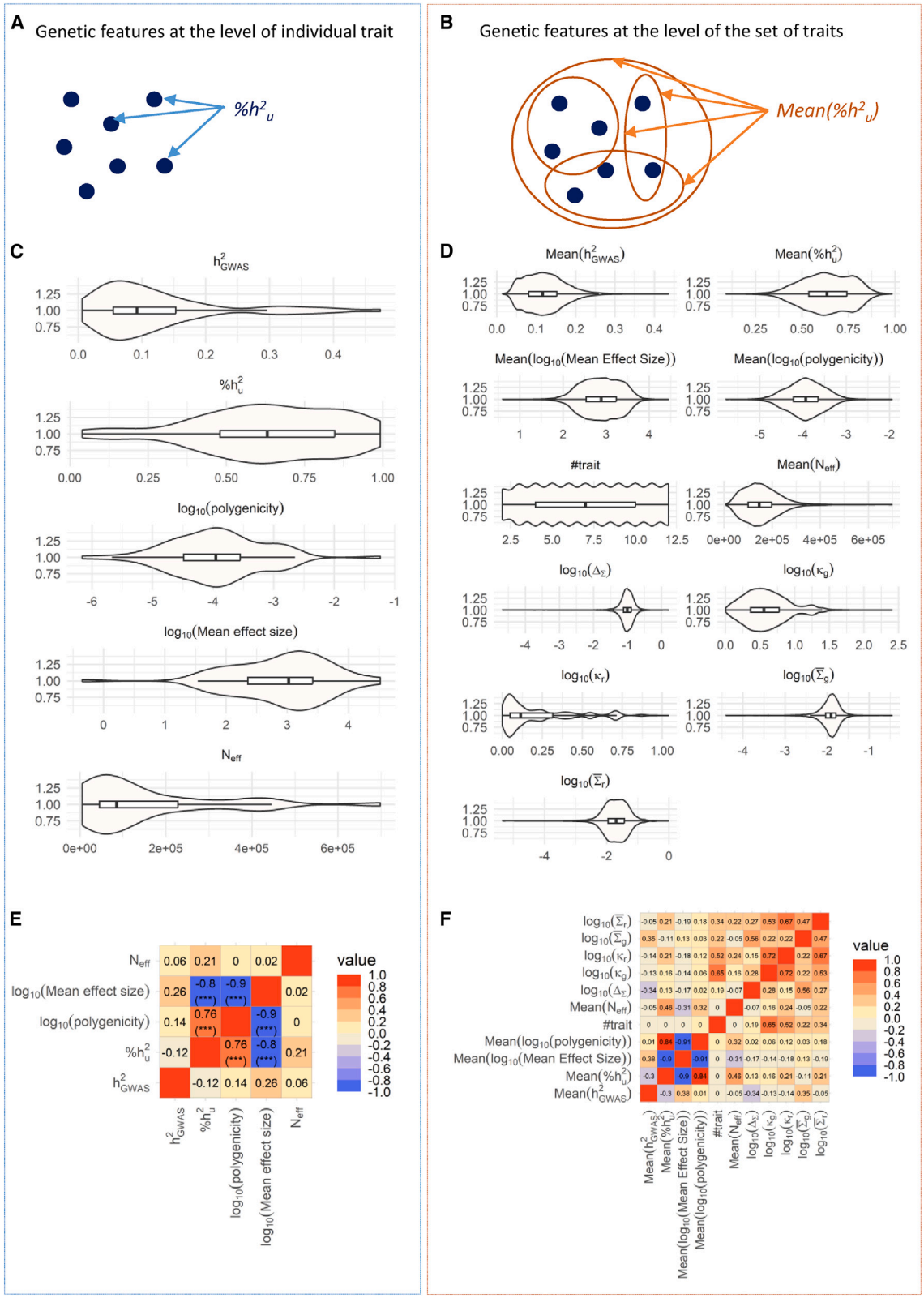

Figure 2. Genetic features characteristics derived from 72 traits and 19,266 random trait sets

Visualization of the investigated features and their relation at the level of individual trait (A, C, and E) and at the level of set of traits (B, D, and F).

(A) Schematic of a genetic feature derived at the level of an individual trait.

(B) Schematic of a genetic feature derived at the level of a set of traits.

(C) Violin plots representing the distribution across the 72 summary statistics of heritability, polygenicity, MES, and sample size of the study.

(legend continued on next page)

$\%h^2_u$ was overall positively associated with multi-trait gain, whereas the mean h^2_{GWAS} was negatively associated. Overall, the association gain of multi-trait test, more specifically the omnibus test, seems driven by genetic correlation, polygenicity, and the number of traits. The predicted gain was significantly correlated with the observed gain on validation data (median Pearson $r = 0.43$, $p < 1.6 \times 10^{-60}$) (Figure S12). A command (*jass predict-gain*) has been added to the command-line toolset of the JASS Python package to allow users to score combinations of traits using this regression model (Data and code availability).

We conducted a series of sensitivity analyses to explore further performances. First, to interpret this model behavior from the perspective of polygenicity and MES, we fitted two models replacing mean $\%h^2_u$ with mean $\log_{10}$(polygenicity) and mean $\log_{10}$(MES) in turn (Figure S13). In these models, mean $\log_{10}$(polygenicity) had a positive contribution to the gain, whereas mean $\log_{10}$(MES) had a negative contribution. In other words, multi-trait tests likely detect new associations in settings where univariate tests perform poorly, confirming the correlation analysis (Figure 3). Additionally, the model suggests that the number of traits and $\log_{10}\overline{\Sigma}_S$ further enhance multi-trait test gain. In contrast with previous observations in simulation studies, $\log_{10}\Delta_\Sigma$ likely diminishes the multi-trait test gain (see Notes S4 and S5 for a hypothetical explanation). Second, we also considered two nonlinear models for comparison purposes, SVR and RFR. As shown in Figure S14, these two models seem to outperform the multivariate linear model on the training datasets. However, they performed similarly or worse on the validation dataset, suggesting a strong overfitting in the training data. Overall, we did not find any benefit in using these more complex models.

Comparison of JASS versus MTAG

We repeated the association screening and the prediction analysis using MTAG, a popular multi-trait approach leveraging genetic correlation among closely related traits to inform GWAS screening. Regarding the association screening, MTAG detected 153,061 new association regions in total across the sets (30% of the number of new associations detected by JASS). On average, MTAG detected eight new loci per set, and at least one new association locus on 63.3% of sets (12,195 of 19,266) compared with 98% of sets for JASS. In 93% of all the trait sets, MTAG detected fewer associations than JASS (Figure 5A). The performance difference further increased when applying JASS also on variants with missing values (in 96% of trait sets MTAG detected fewer associations). Despite these discrepancies in the number of association loci, the number of new association loci in MTAG and JASS were strongly correlated (Pearson $r = 0.72$, $p < 2.2 \times 10^{-308}$), and as well as the *p*-values of MTAG and JASS for the same loci (Pearson $r = 0.75$, $p < 2.2 \times 10^{-308}$). This concordance suggests common determinants for statistical power between the two methods. We next fitted a multivariate linear model to predict MTAG gain from the same six genetic features and training data used for JASS (Figure S15). The most notable difference from JASS was that the number of traits in the set was, consistently across CV folds, negatively associated with the gain of MTAG. Indeed, JASS particularly outperformed MTAG on larger set of traits (Figure 5B).

To test whether the pronounced gain of JASS compared with MTAG on large sets was mostly due to multiple testing correction—as MTAG performs one test per trait in the set of traits, multiple testing correction is applied to MTAG *p* values (Material and methods)—we compared the number of associations detected by JASS and by MTAG without correction (Figure S3). In the absence of multiple testing correction, MTAG type 1 error expectedly increased with the number of traits in the set (e.g., genomic inflation factor λ was 1.3, 1.66, 1.72, and 1.77 for 4 cases with 2, 5, 9, and 12 traits). Despite the increased *p* value inflation for large sets, MTAG detected fewer associations overall than JASS (Figure S3A) on most of the sets (in 71% and 83% of sets when excluding and including variants with partly unavailable summary statistics, respectively). The performance of JASS and MTAG were most similar in small trait sets (Figure S3B), while JASS was particularly advantageous when analyzing set with large number of traits, especially when allowing for variants with missing summary statistics.

To ensure that the higher number of associations detected by JASS was due to a higher statistical power and not to a higher type 1 error rate, we simulated *Z* scores under the null hypothesis of no genetic effects and under the alternative hypothesis (Material and methods). The JASS test is properly calibrated under the null, whereas MTAG showed a slight deflation when adjusted for multiple testing and a substantial inflation when left unadjusted (Figure S16). We simulated Z-scores under the alternative assuming a heritability of 0.25 and a genetic correlation of 0.3 across five traits (Material and methods). JASS had greater statistical power than MTAG under the alternative hypothesis, even when MTAG was left unadjusted for multiple testing (Figure S17).

An informed strategy for trait set selection in multi-trait GWAS

A common and seemingly sound strategy when conducting multi-trait analyses is to use closely related traits. This choice is partly driven by investigators' interest in

(D) Violin plots representing the distribution across the 19,266 sets of traits of the 11 genetic features derived for each trait set.
(E) Pearson correlation among polygenicity, MES, h^2_{GWAS}, $\%h^2_u$, and sample size across 72 traits.
(F) Pearson correlation among the 11 features across 19,266 trait sets. (q-value annotation: *** $<10^{-3}$, ** $<10^{-2}$,* $< 5\text{x}10^{-2}$). Note that Pearson correlations displayed on panel E and F have been rounded up to keep 2 significant numbers.

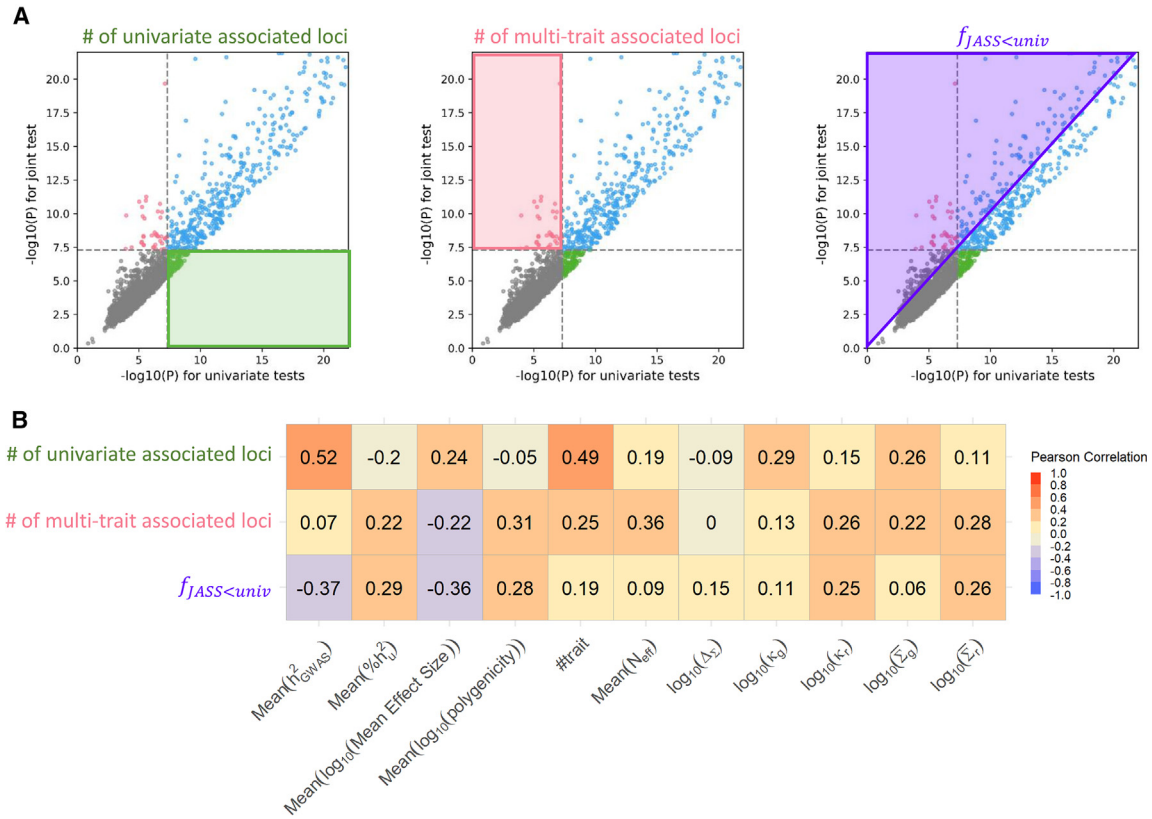

Figure 3. Determinant of JASS gain across trait sets

(A) Illustration of the three metrics to assess univariate and multi-trait GWAS outcomes. On a quadrant plot representing the *p*-value of the multi-trait test with respect to the *p* value of the univariate test, the following areas represent regions where: (green) only the univariate test is significant, (pink) only the multi-trait test is significant, (purple) the multi-trait test is more significant than the univariate test.

(B) Heatmap of the Pearson correlation between the number of univariate association loci, the number of new association loci detected by JASS, the association gain of JASS ($f_{JASS<univ}$) and the 11 genetic features across 19,266 trait sets.

delineating the shared genetic etiology between a disease and closely related phenotypes. This might also arise from the intuitive idea that closely related phenotypes share a fair amount of genetic etiology, which can be leveraged through multi-trait testing. However, its impact on statistical power has not been evaluated. To advise investigators on the best strategy to compose sets, we compared multi-trait gain and the number of new association loci obtained using five strategies. We evaluated three clinically-driven strategies, a random baseline, and a data-driven strategies (see Material and methods): (1) including GWAS from the same clinical group (noted "homogeneous"), (2) including GWAS from two to four clinical groups (noted "low heterogeneity"), (3) including GWAS from five clinical groups or more (noted "high heterogeneity"), (4) select sets at random, and (5) a data-driven approach based on the linear regression predicted gain (Material and methods). The data-driven strategy had a higher gain and larger number of new association loci compared with the other strategies (Figures 6A and 6B). The gain and the number of new association loci increased systematically with the clinical heterogeneity of the traits, and the increases were statistically significant for most pairs of trait selection strategies compared, especially when comparing the data-driven approach with the other approaches ($p < 4.4 \times 10^{-8}$ for gain and $p < 6.7 \times 10^{-7}$ for the number of new association loci). Interestingly, the selecting traits at random significantly outperformed the clinically homogeneous strategy in terms number of new associations discovered. The average number of new association loci detected in validation data equals 14, 20, 32, 22, and 61 for homogeneous, low heterogeneity, high heterogeneity, random baseline, and data-driven sets, respectively. In addition, we examined whether grouping traits based on heritability would lead to similar conclusions (i.e., favoring heterogeneous groups of traits and the data-driven strategy). Consistent with previous results, sets composed of traits with low heritability had a higher gain and a higher number of new associations than sets composed of traits with high heritability (Figure S18). Interestingly, heterogeneous sets in terms of heritability discovered more new associations than low-heritability sets, reinforcing the interest of heterogeneous sets of traits when performing a multi-trait GWAS with JASS. Nevertheless, the data-driven strategy outperformed other strategies in terms of the number of new associations discovered.

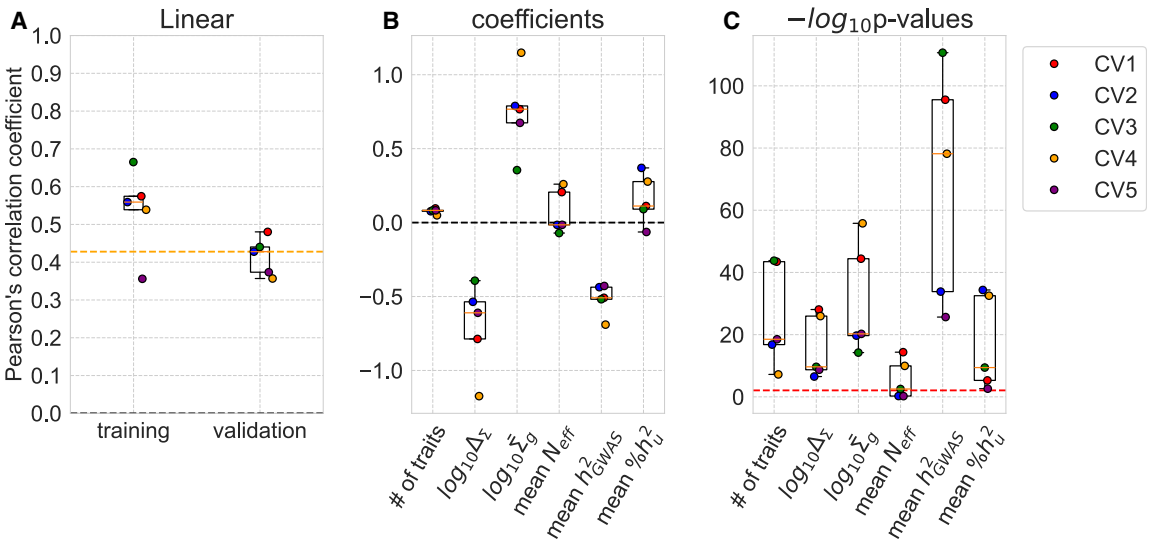

Figure 4. Model prediction power and feature contributions

(A) Boxplots of the prediction power across the 5-fold CVs of the multivariate linear regression model measured as the Pearson's correlation coefficient between the predicted and observed gain. The performance of each CV is represented as a colored dot. Orange dashed line: median correlation coefficient between the predicted and observed gain in the validation data.

(B and C) The boxplots show the coefficients and -$\log_{10}(p$ values) of the six features in the regression model across 5-fold CVs using each corresponding training data. Red dashed line, Bonferroni corrected nominal significance threshold.

We further decomposed the gain of the homogeneous sets by clinical groups (Figures 6C and 6D). Two groups consistently yielded more new association loci than others, namely, "circulatory and respiratory physiological phenomena" and "physiological phenomena." The first group was composed of spirometry traits and asthma, characterized by a large sample size, substantial genetic correlation, and moderate heritability, a favorable setting for JASS. The second group contained anthropometric traits (height, BMI, hip circumference, waist circumference, and waist-to-hip ratio). We next investigated whether multi-trait gain was associated with specific traits and derived the fold enrichment for traits between the top 10% yielding sets (corresponding with >169 new associated loci by JASS) and the least 10% yielding sets. BMI and hip circumference were amongst the top traits (Figure S19). Notably, we found that high yielding trait sets (i.e., those that yielded ≥169 new association regions by JASS) all contained BMI. In contrast, of the remaining trait sets, only 9% of the trait sets contained BMI. Other traits enriched in top sets included spirometry and to a lesser extend mental disorders, arterial pressure, and sleep patterns. BMI genetics is increasingly recognized to be a complex entanglement of metabolic and behavioral factors,[25] and suggests that complex traits that reflect multiple biologic processes may benefit more from multi-trait analysis.

We additionally compared the multi-trait gain and the number of new association loci by MTAG across the four groups of trait sets. The results were rather opposite to what we observed above with JASS: MTAG gain increased as the trait sets became more homogeneous (Figure S20A) ($p < 1.4 \times 10^{-7}$), and the number of new association loci was greater for trait sets of low heterogeneity than high heterogeneity (Figure S20B) ($p = 2.7 \times 10^{-6}$). We then tested whether MTAG outperforms JASS on homogeneous trait sets (Figures S20C and S20D) and found that there was little difference between the two ($p = 0.084$ for gain and $p = 0.063$ for the number of new association loci). We further tested whether MTAG outperforms JASS on homogeneous trait sets of certain clinical groups (Figures S20E and S20F). JASS significantly outperformed on "immune system diseases" and "psychological phenomena" (immune system diseases: $p = 8.2 \times 10^{-4}$ for gain and $p = 6.7 \times 10^{-7}$ for the number of new association loci; psychological phenomena: $p = 2.5 \times 10^{-5}$ for the number of new association loci). In contrast, MTAG outperformed on "musculoskeletal and neural physiological phenomena" (which contains only one trait set). To demonstrate the validity of the new association detected by JASS, we conducted a replication analysis using BMI as a case study. In practice, we tested if we could predict novel associations observed in a larger study (sample size = 683,365 [23]) from multi-trait GWAS applied on the BMI study in the present analysis (sample size = 339,224[26]) (Table S1). Across the 1,776 sets containing BMI, JASS detected 1,167 new associations (after Bonferroni correction, Material and methods) of which 537 corresponded with a new association in the larger BMI GWAS. Eighty-six associations of the larger GWAS were missed by JASS. Hence, JASS was able to flag loci with a high recall (0.86, probability of new association detected in the larger GWAS to be detected by JASS) but a moderate specificity (0.46, probability of a detected loci to be associated in the larger GWAS). This can be explained by the generality of the null hypothesis used in JASS, which requires only one trait in the set to be significant (not necessarily BMI). To improve specificity,

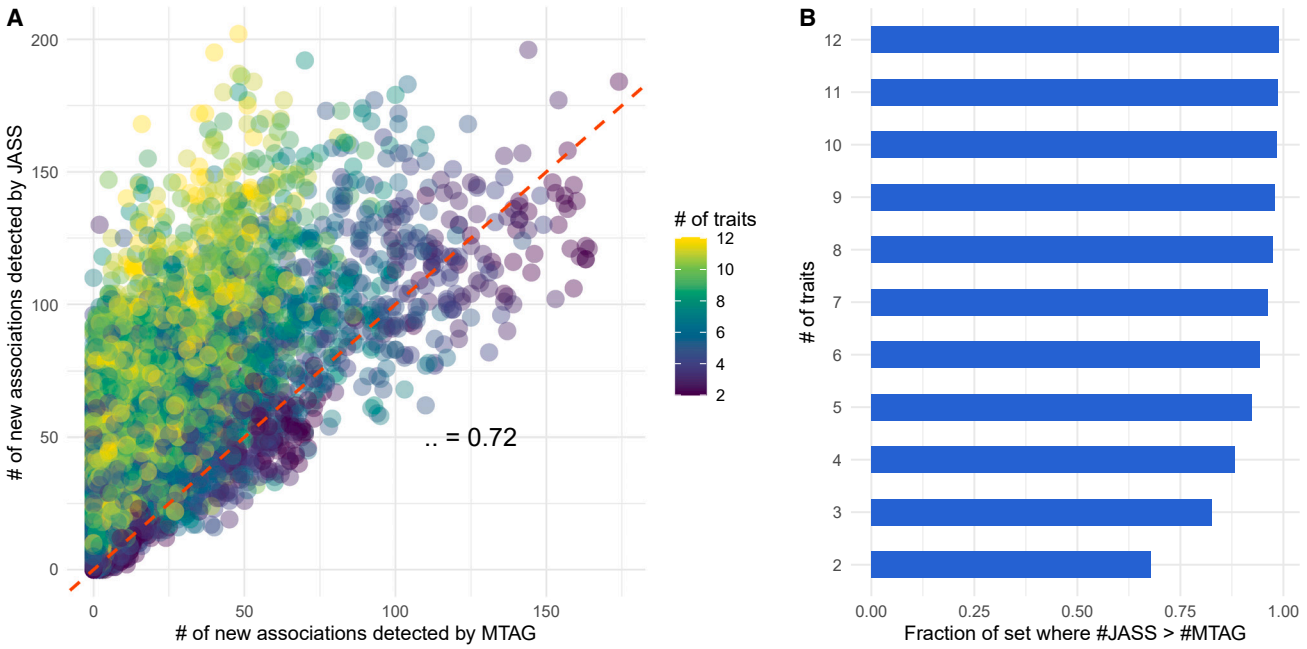

Figure 5. Comparison of MTAG with JASS

(A) The number of new association loci found by JASS with respect to the number of new association loci found MTAG across all the trait sets. Each dot represents a set of traits. Dot colors represent the number of traits in the set.

(B) Fraction of sets where the number of new association loci detected by JASS was superior to the number of associations detected by MTAG stratified by the number of traits in the set.

we fitted a logistic regression predicting if a locus would be associated with BMI in the larger GWAS by combining the number of sets where the locus was associated by JASS and the minimum p value across sets (Figure S21). This model reached an area under the curve of 0.75, an accuracy 0.74, a recall of 46%, and a specificity of 75% when applying a standard probability threshold of 0.5. Based on JASS results, we were able to infer a substantial number of loci associated in a GWAS with twice as many samples as the one used in the present study.

Discussion

This study investigated the genetic features associated with the statistical power of multi-trait GWAS. On average, the power increase relative to the univariate GWAS was substantial: JASS detected new association loci in 98% of 19,266 sets, with an average of 26 new association loci. This power increase appears to be highly associated with the genetic features of trait sets. More specifically, multi-trait gain tends to be higher for sets with (1) a moderate mean heritability (mean h^2_{GWAS}), (2) a smaller mean MES, (3) a larger mean polygenicity, (4) a larger genetic covariance across traits, and (5) a smaller distance between the residual and genetic covariances. We also found that selecting specific traits such as BMI, and more clinically heterogeneous sets, specifically for the omnibus test, can strongly outperform approaches that select clinically homogeneous sets. Finally, we investigated the predictive power of a multivariate linear model that could predict trait sets that most likely benefit from the multi-trait test (median Pearson's $r = 0.43$, $p < 1.6 \times 10^{-60}$), which can be used to astutely select traits to be tested jointly. Our findings provide an approach that can increase the identification of genetic associations using existing GWAS data, with relevance to traits in which genetic signal is scarce. We summarize in Figure S22 a guideline for selecting the best suited methods according to the feature of their data.

A part of our findings—the distance between the residual and genetic covariances negatively contributed to multi-trait gain—is opposite to the trend implied by a previous simulation study.[8] This discrepancy is likely due to constraints on genetic and residual covariances that were not considered in the previous simulation (Notes S4 and S5). With the constraints, the absolute value of the off-diagonal terms of the residual covariance matrix tends to become lower when the genetic and environmental correlations have opposite signs (i.e., a greater distance between the residual and genetic covariances) than when they have the same sign (i.e., a smaller distance). The configuration with the opposite signs can lead to an omnibus null having a larger overlap with the univariate null, which can result in a lower multi-trait gain (Notes S4). This new finding highlights the importance of considering multiple parameters in a real data analysis to notice their interplay and unravel it.

Selecting clinically homogeneous traits is the strategy most commonly used.[4,8,24,27–32] In our previous large-scale analysis,[8] a heterogeneous set yielded the largest number

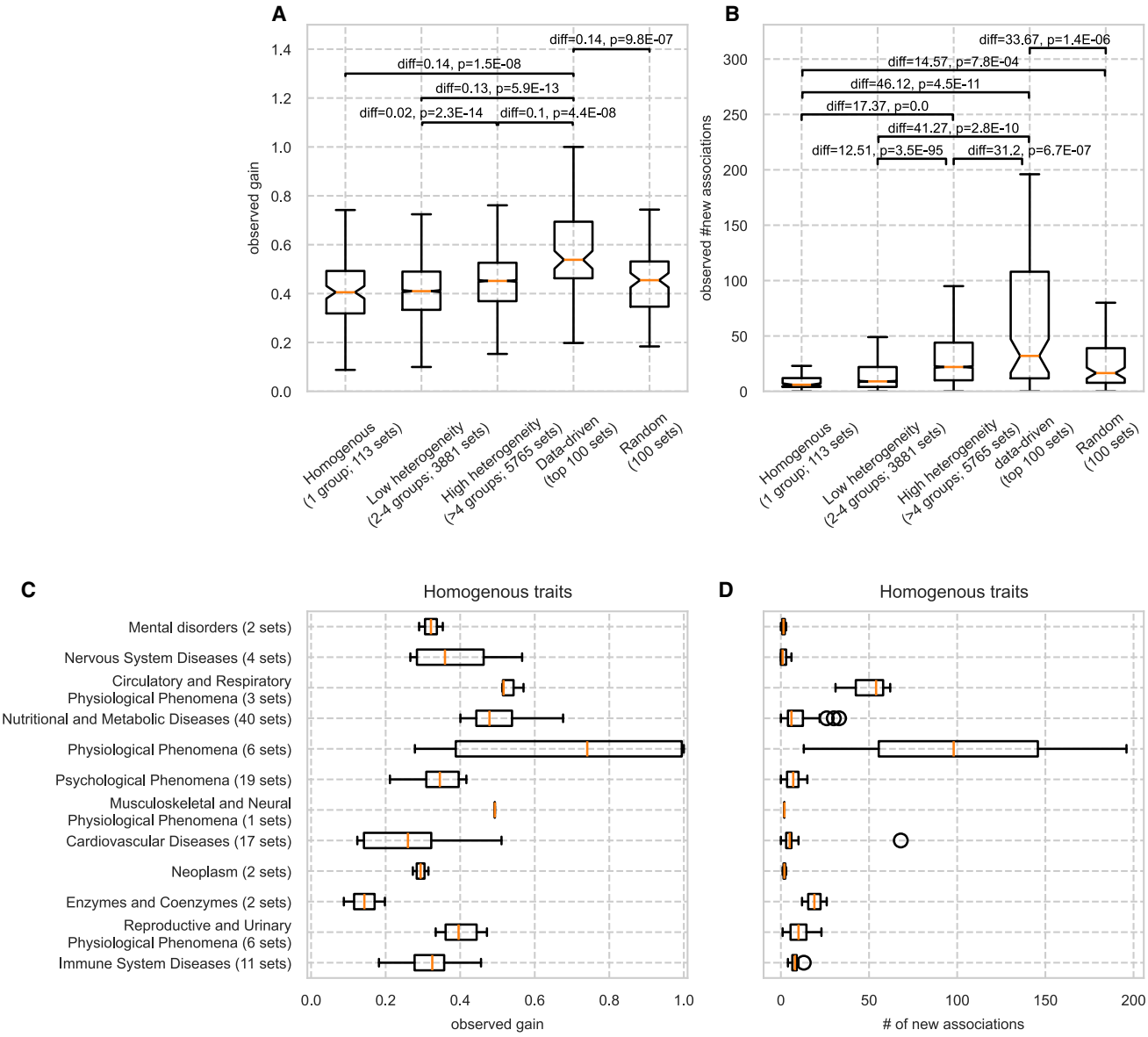

Figure 6. Comparison between clinical and data-driven trait sampling methods

(A and B) Distribution of the gain and the number of new association loci for trait sets selected by four trait selection strategies from the validation data. The *p* values are from the two-sided Welch's t test. Differences in mean values in each pair compared in the test (right-left categories in the order shown on the x axis) are also shown. Note we used a pair of a training data and a validation data to ensure independence for the test. The numbers under the labels on the x axis indicate the number of trait sets from each strategy. The observed JASS gain and the number of new association loci are shown on the y axis.

(C and D) The observed gain and number of new association loci detected for trait sets of the "homogeneous" category, visualized per clinical grouping.

of new association as compared with clinically homogeneous sets. In this work, we further showed that a data-driven strategy is expected to outperform other strategies based on clinical insights. Such sets might capture highly pleiotropic signals hard to detect using univariate GWASs and recommend that investigators compose a heterogeneous set or use our predictive model to build a set of traits.

We compared the results from the standard omnibus test (implemented in JASS) against MTAG, a popular multi-trait GWAS method.[33–36] For the data we used, the omnibus almost systematically outperformed MTAG, with an overall 3-fold increase in the number of loci detected. The difference was particularly striking in larger trait sets. MTAG is built on the hypothesis of homogeneous genetic correlation across genetic variants, and therefore is expected to have maximum power when this assumption is valid. Indeed, we observed that more homogeneous trait sets yielded larger MTAG gain than heterogeneous trait sets (Figures S20A and S20B). Yet, the omnibus generally performed equally well or better than MTAG even on the homogeneous trait sets (Figures S20C–S20F). In contrast, by construction, the omnibus test allows for substantial

heterogeneity, although at the cost of an increase in the degree of freedom. In line with previous work suggesting that genetic correlation might be fairly heterogeneous across the genome,[37] this cost seems to be outweighed by the additional flexibility in capturing heterogeneous multi-trait genetic patterns[8] (Notes S1 and S2). Thus, the omnibus test is recommended for the general identification of variants that impact phenotypes, while MTAG is suitable for identifying variants associated with specific traits (Figure S22).

Our study has some limitations. First, we focused on commonly measured genetic features, but other refined metrics could be used. These include effect size distribution as measured by the alpha parameters.[38] Second, we considered GWAS derived from common diseases and anthropometric traits. Future studies might explore performances using a wider variety of molecular traits, for which GWAS summary statistics are becoming increasingly available. Third, the estimation of the features might be also refined. Here, we used MiXeR[15,16] to estimate most features. However, we observed a dependency of MES and polygenicity on N_{eff}. Improving these metrics could improve the overall analysis and interpretations of the results. Fourth, we focused on European ancestry summary statistics. This decision was motivated by the availability of large GWAS and using one ancestry for LD; however, by doing so, we disposed of many traits with which we could have had a greater variety of genetic features, which might have improved the performance of the predictive model. This focus should not lead the reader to think that multi-trait GWAS is useful only on large sample studies of European ancestry. We recently updated the JASS pipeline to run a multi-ancestry multi-trait GWAS, which was able to detect 367 new association loci in non-European cohorts despite their modest sample sizes.[24] Future work might leverage non-European existing[39] and upcoming biobanks[40] to investigate the validity of our results for non-European ancestries.

In conclusion, this study provides a first overview of what to expect when applying multi-trait tests to a variety of data and how to maximize new discoveries. These insights can be leveraged to discover genetic variants associated with human complex traits and diseases missed by univariate analysis at no cost. Beyond mapping, JASS used on clinically heterogeneous trait sets might offer a way to understand a shared genetic etiology among unexpected traits[41] and contribute to better understanding of pleiotropy.

Data and code availability

Datasets analyzed during the study are available as listed below.

Curated GWAS summary statistics: https://jass.pasteur.fr/ and in Zenodo (https://doi.org/10.5281/zenodo.10356162).

MeSH Tree Structures: https://meshb-prev.nlm.nih.gov/search.

Codes used for the analyses during the study are available in Gitlab repositories as listed below.

Codes for the curation of GWAS summary statistics: https://gitlab.pasteur.fr/statistical-genetics/jass_suite_pipeline.

JASS package including the jass predict-gain command: https://gitlab.pasteur.fr/statistical-genetics/jass.

Codes for the analyses during the study: https://gitlab.pasteur.fr/statistical-genetics/multitrait_power_traitselection.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.xhgg.2024.100319.

Acknowledgments

This research was supported by the Agence Nationale pour la Recherche (GenCAST, ANR-20-CE36-0009). This work has been conducted as part of the INCEPTION program (Investissement d'Avenir grant ANR-16-CONV-0005). M.H.C. was supported by NHLBI (R01HL162813, R01HL153248, R01HL149861, and R01HL147148).

Author contributions

Conceptualization: H.J., H.A.; Methodology: Y.S., H.J., H.A., C.B., A.A., M.H.C.; Software: Y.S., H.J., H.M., B.B., P.L., R.T., C.N., L.T.; Validation: Y.S., H.J., H.A., M.H.C., E.B.; Formal analysis: Y.S., H.J.; Investigation: Y.S., H.J.; Resources: H.A.; Data curation: H.J., C.L., L.T.; Writing (original draft): Y.S., H.J., H.A.; Writing (review & editing): Y.S., R.V., C.B., A.A., M.H.C., E.B., H.A., H.J.; Visualization: Y.S., H.J., H.A.; Supervision: H.J., H.A.; Project administration: H.J., H.A., E.B.; Funding acquisition: H.J., H.A., E.B., M.H.C.

Declaration of interests

M.H.C. has received grant support from Bayer, unrelated to the current work.

Received: January 3, 2024
Accepted: June 11, 2024

Web resources

JASS https://jass.pasteur.fr/

MeSH Browser (for MeSH Tree Structures) https://meshb-prev.nlm.nih.gov/search

Online Mendelian Inheritance in Man https://omim.org/

References

1. Bhattacharjee, S., Rajaraman, P., Jacobs, K.B., Wheeler, W.A., Melin, B.S., Hartge, P., GliomaScan Consortium, Yeager, M., Chung, C.C., Chanock, S.J., and Chatterjee, N. (2012). A subset-based approach improves power and interpretation for the combined analysis of genetic association studies of heterogeneous traits. Am. J. Hum. Genet. *90*, 821–835. https://doi.org/10.1016/j.ajhg.2012.03.015.

2. Julienne, H., Lechat, P., Guillemot, V., Lasry, C., Yao, C., Araud, R., Laville, V., Vilhjalmsson, B., Ménager, H., and Aschard, H. (2020). JASS: command line and web interface for the joint analysis of GWAS results. NAR Genom. Bioinform. *2*, lqaa003. https://doi.org/10.1093/nargab/lqaa003.
3. Qi, G., and Chatterjee, N. (2018). Heritability informed power optimization (HIPO) leads to enhanced detection of genetic associations across multiple traits. PLoS Genet. *14*, e1007549. https://doi.org/10.1371/journal.pgen.1007549.
4. Turley, P., Walters, R.K., Maghzian, O., Okbay, A., Lee, J.J., Fontana, M.A., Nguyen-Viet, T.A., Wedow, R., Zacher, M., Furlotte, N.A., et al. (2018). Multi-trait analysis of genome-wide association summary statistics using MTAG. Nat. Genet. *50*, 229–237. https://doi.org/10.1038/s41588-017-0009-4.
5. Wang, M., Cao, X., Zhang, S., and Sha, Q. (2023). A Clustering Linear Combination (CLC) Method for Multiple Phenotype Association Studies Based on GWAS Summary Statistics. Sci. Rep. *13*, 3389. https://doi.org/10.1038/s41598-023-30415-3.
6. Zhu, X., Feng, T., Tayo, B.O., Liang, J., Young, J.H., Franceschini, N., Smith, J.A., Yanek, L.R., Sun, Y.V., Edwards, T.L., et al. (2015). Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. Am. J. Hum. Genet. *96*, 21–36. https://doi.org/10.1016/j.ajhg.2014.11.011.
7. Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., and Yang, J. (2017). 10 years of GWAS discovery: biology, function, and translation. Am. J. Hum. Genet. *101*, 5–22. https://doi.org/10.1016/j.ajhg.2017.06.005.
8. Julienne, H., Laville, V., McCaw, Z.R., He, Z., Guillemot, V., Lasry, C., Ziyatdinov, A., Nerin, C., Vaysse, A., Lechat, P., et al. (2021). Multitrait GWAS to connect disease variants and biological mechanisms. PLoS Genet. *17*, e1009713. https://doi.org/10.1371/journal.pgen.1009713.
9. Porter, H.F., and O'Reilly, P.F. (2017). Multivariate simulation framework reveals performance of multi-trait GWAS methods. Sci. Rep. *7*, 38837. https://doi.org/10.1038/srep38837.
10. Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. *47*, D1005–D1012. https://doi.org/10.1093/nar/gky1120.
11. 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. Nature *526*, 68–74. https://doi.org/10.1038/nature15393.
12. Julienne, H., Shi, H., Pasaniuc, B., and Aschard, H. (2019). RAISS: robust and accurate imputation from summary statistics. Bioinformatics *35*, 4837–4839. https://doi.org/10.1093/bioinformatics/btz466.
13. Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.-R., ReproGen Consortium; Psychiatric Genomics Consortium; and Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3, and Duncan, L., et al. (2015). An atlas of genetic correlations across human diseases and traits. Nat. Genet. *47*, 1236–1241. https://doi.org/10.1038/ng.3406.
14. Berisa, T., and Pickrell, J.K. (2016). Approximately independent linkage disequilibrium blocks in human populations. Bioinformatics *32*, 283–285. https://doi.org/10.1093/bioinformatics/btv546.
15. Frei, O., Holland, D., Smeland, O.B., Shadrin, A.A., Fan, C.C., Maeland, S., O'Connell, K.S., Wang, Y., Djurovic, S., Thompson, W.K., et al. (2019). Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation. Nat. Commun. *10*, 2417. https://doi.org/10.1038/s41467-019-10310-0.
16. Holland, D., Frei, O., Desikan, R., Fan, C.-C., Shadrin, A.A., Smeland, O.B., Sundar, V.S., Thompson, P., Andreassen, O.A., and Dale, A.M. (2020). Beyond SNP heritability: Polygenicity and discoverability of phenotypes estimated with a univariate Gaussian mixture model. PLoS Genet. *16*, e1008612. https://doi.org/10.1371/journal.pgen.1008612.
17. Harris, C.R., Millman, K.J., Van Der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., et al. (2020). Array programming with NumPy. Nature *585*, 357–362. https://doi.org/10.1038/s41586-020-2649-2.
18. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., and Dubourg, V. (2011). Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.
19. Seabold, S., and Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python, pp. 10–25080.
20. Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B.M., and Gräler, B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ *6*, e5518. https://doi.org/10.7717/peerj.5518.
21. Balabin, R.M., and Smirnov, S.V. (2012). Interpolation and extrapolation problems of multivariate regression in analytical chemistry: benchmarking the robustness on near-infrared (NIR) spectroscopy data. Analyst *137*, 1604–1610. https://doi.org/10.1039/C2AN15972D.
22. National Institutes of Health (2010). National Library of Medicine–Medical Subject Headings.
23. Yengo, L., Sidorenko, J., Kemper, K.E., Zheng, Z., Wood, A.R., Weedon, M.N., Frayling, T.M., Hirschhorn, J., Yang, J., Visscher, P.M.; and GIANT Consortium (2018). Meta-analysis of genome-wide association studies for height and body mass index in∼ 700000 individuals of European ancestry. Hum. Mol. Genet. *27*, 3641–3649. https://doi.org/10.1093/hmg/ddy271.
24. Troubat, L., Fettahoglu, D., Aschard, H., and Julienne, H. (2023). Multi-trait GWAS for diverse ancestries: Mapping the knowledge gap. Preprint at bioRxiv. https://doi.org/10.1101/2023.06.23.546248.
25. Simon, J.J., Skunde, M., Hamze Sinno, M., Brockmeyer, T., Herpertz, S.C., Bendszus, M., Herzog, W., and Friederich, H.-C. (2014). Impaired cross-talk between mesolimbic food reward processing and metabolic signaling predicts body mass index. Front. Behav. Neurosci. *8*, 359. https://doi.org/10.3389/fnbeh.2014.00359.
26. Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R., Powell, C., Vedantam, S., Buchkovich, M.L., Yang, J., et al. (2015). Genetic studies of body mass index yield new insights for obesity biology. Nature *518*, 197–206. https://doi.org/10.1038/nature14177.
27. Fan, C.C., Loughnan, R., Makowski, C., Pecheva, D., Chen, C.-H., Hagler, D.J., Thompson, W.K., Parker, N., van der Meer, D., Frei, O., et al. (2022). Multivariate genome-wide association study on tissue-sensitive diffusion metrics highlights pathways that shape the human brain. Nat. Commun. *13*, 2423. https://doi.org/10.1038/s41467-022-30110-3.

28. Gallois, A., Mefford, J., Ko, A., Vaysse, A., Julienne, H., Ala-Korpela, M., Laakso, M., Zaitlen, N., Pajukanta, P., and Aschard, H. (2019). A comprehensive study of metabolite genetics reveals strong pleiotropy and heterogeneity across time and context. Nat. Commun. *10*, 4788. https://doi.org/10.1038/s41467-019-12703-7.
29. Khunsriraksakul, C., Li, Q., Markus, H., Patrick, M.T., Sauteraud, R., McGuire, D., Wang, X., Wang, C., Wang, L., Chen, S., et al. (2023). Multi-ancestry and multi-trait genome-wide association meta-analyses inform clinical risk prediction for systemic lupus erythematosus. Nat. Commun. *14*, 668. https://doi.org/10.1038/s41467-023-36306-5.
30. Levin, M.G., Tsao, N.L., Singhal, P., Liu, C., Vy, H.M.T., Paranjpe, I., Backman, J.D., Bellomo, T.R., Bone, W.P., Biddinger, K.J., et al. (2022). Genome-wide association and multi-trait analyses characterize the common genetic architecture of heart failure. Nat. Commun. *13*, 6914. https://doi.org/10.1038/s41467-022-34216-6.
31. van der Meer, D., Frei, O., Kaufmann, T., Shadrin, A.A., Devor, A., Smeland, O.B., Thompson, W.K., Fan, C.C., Holland, D., Westlye, L.T., et al. (2020). Understanding the genetic determinants of the brain with MOSTest. Nat. Commun. *11*, 3512. https://doi.org/10.1038/s41467-020-17368-1.
32. Xiong, Z., Gao, X., Chen, Y., Feng, Z., Pan, S., Lu, H., Uitterlinden, A.G., Nijsten, T., Ikram, A., Rivadeneira, F., et al. (2022). Combining genome-wide association studies highlight novel loci involved in human facial variation. Nat. Commun. *13*, 7832. https://doi.org/10.1038/s41467-022-35328-9.
33. Rahmioglu, N., Mortlock, S., Ghiasi, M., Møller, P.L., Stefansdottir, L., Galarneau, G., Turman, C., Danning, R., Law, M.H., Sapkota, Y., et al. (2023). The genetic basis of endometriosis and comorbidity with other pain and inflammatory conditions. Nat. Genet. *55*, 423–436. https://doi.org/10.1038/s41588-023-01323-z.
34. Sveinbjornsson, G., Ulfarsson, M.O., Thorolfsdottir, R.B., Jonsson, B.A., Einarsson, E., Gunnlaugsson, G., Rognvaldsson, S., Arnar, D.O., Baldvinsson, M., Bjarnason, R.G., et al. (2022). Multiomics study of nonalcoholic fatty liver disease. Nat. Genet. *54*, 1652–1663. https://doi.org/10.1038/s41588-022-01199-5.
35. Wang, Y., Tsuo, K., Kanai, M., Neale, B.M., and Martin, A.R. (2022). Challenges and opportunities for developing more generalizable polygenic risk scores. Annu. Rev. Biomed. Data Sci. *5*, 293–320. https://doi.org/10.1146/annurev-biodatasci-111721-074830.
36. Wang, Z., Emmerich, A., Pillon, N.J., Moore, T., Hemerich, D., Cornelis, M.C., Mazzaferro, E., Broos, S., Ahluwalia, T.S., Bartz, T.M., et al. (2022). Genome-wide association analyses of physical activity and sedentary behavior provide insights into underlying mechanisms and roles in disease prevention. Nat. Genet. *54*, 1332–1344. https://doi.org/10.1038/s41588-022-01165-1.
37. Werme, J., van der Sluis, S., Posthuma, D., and de Leeuw, C.A. (2022). An integrated framework for local genetic correlation analysis. Nat. Genet. *54*, 274–282. https://doi.org/10.1038/s41588-022-01017-y.
38. Miao, L., Jiang, L., Tang, B., Sham, P.C., and Li, M. (2023). Dissecting the high-resolution genetic architecture of complex phenotypes by accurately estimating gene-based conditional heritability. Am. J. Hum. Genet. *110*, 1534–1548. https://doi.org/10.1016/j.ajhg.2023.08.006.
39. Ishigaki, K., Akiyama, M., Kanai, M., Takahashi, A., Kawakami, E., Sugishita, H., Sakaue, S., Matoba, N., Low, S.-K., Okada, Y., et al. (2020). Large-scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases. Nat. Genet. *52*, 669–679. https://doi.org/10.1038/s41588-020-0640-3.
40. (2019). The "All of Us" Research Program. N. Engl. J. Med. *381*, 668–676. https://doi.org/10.1056/NEJMsr1809937.
41. Hackinger, S., and Zeggini, E. (2017). Statistical methods to detect pleiotropy in human complex traits. Open Biol. *7*, 170125. https://doi.org/10.1098/rsob.170125.