

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

2 discoverability

3 Yuka Suzuki^{1*}, Hervé Ménager², Bryan Brancotte², Raphaël Vernet³, Cyril Nerin¹, Christophe Boetto¹,
4 Antoine Auvergne¹, Christophe Linhard³, Rachel Torchet², Pierre Lechat², Lucie Troubat¹, Michael H.
5 Cho^{4,5}, Emmanuelle Bouzigon³, Hugues Aschard^{1*}, Hanna Julienne^{1,2*}

6 *Corresponding authors : yuka.suzuki@pasteur.fr, hugues.aschard@pasteur.fr, hanna.julienne@pasteur.fr

7 ¹*Institut Pasteur, Université Paris Cité, Department of Computational Biology, Paris, 75015 France*

8 ²*Institut Pasteur, Université Paris Cité, Bioinformatics of Biostatistics Hub, F-75015 Paris, France*

9 ³*Université Paris Cité, Institut National de la Santé et de la Recherche Médicale (INSERM), UMR-1124, Group of
10 Genomic Epidemiology of Multifactorial Diseases, Paris, France*

11 ⁴*Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, 181
12 Longwood Ave, Boston, MA, 02115, USA*

13 ⁵*Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Harvard Medical School,
14 Boston, MA, USA*

15 Abstract

16 Since the first Genome-Wide Association Studies (GWAS), thousands of variant-trait associations
17 have been discovered. However, the sample size required to detect additional variants using standard
18 univariate association screening is increasingly prohibitive. Multi-trait GWAS offers a relevant
19 alternative: it can improve statistical power and lead to new insights about gene function and the joint
20 genetic architecture of human phenotypes. Although many methodological hurdles of multi-trait
21 testing have been discussed, the strategy to select trait, among overwhelming possibilities, has been
22 overlooked. In this study, we conducted extensive multi-trait tests using JASS (Joint Analysis of
23 Summary Statistics) and assessed which genetic features of the analysed sets were associated with an

24 increased detection of variants as compared to univariate screening. Our analyses identified multiple
25 factors associated with the gain in the association detection in multi-trait tests. Together, these factors
26 of the analysed sets are predictive of the gain of the multi-trait test (Pearson's ρ equal to 0.43 between
27 the observed and predicted gain, $P < 1.6 \times 10^{-60}$). Applying an alternative multi-trait approach (MTAG,
28 multi-trait analysis of GWAS), we found that in most scenarios but particularly those with larger
29 numbers of traits, JASS outperformed MTAG. Finally, we benchmark several strategies to select set of
30 traits including the prevalent strategy of selecting clinically similar traits, which systematically
31 underperformed selecting clinically heterogenous traits or selecting sets that issued from our data-
32 driven models. This work provides a unique picture of the determinant of multi-trait GWAS statistical
33 power and outline practical strategies for multi-trait testing.

34

35 **Introduction**

36 Despite a continuous increase of the sample size of Genome-Wide Association Studies (GWAS),
37 many genetic variants underlying human complex traits and diseases remain undetected. To increase
38 statistical power and detection of associations at low cost, investigators have developed various multi-
39 trait approaches based on GWAS summary statistics^{1–6}. Although other factors might also be involved,
40 few studies investigated how the choice of phenotypes impacts the gain in the power of multi-trait
41 approaches. In the standard univariate GWAS, statistical power mostly depends on minor allele
42 frequency, sample size, the size of genetic effect, and polygenicity (the number of causal variants)⁷. In
43 the multi-trait GWAS, it additionally depends on complex characteristics of the set of traits: their
44 shared aetiology, their genetic correlation, and the number of traits in the set. As previously described,
45 the increased power of multi-trait test partly comes from adjusting for the correlation across GWAS
46 studies due to sample overlaps and genetic relationships across phenotypes^{4,8,9}. Previous works
47 explored this question using simulated data^{8,9}. However, simulations are limited in their scope, and
48 more studies are needed to better characterise scenarios increasing the gain of multi-trait tests.

49 Here we empirically examined how trait characteristics impact the statistical power of multi-trait
50 GWAS. We performed our analyses on 72 curated GWAS summary statistics and analysed the impact
51 of 11 genetic features describing both individual and collective characteristics of sets of GWAS. Among
52 available multi-trait methods, we used a standard k -degree of freedom joint test (omnibus test) of k
53 GWAS implemented in JASS² as our primary analysis. The JASS package and its associated tools solve
54 all practical issues commonly encountered in multi-trait GWAS analyses, including missing data and
55 computational efficiency, allowing for a large-scale power analysis on real data. We inspected the
56 association between these genetic features and the gain in the association detection in multi-trait
57 GWAS over univariate GWAS. We then assessed how well these genetic features predict the gain. We
58 further compared JASS and MTAG, an alternative multi-trait approach, in terms of the impact of the
59 trait selection strategy on the gain of multi-trait GWAS.

60

61 **Results**

62 **Study Overview**

63 We conducted a series of analyses to identify features that influence the gain in association
64 detection of multi-trait GWAS relative to univariate GWAS. The key principles and main steps of the
65 study are depicted in **Figure 1**. We used 72 curated GWAS summary statistics spanning a range of
66 clinical domains (**Table S1**) and considered 11 features describing the univariate and multivariate
67 genetic architecture of the phenotypes. Five of them characterise single GWAS: mean genetic effect
68 size (MES), polygenicity, effective sample size (N_{eff}), linear additive heritability of common variants
69 (h_{GWAS}^2), the proportion of uncaptured linear additive heritability of common variants (% h_u^2). The
70 remaining six features characterise sets of GWAS. It includes the number of GWAS analysed jointly,
71 and five metrics related to the genetic (Σ_g , **Table S2** and **Fig. S1**) and residual (Σ_r , **Table S3** and **Fig.**
72 **S2**) covariance matrices, where the latter represents the covariance between the Z-score statistics of
73 two GWAS due to sample overlap and phenotypic correlation. Those five features are the mean of the
74 off-diagonal terms denoted as $\bar{\Sigma}_g$ and $\bar{\Sigma}_r$; the conditional numbers of Σ_g and Σ_r denoted as κ_g and
75 κ_r , which we used as a measure of multi-collinearity; and the average difference between the two
76 matrices Δ_Σ . Δ_Σ is expected to be a key driver of the power of multi-trait test^{8,9} (**Supplementary note**
77 **1, 2, Fig. S3**).

78 Given the 72 traits, there are over 4.7×10^{21} possible sets of two to 72 traits. In this study, we used
79 19,266 unique random sets of two to 12 GWAS drawn using a stratified sampling conditional on mean
80 effect size, heritability and the number of traits to maximize the range of the genetic features studied
81 (see **Material and methods, Supplementary note 3**). For each set, we derived the average of the single
82 GWAS feature (polygenicity, MES, N_{eff} , h_{GWAS}^2 , and % h_u^2) and the six set features ($\bar{\Sigma}_g$, $\bar{\Sigma}_r$, κ_g , κ_r , Δ_Σ ,
83 and the number of GWAS selected). Note that we used the log of polygenicity, MES, $\bar{\Sigma}_g$, $\bar{\Sigma}_r$, κ_g , κ_r ,

84 and Δ_{Σ} when investigating their association with other variables due to their skewed distributions (**Fig.**
85 **S4**). Then, we compared multi-trait and univariate GWAS results based on the minimum *P*-value of the
86 multi-trait test across all variants within LD-independent loci (P_{JASS}) and the minimum *P*-value of
87 univariate tests across all variants and all GWAS within the same loci (P_{uni}). We used two metrics: i)
88 the proportion of significant loci found associated at genome-wide level ($P < 5 \times 10^{-8}$) by JASS and
89 missed by the univariate GWAS, and ii) the fraction of loci where P_{JASS} was smaller than P_{uni}
90 ($f_{JASS < univ}$).

91 We first describe the distribution of the genetic features and their correlation. Second, we applied
92 JASS to each set and compared multi-trait and univariate GWAS association results (**Fig. 1B**). Third, we
93 build a predictive model of the association gain using a five-fold cross-validation approach (**Fig. 1C**).
94 For each cross-validation, we pulled a training set of 1,980 out of the 19,266 unique random sets and
95 conducted a regression analysis to estimate the contribution of a subset of genetic features of trait
96 sets using a joint model of the features. We measured the correlation between observed and
97 predicted gain in an independent validation dataset. Fourth, we compared the association gain from
98 JASS against MTAG⁴, a popular multi-trait approach that leverages genetic correlation across traits to
99 boost statistical power. Finally, we compare common trait selection strategies—e.g. choosing clinically
100 homogenous or heterogeneous traits—in their impacts on the association gain to evaluate our model
101 prediction and provide a practical strategy in multi-trait test (**Fig. 1D**).

102

103 ***Distribution of the genetic features across GWAS trait sets***

104 The individual genetic features of the 72 studied traits were distributed as follows. The sample sizes
105 ranged from 5,318 to 697,828 with a median of 85,559. Heritability (h_{GWAS}^2) ranged from 1% to 48%
106 with a median at 10% (**Fig. 2A**) and was consistent across the software used for its estimation (**Fig. S5**).
107 Polygenicity and MES were highly variable: Polygenicity (i.e. the estimated number of causal variants)
108 ranged from 6.9 (Estimated Glomerular Filtration Rate from Cystatin C) to 570,102 (variability sleep

109 duration) with a median of 1,110, and MES ranged from 4.48×10^{-8} (variability sleep duration) to $3.3 \times$
110 10^{-3} (Fasting proinsulin) with a median of 1.0×10^{-4} . The distribution of the average these parameters
111 across the 19,266 random sets is presented in **Figure 2B** along with the GWAS set features. The latter
112 metrics ranged (in 25-75 percentiles) as -2.04 to -1.79 for $\log_{10} \bar{\Sigma}_g$, -1.96 to -1.44 for $\log_{10} \bar{\Sigma}_r$, 0.36 to
113 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
114 in $\log_{10} \Delta_\Sigma$ was limited (given the theoretical upper bound of $\log_{10} \Delta_\Sigma$ is $\log_{10} 2 \cong 0.3$).

115 We measured the correlation across features at the trait and set levels (**Fig. 2C, 2D**). The
116 proportion of uncaptured linear additive heritability of common variants ($\%h_u^2$) was strongly
117 associated with \log_{10} polygenicity ($\rho = 0.76, P = 8.8 \times 10^{-15}$) and with \log_{10} MES ($\rho = -0.8, P = 3.5 \times$
118 10^{-17}) (**Fig. 2C, S6**), in agreement with previous reports showing that univariate GWAS performs better
119 for traits with a larger MES and a smaller polygenicity⁷. As expected, the means of individual trait
120 feature were correlated in the same way across sets as across traits (e.g. mean $\%h_u^2$ and mean
121 polygenicity are highly correlated as are $\%h_u^2$ and polygenicity, **Fig. 2D**). Metrics of genetic and
122 covariance matrices were positively correlated to each other, i.e. $\log_{10} \bar{\Sigma}_g$ with $\log_{10} \bar{\Sigma}_r$ and $\log_{10} \Delta_\Sigma$,
123 $\log_{10} \kappa_g$ with $\log_{10} \kappa_r$, and $\log_{10} \bar{\Sigma}_r$ with $\log_{10} \kappa_r$ and $\log_{10} \kappa_g$, which further correlated with the
124 number of traits.

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

126 We applied JASS to all 19,266 sets and quantified the gain in association detection of the multi-
127 trait against univariate test. On average, in a set, JASS detected 26 new loci while 285 were previously
128 associated with the univariate tests, i.e. a 1.1-fold increase in the total number of loci detected (**Fig.**
129 **S7A**). JASS gain was maximal for sets with a relatively low number (< 300 loci) of previously detected
130 loci (**Fig. S7B**). JASS detected at least one new association in 98% of the trait sets (18,787 sets, **Fig.**
131 **S7C**), and 508,829 new associations in total (note that these can be overlapping loci). These numbers
132 are obtained when applying JASS on variants with beta coefficients available for all traits in the set.
133 When analysing all variants including those with missing values as allowed by JASS (see **Material and**

134 **methods**), about 1.4 times more new associations were detected (693,382 new associations in total
135 by JASS including variants with missing values, **Fig. S7D**).

136 We assessed the marginal relationship between genetic features and the number of loci detected
137 by the univariate and multi-trait GWAS tests, and the association gain of multi-trait test. **Figure 3**
138 presents the correlation between each feature and the number of univariate and multi-trait GWAS
139 associated loci. As expected, the number of univariate associated loci was positively correlated with
140 the number of traits, mean N_{eff} , the mean h^2_{GWAS} and mean MES, and negatively correlated with mean
141 uncaptured linear additive heritability (mean $\%h_u^2$). Multi-trait GWAS gain over univariate GWAS was
142 positively associated with mean polygenicity and mean $\%h_u^2$ while negatively associated with mean
143 MES. Mean h^2_{GWAS} show opposite effect, being slightly positively associated with gain, but negatively
144 associated with gain. Overall, this suggests that the multi-trait test can be highly complementary to
145 the univariate test, performing better in situations where the univariate tests display low power. We
146 noted in a recent study¹⁰ that a high multicollinearity of the matrix underlying the null hypothesis (Σ_r)
147 can lead to a lack of robustness of the omnibus test¹⁰. We checked how the condition number κ_r was
148 related to JASS gain (**Fig. 3B**, **Fig. S8**). The condition number stayed in a reasonable range for 99% of
149 sets (min=1, max=11). The rest (1%) of sets were flagged for having singular residual matrices.
150 Furthermore, note that the trait sets contained overlapping traits and are therefore not fully
151 independent from each other. To robustly assess the impacts of genetic features on the multi-trait
152 gain by addressing this issue, below we conducted regression analyses with a cross validation scheme
153 establishing independence between training and validation data (**Fig. S9**).

154 ***Predicting multi-trait test gain from genetic features***

155 We investigated the predictive power of the association gain ($f_{JASS<univ}$) from a joint modelling of
156 the genetic features (**Fig. 4**). Note that some features are almost linear combinations of other ones
157 (e.g. Δ_Σ is proportional to the difference between $\bar{\Sigma}_g$ and $\bar{\Sigma}_r$). To avoid extreme collinearity and
158 ensure parsimony, we selected six moderately correlated features: the number of traits, $\log_{10} \Delta_\Sigma$,

159 $\log_{10} \bar{\Sigma}_g$, mean N_{eff} , mean h^2_{GWAS} , and mean $\%h_u^2$ (see **Material and methods**). We favoured mean
160 $\%h_u^2$ over mean MES and mean polygenicity since mean $\%h_u^2$ captures both $\log_{10}(\text{MES})$ and
161 $\log_{10}(\text{polygenicity})$ (**Fig. 2**). We assessed performances using a five-fold cross validation (CV). For each
162 CV the model parameters were derived on a training data, and the prediction accuracy was derived in
163 an independent data (**Material and methods, Fig. S9**). All six features were highly associated to the
164 multi-trait gain (**Table 1**). Mean $\%h_u^2$ was overall positively associated to multi-trait gain, whereas
165 mean h^2_{GWAS} was negatively associated. Overall, the association gain of multi-trait test, more
166 specifically the omnibus test, seems driven by genetic correlation, polygenicity, and the number of
167 traits. The predicted gain was significantly correlated with the observed gain on validation data
168 (median Pearson $\rho=0.43$, $P < 1.6 \times 10^{-60}$, **Fig. S10**).

169 We conducted a series of sensitivity analyses to explore further performances. First, to interpret
170 this model behaviour from the perspective of polygenicity and MES, we fitted two models replacing
171 mean $\%h_u^2$ with mean $\log_{10}(\text{polygenicity})$ and mean $\log_{10}(\text{MES})$ in turn (**Fig. S11**). In these models,
172 mean $\log_{10}(\text{polygenicity})$ had a positive contribution to the gain, whereas mean $\log_{10}(\text{MES})$ had a
173 negative contribution. In other words, multi-trait tests likely detect new associations in settings where
174 univariate tests perform poorly, confirming the correlation analysis (**Fig. 3**). Additionally, the model
175 suggested the number of traits and $\log_{10} \bar{\Sigma}_g$ further enhance multi-trait test gain. Contrary to previous
176 observation on simulation studies, $\log_{10} \Delta_\Sigma$ likely diminishes the multi-trait test gain (see
177 **Supplementary note 4** for a hypothetical explanation). Second, we also considered two nonlinear
178 models for comparison purposes, support vector regression and random forest regression. As showed
179 in **Figure S12**, these two models appear to outperform the multivariate linear model on the training
180 datasets. However, they performed similarly or worse on the test dataset, suggesting a strong
181 overfitting in the training data. Overall, we did not find any benefit in using these more complex
182 models.

183 **Comparison of JASS versus MTAG**

184 We repeated the association screening and the prediction analysis using MTAG (Multi-Trait
185 Analysis of GWAS), a popular multi-trait approach leveraging genetic correlation among closely related
186 traits to inform GWAS screening. Regarding the association screening, MTAG detected 153,061 new
187 association regions in total across the sets (30% of the number of new associations detected by JASS).
188 On average, MTAG detected eight new association loci per set and at least one new association locus
189 on 63.3% of sets (12,195 out of 19,266) compared to 98% of sets for JASS. In 93% of all the trait sets,
190 MTAG detected fewer associations than JASS did (**Fig. 5A**). The performance difference further
191 increased when applying JASS also on variants with missing values (in 96% of trait sets MTAG detected
192 fewer associations). Despite these discrepancies in the number of association loci, the number of new
193 association loci in MTAG and JASS were strongly correlated (Pearson $\rho = 0.72$, $P < 2.2 \times 10^{-308}$), and as
194 well as the P -values of MTAG and JASS for the same loci (Pearson $\rho = 0.75$, $P < 2.2 \times 10^{-308}$). This
195 concordance suggests common determinants for statistical power between the two methods. We next
196 fitted a multivariate linear model to predict MTAG gain from the same six genetic features and training
197 data used for JASS (**Fig. S13**). The most notable difference from JASS was that the number of traits in
198 the set was, consistently across cross-validation folds, negatively associated with the gain of MTAG.
199 Indeed, JASS particularly outperformed MTAG on larger set of traits (**Fig. 5B**).

200 To test whether the pronounced gain of JASS compared with MTAG on large sets was mostly due
201 to multiple testing correction—as MTAG performs one test by trait present in the set of traits, multiple
202 testing correction is applied to MTAG P -values (**Material and methods**)—we compared the number of
203 associations detected by JASS and by MTAG without correction (**Fig. S14**). In the absence of multiple
204 testing correction, MTAG type 1 error expectedly increased with the number of traits in the set (e.g.
205 genomic inflation factor λ was 1.3, 1.66, 1.72, and 1.77 for four cases with 2, 5, 9, and 12 traits).
206 Despite the increased P -value inflation for large sets, MTAG detected fewer associations overall (**Fig.**
207 **S14A**) than JASS on most of the sets (in 71% and 83% of sets when excluding and including variants
208 with partly unavailable summary statistics, respectively). The performance of JASS and MTAG were

209 most similar in small trait sets (**Fig. S14B**), while JASS was particularly advantageous when analysing
210 large trait set especially when allowing for variants with summary statistics partly unavailable.

211

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

213 A common and seemingly sound strategy when conducting multi-trait analyse is to use closely
214 related traits. This choice is partly driven by investigators' interest in delineating the shared genetic
215 aetiology between a disease and closely related phenotypes. This might also arise from the intuitive
216 idea that closely related phenotypes share a fair amount of genetic aetiology that the multi-test could
217 leverage. However, its impact on statistical power has not been evaluated. To advise investigators on
218 the best strategy to compose sets, we compared multi-trait gain and the number of new association
219 loci obtained using this clinically-driven strategy (referred as "*homogenous*", see **Material and**
220 **methods**) to three alternatives: i) including GWAS from two to four clinical groups (noted "*low*
221 *heterogeneity*"), ii) including GWAS from five clinical groups or more (noted "*high heterogeneity*"),
222 and iii) a data-driven approach based on the linear regression predicted gain (**Material and methods**).
223 The data-driven strategy had a higher gain and larger number of new association loci by JASS
224 compared to the other strategies (**Fig. 6A and 6B**). The gain and the number of new association loci
225 increased systematically with the clinical heterogeneity of the traits, and the increases were
226 statistically significant for most pairs of trait selection strategies compared, especially when comparing
227 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
228 number of new association loci). The average number of new association loci detected in validation
229 data equalled 14, 20, 32, and 61 for homogeneous, low heterogeneity, high heterogeneity, and data-
230 driven sets, respectively.

231 We further decomposed the gain of the homogenous sets by clinical groups (**Fig. 6C, 6D**). Two
232 groups consistently yielded more new association loci than others, namely: 'Circulatory and
233 Respiratory Physiological Phenomena' and 'Physiological Phenomena.' The first group was composed

234 of spirometry traits and asthma, characterised by a large sample size, substantial genetic correlation,
235 and moderate heritability: a favourable setting for JASS. The second group contained anthropometric
236 traits (Height, BMI, Hip circumference, waist circumference, and Waist hip ratio). We next investigated
237 whether multi-trait gain was associated with specific traits and derived the fold enrichment for traits
238 between the top 10% yielding sets (corresponding to >169 new associated loci by JASS) and the least
239 10% yielding sets. BMI and hip circumference were amongst the top traits (**Fig. S15**). Notably, we
240 found that high yielding trait sets (i.e., those that yielded 169 or more new association regions by JASS)
241 all contained BMI. In contrast, out of the remaining trait sets, only 9% of the trait sets contained BMI.
242 Other traits enriched in top sets included spirometry and to a lesser extend mental disorders, arterial
243 pressure, and sleep patterns. BMI genetics is increasingly recognized to be a complex entanglement
244 of metabolic and behavioural factors¹¹, and suggests that complex traits that reflect multiple biologic
245 processes may benefit more from multi-trait analysis.

246 We additionally compared the multi-trait gain and the number of new association loci by MTAG
247 across the four groups of trait sets. The results were rather opposite to what we observed above with
248 JASS: MTAG gain increased as the trait sets became more homogenous (**Fig. S17A**, $P < 1.4 \times 10^{-7}$), and
249 the number of new association loci was greater for trait sets of low heterogeneity than high
250 heterogeneity (**Fig. S17B**, $P = 2.7 \times 10^{-6}$). We then tested whether MTAG outperforms JASS on
251 homogenous trait sets (**Fig. S17C,D**) and found that there was little difference between the two
252 ($P=0.084$ for gain, $P=0.063$ for the number of new association loci). We further tested whether MTAG
253 outperforms JASS on homogenous trait sets of certain clinical groups (**Fig. S17E,F**). JASS significantly
254 outperformed on ‘Immune System Diseases’ and ‘Psychological Phenomena’ (‘Immune System
255 Diseases’: $P = 8.2 \times 10^{-4}$ for gain and $P = 6.7 \times 10^{-7}$ for the number of new association loci, ‘Psychological
256 Phenomena’: $P = 2.5 \times 10^{-5}$ for the number of new association loci). In contrast, MTAG outperformed
257 on ‘Musculoskeletal and Neural Physiological Phenomena’ (which contains only one trait set).

258 While JASS type 1 error in JASS is properly calibrated, this large number of new associations could
259 be spurious, and irrelevant to biology. To ensure that new associations detected by JASS are relevant,

260 we focussed on BMI and tested if we could predict novel associations observed in a larger study
261 (sample size= 683,365 ¹²) from multi-trait GWAS applied on the BMI study in the present analysis
262 (sample size= 339,224 ¹³, **Table S1**). Across the 1,776 sets containing BMI, JASS detected 1,167 new
263 associations (after Bonferroni correction, **Material and methods**) of which 537 corresponded to a new
264 association in the larger BMI GWAS. 86 associations of the larger GWAS were missed by JASS. Hence,
265 JASS was able to flag loci with a high recall (0.86, probability of new association detected in the larger
266 GWAS to be detected by JASS) but a moderate specificity (0.46, probability of a detected loci to be
267 associated in the larger GWAS). This can be explained by the generality of the null hypothesis used in
268 JASS which requires only one trait in the set to be significant (not necessarily BMI). To improve
269 specificity, we fitted a logistic regression predicting if a locus would be associated with BMI in the
270 larger GWAS by combining the number of sets where the locus was associated by JASS and the
271 minimum *P*-value across sets (**Fig. S16**). This model reached an AUC of 0.75, an accuracy 0.74, a recall
272 of 46% and a specificity of 75% when applying a standard probability threshold of 0.5. Based on JASS
273 results, we were able to infer a substantial number of loci associated in a GWAS with twice as many
274 samples as the one used in the present study.

275 **Discussion**

276 This study investigated the genetic features associated with the statistical power of multi-trait
277 GWAS. On average, the power increase relative to the univariate GWAS was substantial: JASS detected
278 new association loci in 98% of 19,266 sets, with an average of 26 new association loci. This power
279 increase appears to be highly associated with the genetic features of trait sets. More specifically,
280 multi-trait gain tends to be higher for sets with 1) a moderate mean heritability (mean h_{GWAS}^2), 2) a
281 smaller mean MES, 3) a larger mean polygenicity, 4) a larger genetic covariance across traits, and 5) a
282 smaller distance between the residual and genetic covariance. Although some features, such as
283 increased distance between genetic and residual correlation, have been previously found associated
284 with the gain in association detection, this analysis suggests a negative or negligible gain in real data.

285 This might be explained by the complex correlation across features, highlighting the importance of
286 considering multiple features jointly or by variations of genetic architecture being narrower in real
287 data than assumed by previous simulation studies. We also found that selecting specific traits such as
288 BMI, and more clinically heterogenous sets, specifically for the omnibus test, can strongly outperform
289 approaches that select clinically homogenous sets. Finally, we investigated the predictive power of a
290 multivariate linear model that could predict trait sets that most likely benefit from the multi-trait test
291 (median Pearson's $p=0.43$, $P < 1.6 \times 10^{-60}$), which can be used to astutely select traits to be tested
292 jointly. Our findings provide an approach that can increase the identification of genetic associations
293 using existing GWAS data, with relevance to traits in which genetic signal is scarce. We summarize in
294 **Figure S18** a guideline for selecting the best suited methods according to the feature of their data.

295 Selecting clinically homogenous traits is the strategy most commonly used^{4,8,10,14-19}. In our previous
296 large-scale analysis⁸, an heterogeneous set yielded the largest number of new association as
297 compared to clinically homogenous sets. In this work we further showed that a data-driven strategy
298 is expected to outperform other strategies based on clinical insights. Such sets might capture highly
299 pleiotropic signals hard to detect using univariate GWAS and recommend that investigators compose
300 a heterogeneous set or use our predictive model to build a set of traits.

301 We compared the results from the standard omnibus test (implemented in JASS) against MTAG, a
302 popular multi-trait GWAS method²⁰⁻²³. For the data we used, the omnibus almost systematically
303 outperformed MTAG, with an overall three-fold increase in the number of loci detected. The
304 difference was particularly striking on larger trait sets. MTAG is built on the hypothesis of
305 homogeneous genetic correlation across genetic variants, and therefore is expected to have maximum
306 power when this assumption is valid. Indeed, we observed that more homogenous trait sets yielded
307 larger MTAG gain than heterogeneous trait sets (**Fig. S17A,B**). Yet, the omnibus generally performed
308 equally well or better than MTAG even on the homogenous trait sets (**Fig. S17C-F**). In contrast, by
309 construction, the omnibus test allows for substantial heterogeneity, although at the cost of an
310 increase in the degree of freedom. In line with previous work suggesting that genetic correlation might

311 be fairly heterogenous across the genome²⁴, this cost appears to be outweighed by the additional
312 flexibility in capturing heterogeneous multi-trait genetic patterns⁸ (**Supplementary notes 1, 2**). Thus,
313 the omnibus test is recommended for general identification of variants that impact phenotypes, while
314 MTAG is suitable for identifying variants associated with specific traits (**Fig. S18**).

315 Our study has some limitations. First, we focused on commonly measured genetic features, but
316 based on these results, a number of other refined metrics could be used. These include effect size
317 distribution as measured by the alpha parameters²⁵. Second, we considered GWAS derived from
318 common diseases and anthropometric traits. Future studies might explore performances using a wider
319 variety of molecular traits, for which GWAS summary statistics are becoming increasingly available.
320 Third, the estimation of the features might be also refined. Here, we used MiXeR^{26,27} to estimate most
321 features. However, we observed a dependency of MES and polygenicity on N_{eff} . Improving these
322 metrics could improve the overall analysis and interpretations of the results. Fourth, we focused on
323 European ancestry summary statistics. This decision was motivated by the availability of large GWAS
324 and using one ancestry for linkage disequilibrium; however, by doing so, we disposed of many traits
325 with which we could have had a wider variety of genetic features, which might have improved the
326 performance of the predictive model. This focus should not lead the reader to think that multi-trait
327 GWAS is useful only on large sample studies of European ancestry. We actually recently updated the
328 JASS pipeline to run a Multi-ancestry Multi-trait GWAS, which was able to detect 367 new association
329 loci, despite the modest sample size of the non-European cohorts used¹⁰. Future work might leverage
330 non-European existing²⁸ and upcoming biobanks²⁹ to investigate the validity of our results for non-
331 European ancestries.

332 In conclusion, this study provides a first overview of what to expect when applying multi-trait tests
333 to a variety of data and how to maximise new discoveries. These insights can be leveraged to discover
334 genetic variants associated with human complex traits and diseases missed by univariate analysis at
335 no cost. Beyond mapping, JASS used on clinically heterogeneous trait sets might offer a way to

336 understand a shared genetic aetiology among unexpected traits³⁰ and contribute to deeper
337 understanding of pleiotropy.

338 **Acknowledgments**

339 This research was supported by the Agence Nationale pour la Recherche (GenCAST, ANR-20-CE36-
340 0009). This work has been conducted as part of the INCEPTION program (Investissement d'Avenir
341 grant ANR-16-CONV-0005). MHC was supported by R01HL162813, R01HL153248, R01HL149861, and
342 R01HL147148.

343 **Author contributions**

344 Conceptualization: H.J., H.A.
345 Methodology: Y.S., H.J., H.A., C.B., A.A., M.H.C.
346 Software: Y.S., H.J., H.M., B.B., P.L., R.T., C.N., L.T.
347 Validation: Y.S., H.J., H.A., M.H.C., E.B.
348 Formal analysis: Y.S., H.J.
349 Investigation: Y.S., H.J.
350 Resources: H.A.
351 Data Curation: H.J., C.L., L.T.
352 Writing (original draft): Y.S., H.J., H.A.
353 Writing (review & editing): Y.S., R.V., C.B., A.A., M.H.C., E.B., H.A., H.J.
354 Visualization: Y.S., H.J., H.A.
355 Supervision: H.J., H.A.
356 Project administration: H.J., H.A., E.B.
357 Funding acquisition: H.J., H.A., E.B., M.H.C.

358 **Declaration of interests**

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

360 **Web resources**

361 JASS <https://jass.pasteur.fr/>

362 MeSH Browser <https://meshb-prev.nlm.nih.gov/search>

363 **Data and Code availability**

364 https://gitlab.pasteur.fr/statistical-genetics/jass_suite_pipeline

365 <https://gitlab.pasteur.fr/statistical-genetics/jass>

366 https://gitlab.pasteur.fr/statistical-genetics/multitrait_power_traitselection

367 **Material and Methods**

368 ***Database of curated summary statistics***

369 We assembled a database of 72 genome-wide GWAS summary statistics of quantitative traits and
370 diseases conducted in European ancestry population pulled from the GWAS catalogue³¹ and a variety
371 of publicly available meta-analyses. We cleaned, harmonised and imputed each study using our
372 previously developed pipeline². In brief, the process includes the following steps: 1) alignment of each
373 GWAS to the 1000G GRCh37 reference panel³², 2) imputation of missing summary statistics using
374 RAISS³³, 3) computation of the heritability, genetic and residual covariance matrices, referred further
375 as h_{GWAS}^2 , Σ_g and Σ_r , using LD-score regression³⁴, 4) aggregation of curated GWAS in a unique entry
376 file used as input for JASS. We filtered all GWAS with negative heritability, resulting in a total of 72
377 traits (**Table S1**). Curated GWAS summary statistics used in the analysis are available on the JASS
378 webserver <https://jass.pasteur.fr/>.

379 **Joint test and association gain**

380 Multi-trait analyses were conducted using the omnibus test implemented in the JASS package^{2,8}.
381 For a set of k GWAS, the omnibus statistics is defined as $T_{omni} = \mathbf{z}^t \boldsymbol{\Sigma}_r \mathbf{z}$ where \mathbf{z} is the vector of Z-
382 scores across traits $\mathbf{z} = (z_1 \dots z_k)$ and $\boldsymbol{\Sigma}_r$ is the residual Z-score covariance derived using the LD-score
383 regression. Under the null hypothesis of no association with any of the k phenotypes, T_{omni} follows a
384 χ^2 distribution with k degree of freedom. To maximize data usage, the default setting of JASS uses all
385 variants even those with missing association statistics. In this case, JASS returns association P -value
386 based on the subset of Z-scores available. In contrast, the *–remove-nans* option removes variants with
387 incomplete data. Here, we used *–remove-nans* option as the primary analysis for a better
388 characterisation of trait sets and for a fair comparison with the default setting of MTAG⁴ that do not
389 allow for missing statistics, whereas we also provide some results from the default setting of JASS as
390 an additional information.

391 All power comparisons were conducted at a locus level. The entire genome was split into a total of
392 1,703 quasi-independent loci defined based on linkage disequilibrium (LD)-independent blocks, as
393 proposed by Berisa and Pickrell³⁵. For both multi-trait and univariate analyses, we obtained the
394 minimum P -value across variants in each locus. The gain of the multi-trait test was derived as the
395 fraction of loci whose P -values were smaller than corresponding P -values in univariate GWAS
396 corrected for the number of traits jointly analysed:

$$397 \quad gain = f_{multi < univ} = \sum_i (P_{multi.i} < P_{uni.i} \cdot k) / R \quad (1)$$

398 where R is the number of loci ($R=1,703$), k is the number of traits (the number of GWAS studies)
399 jointly analysed, $P_{multi.i}$ is the minimum P -value of the multi-trait test in region i , and $P_{uni,i}$ is the
400 minimum P -value of the univariate tests across all GWAS analysed in locus i .

401

402 ***Estimated and derived genetic features***

403 We investigated the effect of both single and multi-trait features. Single GWAS features include
404 the effective sample size (N_{eff}), linear additive common variants heritability (h_{GWAS}^2), polygenicity,
405 MES, and the proportion of uncaptured linear additive common variants heritability ($\%h_u^2$). Multi-trait
406 GWAS features include the number of traits in a trait set (k), the average of the off-diagonal of genetic
407 covariance and the residual covariance ($\bar{\Sigma}_g$ and $\bar{\Sigma}_r$, respectively), condition numbers of genetic
408 covariance matrix and residual covariance matrix (κ_g and κ_r , respectively) across traits, and average
409 distance between the genetic and residual correlation matrices (Δ_{Σ}). All parameters were aggregated
410 to form a vector of 11 features per trait set. For the single GWAS parameters, MES, polygenicity, N_{eff} ,
411 h_{GWAS}^2 , and $\%h_u^2$, we calculated mean values across each set of traits.

412 Polygenicity and heritability (h_{GWAS}^2) were estimated using MiXeR^{26,27}, with the 1000 Genomes
413 Phase3 reference panel provided along the MiXeR package containing approximately 10 million
414 common variants³². Following the authors recommendation, we defined the parameter for effective
415 sample size as $N_{eff} = 1/(1/N_{case} + 1/N_{controls})$. For comparison purposes, we also estimated
416 h_{GWAS}^2 using the LDscore regression³⁴, and the two metrics were consistent (Pearson $\rho=0.86$, **Fig. S5**).
417 The estimated polygenicity by MiXeR showed a dependency on the GWAS sample size, with about a
418 10-fold increase of the polygenicity for an increase of 500,000 of the sample size (**Fig. S19**). We
419 therefore adjusted polygenicity by taking the residuals of linear regression between \log_{10} polygenicity
420 and N_{eff} : $\log_{10} \text{polygenicity (adjusted)} = \log_{10} \text{polygenicity (mixer)} - \alpha N_{eff}$, where α was
421 estimated by a linear regression $\log_{10} \text{polygenicity (mixer)} \sim \alpha N_{eff} + \varepsilon$. We obtained 'adjusted
422 polygenicity' as $10^{[\log_{10} \text{polygenicity (adjusted)}]}$. We computed MES as $\frac{h_{GWAS}^2}{\text{polygenicity (adjusted)}}$.

423 The proportion of uncaptured linear additive common variants heritability ($\%h_u^2$) was derived as
424 $(h_{GWAS}^2 - h_{GWAS-hits}^2)/h_{GWAS}^2$, where $h_{GWAS-hits}^2$ denotes the heritability accounted by the
425 univariate GWAS association loci. It was derived using the lead variants from each locus reaching

426 genome-wide significance ($P < 5 \times 10^{-8}$): $h_{GWAS-hits}^2 = \sum_{i \in I} \beta_i^2$, where $\beta_i = z_i / \sqrt{N_{eff}}$. We excluded
427 loci with lead variant whose $abs(\beta) > 0.19$, because $h_{GWAS-hits}^2$ tended to become larger than h_{GWAS}^2 .
428 The mean genetic covariance and mean residual covariance were defined as the mean of the
429 absolute value of the upper off-diagonal elements of the genetic and residual covariance matrices (Σ_g
430 and Σ_r), i.e. $\bar{\sigma}_g = \sum_{i,j;i < j} \sigma_{gij} / \sum_{i,j;i < j} 1$, and $\bar{\sigma}_r = \sum_{i,j;i < j} \sigma_{rij} / \sum_{i,j;i < j} 1$, respectively, where
431 σ_{gij} and σ_{rij} are ij elements of Σ_g and Σ_r , and k is the number of traits. The condition numbers of
432 genetic and residual covariance matrices were computed as: $\kappa_g = \sqrt{\max(\lambda_{g,i}) / \min(\lambda_{g,i})}$, and $\kappa_r =$
433 $\sqrt{\max(\lambda_{r,i}) / \min(\lambda_{r,i})}$ where $\lambda_{g,i}$ and $\lambda_{r,i}$ are the eigenvalues of the genetic and residual
434 covariance matrices. We used `numpy.linalg.eig`³⁶ for the eigen decomposition and assigned an infinite
435 value to κ when the minimum eigenvalue was negative or close to zero. The average distance Δ_Σ was
436 defined as the mean over the absolute values of pairwise difference between the corresponding upper
437 off-diagonal elements in genetic and residual correlation matrices $\Delta_\Sigma =$
438 $\sum_{i,j;i < j} |\rho_{gij} - \rho_{rij}| / \sum_{i,j;i < j} 1$, where ρ_{gij} and ρ_{rij} indicate the ij elements in the genetic and
439 residual correlation matrices.

440

441 ***Assessment of features associated with multi-trait association gain***

442 The contribution of features to multi-trait gain was estimated using a five-fold cross validation. For
443 each round of cross validation, the 72 GWAS were randomly split into a training and validation data,
444 each including 36 GWAS. Within each cross validation, we generated 1,980 unique random trait sets,
445 containing 2 to 12 traits, for each of the training and validation data: random sampling out of the 36
446 traits (660 sets generated), random sampling out of traits with common SNP heritability below median
447 (330 sets) and above median (330 sets); random sampling out of traits with MES below median (330
448 sets) and above median (330 sets). For this stratification of traits by the median of MES, we used
449 polygenicity / h_{GWAS}^2 estimated by MiXeR without adjusting for the effective sample size. For each set,

450 we ran JASS and derived the 11 features of interest ($F_i, i = 1, \dots, 11$) and the multi-trait gain. We used
451 19,793 trait sets out of the total 19,800 ($=1,980 \times 2 \times 5$) sampled for which the whole analysis process
452 completed without error. Errors include cases where there was no association detected by both the
453 univariate and joint tests.

454 Moving to the multivariate regression analysis, we selected six out of the 11 features based on
455 collinearity analysis (as described below). The six features and the multi-trait gain were standardised
456 into a range between 0 and 1 using *MinMaxScaler* in scikit-learn³⁷. This standardisation was applied at
457 once on the entire dataset including training and validation data across the five-fold cross validation
458 sets.

459 We used the training data to estimate the joint effect $\hat{\delta}_i$ of each feature i from a multiple
460 regression: $gain_{train} \sim \sum_i \delta_i F_{train,i}$. This was conducted using the *OLS* function in statsmodels in
461 Python³⁸. We report Pearson's correlation coefficient as a metric of predictive power.

462

463 ***Collinearity and selection of features***

464 We observed collinearity among some of the 11 features of traits across 19,266 unique trait sets.
465 $\log_{10} \kappa_r$, $\log_{10} \kappa_g$, $\log_{10} \bar{\Sigma}_r$, and the number of traits were highly correlated (Pearson $\rho > 0.65$).
466 Likewise, mean $\%h^2_u$, mean $\log_{10}\text{MES}$, and mean $\log_{10}\text{polygenicity}$ were highly correlated ($\text{abs}(\rho) >$
467 0.8). These correlated features capture redundant characteristics of traits. Thus, we selected one out
468 of each correlated features: the number of traits and mean $\%h^2_u$. We chose the number of traits
469 because it had the smallest P -value in a multivariate linear regression with all the features included,
470 where inf values in condition number were replaced with their non- inf max value. We chose
471 mean($\%h^2_u$) because it captures both mean($\log_{10}\text{MES}$) ($\rho=-0.90$) and mean($\log_{10}\text{polygenicity}$) ($\rho=0.84$).
472 The estimation of mean($\%h^2_u$) is also more straightforward than mean($\log_{10}\text{MES}$) and
473 mean($\log_{10}\text{polygenicity}$). After this pre-selection, we built models with the remaining six features as
474 described in the previous section.

475 **Non-linear models**

476 We considered two alternative non-linear models for prediction purposes: support vector
477 regression (SVR), and random forest regression (RFR). SVR and RFR are regression approaches that
478 allow for non-linear relationships. SVR's goal is to find a hyperplane (or line, in the case of two-
479 dimensional data) that best fits the data. It is effective at handling non-linear and complex data by
480 using the kernel trick—mapping data with a kernel function into a higher-dimensional space where it
481 is easier to find the best-fit hyperplane. SVR also penalises the complexity of the model and gives the
482 flexibility in how much error is acceptable. Random forest regression performs a regression using
483 decision trees. It generates multiple trees, fits each to a random subset of training data, and averages
484 the predictions across trees as the final prediction. While the random sampling and averaging
485 supposedly makes the model robust to outliers, RFR's performance relies on a high quality of training
486 data; the training data needs to cover a wide range as RFR does not work well for extrapolation, and
487 RFR leads to biased predictions when the training data is sampled in a biased way³⁹. In contrast, SVR
488 is suggested to be capable of extrapolation⁴⁰. We used the scikit-learn³⁷ python implementation of
489 RFR and of the SVR. We fitted SVR and RFR models to the *gain_{train}* including hyperparameters.
490 Hyperparameters were tuned using *RandomizedSearchCV* in scikit-lean across the following range:
491 SVR's kernel=[linear, rbf, sigmoid, poly], C=[1,10,50,100], epsilon=[10⁻³,10⁻²,10⁻¹,1], degree=[2,3,4],
492 and RFR's n_estimators=[5,20,50,100], max_features=['auto','sqrt'], max_depth=[12 values ranging
493 from 2 to 100], min_samples_split=[2,5,10], min_samples_leaf=[1,2,4], bootstrap=[True, False].

494

495 **MTAG analyses**

496 For comparison purposes, we repeated the multi-trait test and the prediction analyses with the
497 MTAG approach⁴. MTAG uses a weighted sum of Z-score (**Supplementary note 1**). Trait weights are
498 derived using the generalised method of moments, as $(\hat{\beta}_j \hat{\beta}_j' - \Omega - \Sigma_j) = 0$, where Ω is the genetic
499 covariance matrix to be estimated for the weights, and Σ_j is the genetic covariance matrix estimated

500 using the LD score regressions by Bulik-Sullivan et al³⁴. The model assumes that the genetic covariance
501 matrix is homogenous across variants. We used MTAG with its default setting, which considered only
502 complete cases. We ran MTAG for the same trait sets used for the analysis with JASS (all data used in
503 the five-fold cross validation). MTAG outputs *P*-values for each trait in each trait set, whereas JASS
504 gives a single *P*-value for a trait set. To account for the number of tests run by MTAG for one set, we
505 obtained minimum *P*-values by MTAG across traits and variants in each locus and multiplied the
506 minimum *P*-values by the number of traits in the set. For the comparison of the association gain
507 between MTAG and JASS, we used both the minimum *P*-values with and without multiple test
508 correction (Fig. 5, and Fig. S14).

509

510 ***Comparison of strategies for trait selection.***

511 To compare the performances of trait selection strategies, we classified the 19,266 unique sets into
512 clinically homogenous, low heterogeneity, high heterogeneity, or high predicted gain according to our
513 predictive model. To assess clinical homogeneity, we first classified the 72 traits into clinical groups
514 using the broadest categories in the MeSH Tree Structures⁴¹. The grouping was further refined based
515 on clinician's insights (Table S1). We labeled each set of traits as 'homogeneous' if all the traits are in
516 the same clinical group, 'low heterogeneity' if trait belonged to two to four clinical groups, 'high
517 heterogeneity' if traits spanned five or more clinical groups. For the 'data-driven' method, we selected
518 100 sets of traits by CV fold that had the largest gains predicted by the linear model with aggregated
519 coefficients across CV folds (Table 1). We used the Welch's *t*-test to evaluate the differences in gain
520 and the number of new associated loci. We used a two-sided Welch's *t*-tests to determine whether
521 the data driven method achieves a greater association gain than other methods, and whether jointly
522 analysing clinically heterogenous traits achieves a greater association gain than jointly analysing
523 clinically homogenous traits. For this test, we used a pair of training data and validation data that are
524 mutually exclusive to ensure independence.

525

526 ***Evaluation of the relevance of new association***

527 To evaluate the relevance of new associations detected by JASS (i.e. if most of them were true
528 positive), we attempted to predict loci discovered in a recent large meta-analysis on BMI (sample size
529 of 683,365 on average across ~2.3 million variants¹²) from the results of multi-trait GWAS applied on
530 1,776 trait sets containing a smaller study of BMI (sample size of 339,224, **Table S1**). First, we
531 compared loci detected by JASS (after a Bonferroni correction to account for the number of sets) and
532 in the larger GWAS using the standard genome wide significance threshold of 5×10^{-8} . Second, we
533 fitted a logistic regression to predict associated loci in the larger GWAS by combining JASS *P*-value and
534 the number of sets where the loci was considered associated with JASS. We use odd number
535 chromosomes to fit the logistic regression and evaluated its performances on even number
536 chromosomes.

537

538 **References**

- 539 1. Bhattacharjee, S., Rajaraman, P., Jacobs, K.B., Wheeler, W.A., Melin, B.S., Hartge, P., Yeager, M.,
540 Chung, C.C., Chanock, S.J., and Chatterjee, N. (2012). A subset-based approach improves power
541 and interpretation for the combined analysis of genetic association studies of heterogeneous
542 traits. *The American Journal of Human Genetics* 90, 821–835. 10.1016/j.ajhg.2012.03.015.
- 543 2. Julienne, H., Lechat, P., Guillemot, V., Lasry, C., Yao, C., Araud, R., Laville, V., Vilhjalmsson, B.,
544 Ménager, H., and Aschard, H. (2020). JASS: command line and web interface for the joint analysis
545 of GWAS results. *NAR genomics and bioinformatics* 2, lqaa003. 10.1093/nargab/lqaa003.
- 546 3. Qi, G., and Chatterjee, N. (2018). Heritability informed power optimization (HIPO) leads to
547 enhanced detection of genetic associations across multiple traits. *PLoS genetics* 14, e1007549.
548 10.1371/journal.pgen.1007549.
- 549 4. Turley, P., Walters, R.K., Maghzian, O., Okbay, A., Lee, J.J., Fontana, M.A., Nguyen-Viet, T.A.,
550 Wedow, R., Zacher, M., Furlotte, N.A., et al. (2018). Multi-trait analysis of genome-wide
551 association summary statistics using MTAG. *Nat Genet* 50, 229–237. 10.1038/s41588-017-0009-
552 4.

553 5. Wang, M., Cao, X., Zhang, S., and Sha, Q. (2022). sCLC: a Clustering Linear Combination (CLC)
554 Method for Multiple Phenotype Association Studies Based on GWAS Summary Statistics. Preprint,
555 10.21203/rs.3.rs-2284388/v1 10.21203/rs.3.rs-2284388/v1.

556 6. Zhu, X., Feng, T., Tayo, B.O., Liang, J., Young, J.H., Franceschini, N., Smith, J.A., Yanek, L.R., Sun,
557 Y.V., and Edwards, T.L. (2015). Meta-analysis of correlated traits via summary statistics from
558 GWASs with an application in hypertension. *The American Journal of Human Genetics* *96*, 21–36.
559 10.1016/j.ajhg.2014.11.011.

560 7. Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., and Yang, J. (2017).
561 10 years of GWAS discovery: biology, function, and translation. *The American Journal of Human
562 Genetics* *101*, 5–22. 10.1016/j.ajhg.2017.06.005.

563 8. Julienne, H., Laville, V., McCaw, Z.R., He, Z., Guillemot, V., Lasry, C., Ziyatdinov, A., Nerin, C.,
564 Vaysse, A., and Lechat, P. (2021). Multitrait GWAS to connect disease variants and biological
565 mechanisms. *PLoS genetics* *17*, e1009713. 10.1371/journal.pgen.1009713.

566 9. Porter, H.F., and O'Reilly, P.F. (2017). Multivariate simulation framework reveals performance of
567 multi-trait GWAS methods. *Sci Rep* *7*, 38837. 10.1038/srep38837.

568 10. Troubat, L., Fettahoglu, D., Aschard, H., and Julienne, H. (2023). Multi-trait GWAS for diverse
569 ancestries: Mapping the knowledge gap. *bioRxiv*, 2023–06. 10.1101/2023.06.23.546248.

570 11. Simon, J.J., Skunde, M., Hamze Sinno, M., Brockmeyer, T., Herpertz, S.C., Bendszus, M., Herzog,
571 W., and Friederich, H.-C. (2014). Impaired cross-talk between mesolimbic food reward processing
572 and metabolic signaling predicts body mass index. *Frontiers in Behavioral Neuroscience* *8*, 359.
573 10.3389/fnbeh.2014.00359.

574 12. Yengo, L., Sidorenko, J., Kemper, K.E., Zheng, Z., Wood, A.R., Weedon, M.N., Frayling, T.M.,
575 Hirschhorn, J., Yang, J., and Visscher, P.M. (2018). Meta-analysis of genome-wide association
576 studies for height and body mass index in ~ 700000 individuals of European ancestry. *Human
577 molecular genetics* *27*, 3641–3649. 10.1093/hmg/ddy271.

578 13. Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R., Powell, C., Vedantam, S.,
579 Buchkovich, M.L., and Yang, J. (2015). Genetic studies of body mass index yield new insights for
580 obesity biology. *Nature* *518*, 197–206. 10.1038/nature14177.

581 14. Fan, C.C., Loughnan, R., Makowski, C., Pecheva, D., Chen, C.-H., Hagler, D.J., Thompson, W.K.,
582 Parker, N., van der Meer, D., Frei, O., et al. (2022). Multivariate genome-wide association study
583 on tissue-sensitive diffusion metrics highlights pathways that shape the human brain. *Nat
584 Commun* *13*, 2423. 10.1038/s41467-022-30110-3.

585 15. Gallois, A., Mefford, J., Ko, A., Vaysse, A., Julienne, H., Ala-Korpela, M., Laakso, M., Zaitlen, N.,
586 Pajukanta, P., and Aschard, H. (2019). A comprehensive study of metabolite genetics reveals
587 strong pleiotropy and heterogeneity across time and context. *Nat Commun* *10*, 4788.
588 10.1038/s41467-019-12703-7.

589 16. Khunsriraksakul, C., Li, Q., Markus, H., Patrick, M.T., Sauteraud, R., McGuire, D., Wang, X., Wang,
590 C., Wang, L., Chen, S., et al. (2023). Multi-ancestry and multi-trait genome-wide association meta-
591 analyses inform clinical risk prediction for systemic lupus erythematosus. *Nat Commun* *14*, 668.
592 10.1038/s41467-023-36306-5.

593 17. Levin, M.G., Tsao, N.L., Singhal, P., Liu, C., Vy, H.M.T., Paranjpe, I., Backman, J.D., Bellomo, T.R.,
594 Bone, W.P., Biddinger, K.J., et al. (2022). Genome-wide association and multi-trait analyses
595 characterize the common genetic architecture of heart failure. *Nat Commun* *13*, 6914.
596 10.1038/s41467-022-34216-6.

597 18. van der Meer, D., Frei, O., Kaufmann, T., Shadrin, A.A., Devor, A., Smeland, O.B., Thompson, W.K.,
598 Fan, C.C., Holland, D., Westlye, L.T., et al. (2020). Understanding the genetic determinants of the
599 brain with MOSTest. *Nat Commun* *11*, 3512. 10.1038/s41467-020-17368-1.

600 19. Xiong, Z., Gao, X., Chen, Y., Feng, Z., Pan, S., Lu, H., Uitterlinden, A.G., Nijsten, T., Ikram, A.,
601 Rivadeneira, F., et al. (2022). Combining genome-wide association studies highlight novel loci
602 involved in human facial variation. *Nat Commun* *13*, 7832. 10.1038/s41467-022-35328-9.

603 20. Rahmioglu, N., Mortlock, S., Ghiasi, M., Møller, P.L., Stefansdottir, L., Galarneau, G., Turman, C.,
604 Danning, R., Law, M.H., and Sapkota, Y. (2023). The genetic basis of endometriosis and
605 comorbidity with other pain and inflammatory conditions. *Nature genetics* *55*, 423–436.
606 10.1038/s41588-023-01323-z.

607 21. Sveinbjornsson, G., Ulfarsson, M.O., Thorolfsdottir, R.B., Jonsson, B.A., Einarsson, E.,
608 Gunnlaugsson, G., Rognvaldsson, S., Arnar, D.O., Baldvinsson, M., and Bjarnason, R.G. (2022).
609 Multiomics study of nonalcoholic fatty liver disease. *Nature Genetics* *54*, 1652–1663.
610 10.1038/s41588-022-01199-5.

611 22. Wang, Y., Tsuo, K., Kanai, M., Neale, B.M., and Martin, A.R. (2022). Challenges and opportunities
612 for developing more generalizable polygenic risk scores. *Annual review of biomedical data science*
613 *5*, 293–320. 10.1146/annurev-biodatasci-111721-074830.

614 23. Wang, Z., Emmerich, A., Pillon, N.J., Moore, T., Hemerich, D., Cornelis, M.C., Mazzaferro, E., Broos,
615 S., Ahluwalia, T.S., and Bartz, T.M. (2022). Genome-wide association analyses of physical activity
616 and sedentary behavior provide insights into underlying mechanisms and roles in disease
617 prevention. *Nature genetics* *54*, 1332–1344. 10.1038/s41588-022-01165-1.

618 24. Werme, J., van der Sluis, S., Posthuma, D., and de Leeuw, C.A. (2022). An integrated framework
619 for local genetic correlation analysis. *Nature genetics* *54*, 274–282. 10.1038/s41588-022-01017-y.

620 25. Miao, L., Jiang, L., Tang, B., Sham, P.C., and Li, M. (2023). Dissecting the high-resolution genetic
621 architecture of complex phenotypes by accurately estimating gene-based conditional heritability.
622 *The American Journal of Human Genetics* *110*, 1534–1548. 10.1016/j.ajhg.2023.08.006.

623 26. Frei, O., Holland, D., Smeland, O.B., Shadrin, A.A., Fan, C.C., Maeland, S., O'Connell, K.S., Wang, Y.,
624 Djurovic, S., and Thompson, W.K. (2019). Bivariate causal mixture model quantifies polygenic
625 overlap between complex traits beyond genetic correlation. *Nature communications* *10*, 2417.
626 10.1038/s41467-019-10310-0.

627 27. Holland, D., Frei, O., Desikan, R., Fan, C.-C., Shadrin, A.A., Smeland, O.B., Sundar, V.S., Thompson,
628 P., Andreassen, O.A., and Dale, A.M. (2020). Beyond SNP heritability: Polygenicity and
629 discoverability of phenotypes estimated with a univariate Gaussian mixture model. *PLoS Genetics*
630 *16*, e1008612. 10.1371/journal.pgen.1008612.

631 28. Ishigaki, K., Akiyama, M., Kanai, M., Takahashi, A., Kawakami, E., Sugishita, H., Sakaue, S., Matoba,
632 N., Low, S.-K., Okada, Y., et al. (2020). Large-scale genome-wide association study in a Japanese
633 population identifies novel susceptibility loci across different diseases. *Nat Genet* *52*, 669–679.
634 10.1038/s41588-020-0640-3.

635 29. The “All of Us” Research Program (2019). *New England Journal of Medicine* 381, 668–676.
636 10.1056/NEJMsr1809937.

637 30. Hackinger, S., and Zeggini, E. (2017). Statistical methods to detect pleiotropy in human complex
638 traits. *Open biology* 7, 170125. 10.1098/rsob.170125.

639 31. Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon,
640 A., Morales, J., Mountjoy, E., and Sollis, E. (2019). The NHGRI-EBI GWAS Catalog of published
641 genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic acids
642 research* 47, D1005–D1012. 10.1093/nar/gky1120.

643 32. 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. *Nature*
644 526, 68. 10.1038/nature15393.

645 33. Julienne, H., Shi, H., Pasaniuc, B., and Aschard, H. (2019). RAISS: robust and accurate imputation
646 from summary statistics. *Bioinformatics* 35, 4837–4839. 10.1093/bioinformatics/btz466.

647 34. Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.-R., Duncan, L., Perry, J.R.,
648 Patterson, N., and Robinson, E.B. (2015). An atlas of genetic correlations across human diseases
649 and traits. *Nature genetics* 47, 1236–1241. 10.1038/ng.3406.

650 35. Berisa, T., and Pickrell, J.K. (2016). Approximately independent linkage disequilibrium blocks in
651 human populations. *Bioinformatics* 32, 283. 10.1093/bioinformatics/btv546.

652 36. Harris, C.R., Millman, K.J., Van Der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser,
653 E., Taylor, J., Berg, S., and Smith, N.J. (2020). Array programming with NumPy. *Nature* 585, 357–
654 362. 10.1038/s41586-020-2649-2.

655 37. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M.,
656 Prettenhofer, P., Weiss, R., and Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *the
657 Journal of machine Learning research* 12, 2825–2830.

658 38. Seabold, S., and Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with
659 python. In (Austin, TX), pp. 10–25080.

660 39. Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B., and Gräler, B. (2018). Random forest as a
661 generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ* 6,
662 e5518. 10.7717/peerj.5518.

663 40. Balabin, R.M., and Smirnov, S.V. (2012). Interpolation and extrapolation problems of multivariate
664 regression in analytical chemistry: benchmarking the robustness on near-infrared (NIR)
665 spectroscopy data. *Analyst* 137, 1604–1610. 10.1039/C2AN15972D.

666 41. National Institutes of Health (2010). National Library of Medicine—Medical Subject Headings.

667

668 **Tables and Figures**

669 **Table 1. Coefficients of the multivariate linear regression models from five-fold cross**

670 **validations.** Mean genetic residual distance and mean genetic covariance were \log_{10}

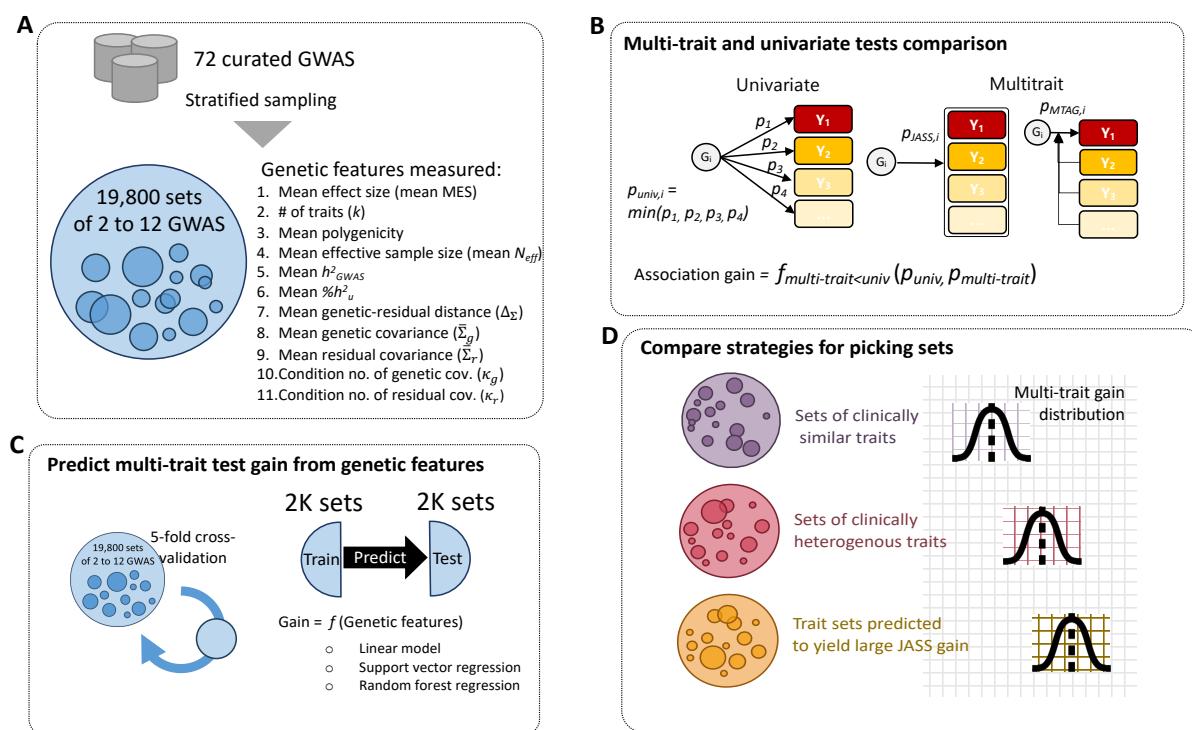
671 transformed. All the features were scaled using *MinMaxScaler* in *scikit-learn*³⁷.

	CV1	CV2	CV3	CV4	CV5	Mean
# of traits	0.096 (p= 3.447x10 ⁻⁴⁴)	0.077 (p= 1.592x10 ⁻¹⁷)	0.082 (p= 1.638x10 ⁻⁴⁴)	0.049 (p= 6.148x10 ⁻⁸)	0.083 (p= 3.071x10 ⁻¹⁹)	0.077 (std=0.017)
Mean $\log_{10} \Delta_{\Sigma}$	-0.787 (p= 8.338x10 ⁻²⁹)	-0.536 (p= 8.338x10 ⁻²⁹)	-0.393 (p= 2.242x10 ⁻¹⁰)	-1.174 (p= 1.086x10 ⁻²⁶)	-0.610 (p= 1.947x10 ⁻⁰⁹)	-0.700 (std=0.301)
Mean $\log_{10} \bar{\Sigma}_g$	0.767 (p= 3.813x10 ⁻⁴⁵)	0.789 (p= 2.055x10 ⁻²⁰)	0.355 (p= 5.990x10 ⁻¹⁵)	1.149 (p= 1.604x10 ⁻⁵⁶)	0.675 (p= 5.636x10 ⁻²¹)	0.747 (std=0.284)
Mean N_{eff}	0.206 (p= 4.342x10 ⁻¹⁵)	-0.016 (p= 0.554)	-0.071 (p= 0.003)	0.260 (p= 1.099x10 ⁻¹⁰)	-0.015 (p= 0.598)	0.073 (std=0.149)
Mean h^2_{GWAS}	-0.509 (p= 2.937x10 ⁻⁹⁶)	-0.437 (p= 1.455x10 ⁻³⁴)	-0.518 (p= 2.216x10 ⁻¹¹¹)	-0.690 (p= 6.670x10 ⁻⁷⁹)	-0.429 (p= 2.225x10 ⁻²⁶)	-0.516 (std=0.105)
Mean $\%h^2_u$	0.112 (p= 5.158x10 ⁻⁶)	0.370 (p= 4.353x10 ⁻³⁵)	0.091 (p= 3.877x10 ⁻¹⁰)	0.277 (p= 3.022x10 ⁻³³)	-0.064 (p= 0.002)	0.157 (std=0.169)

672

673 **Figure 1. Study overview**

674 We conducted a power analysis on real data to understand in which setting a standard multi-trait
 675 test—the omnibus test—outperforms univariate GWAS. A) To assemble our real data, we curated 72
 676 GWAS summary statistics and formed about 20k sets of traits by random sampling. Each set of traits
 677 was characterised by assessing key genetic features such as polygenicity, mean effect size (MES), and
 678 heritability. (B) For each set of traits, we ran omnibus test using JASS and computed the association
 679 gain compared to the univariate test. We defined this association gain as the number of LD-
 680 independent loci where the omnibus yields a smaller P -value than the univariate test. We repeated
 681 the analysis using MTAG, a popular multi-trait approach. (C) To investigate which genetic features
 682 (highlighted in A) explain the JASS (omnibus) association gain, we applied statistical models to predict
 683 the gain as a function of genetic features. Several models were benchmarked to optimise prediction
 684 performances. (D) To suggest a practical strategy for selecting traits that yield a large association gain,
 685 we compared the performance of JASS on trait sets that are clinically similar, clinically heterogenous,
 686 and that were predicted to have a large gain by the predictive model highlighted in (C).

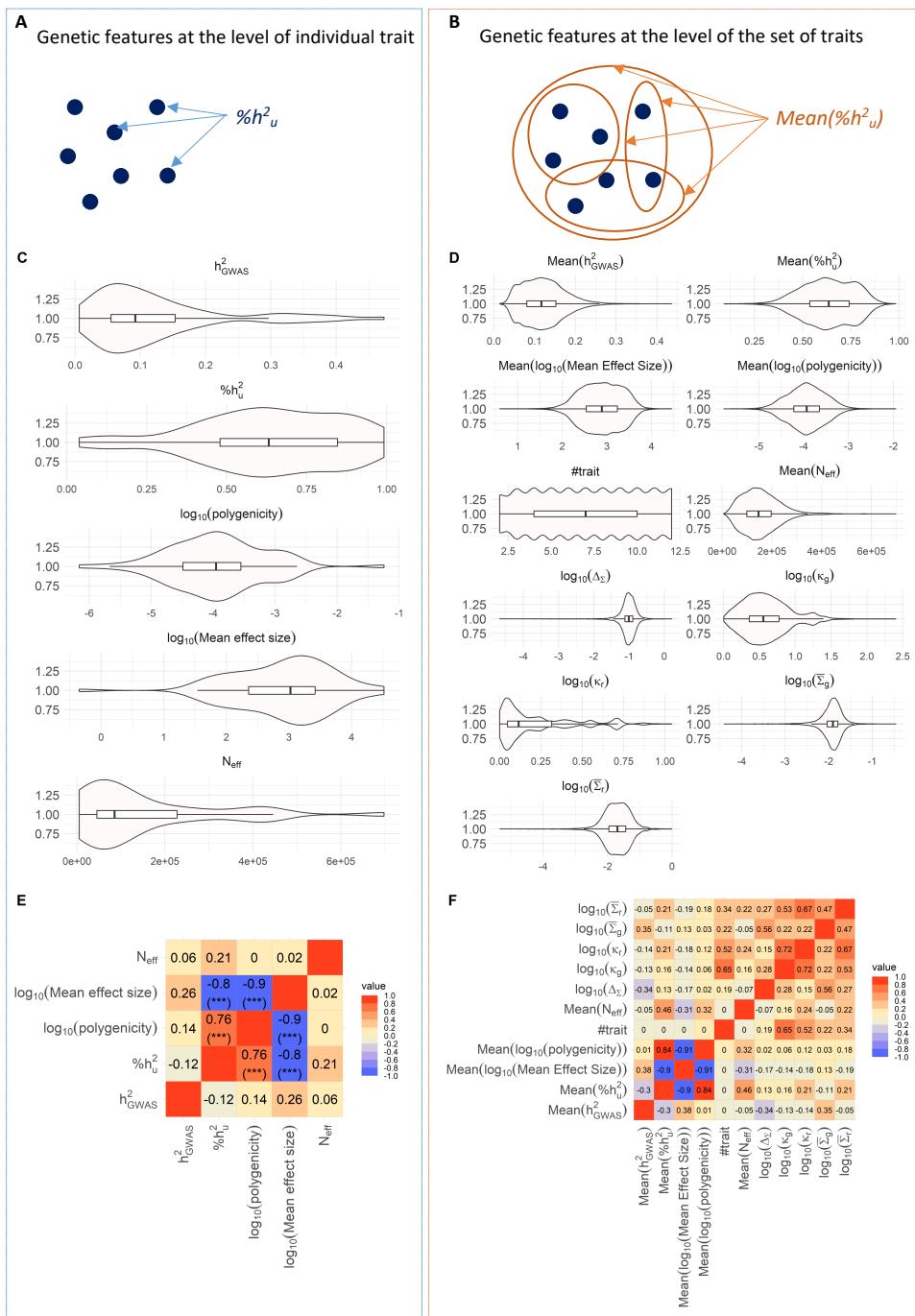


687

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

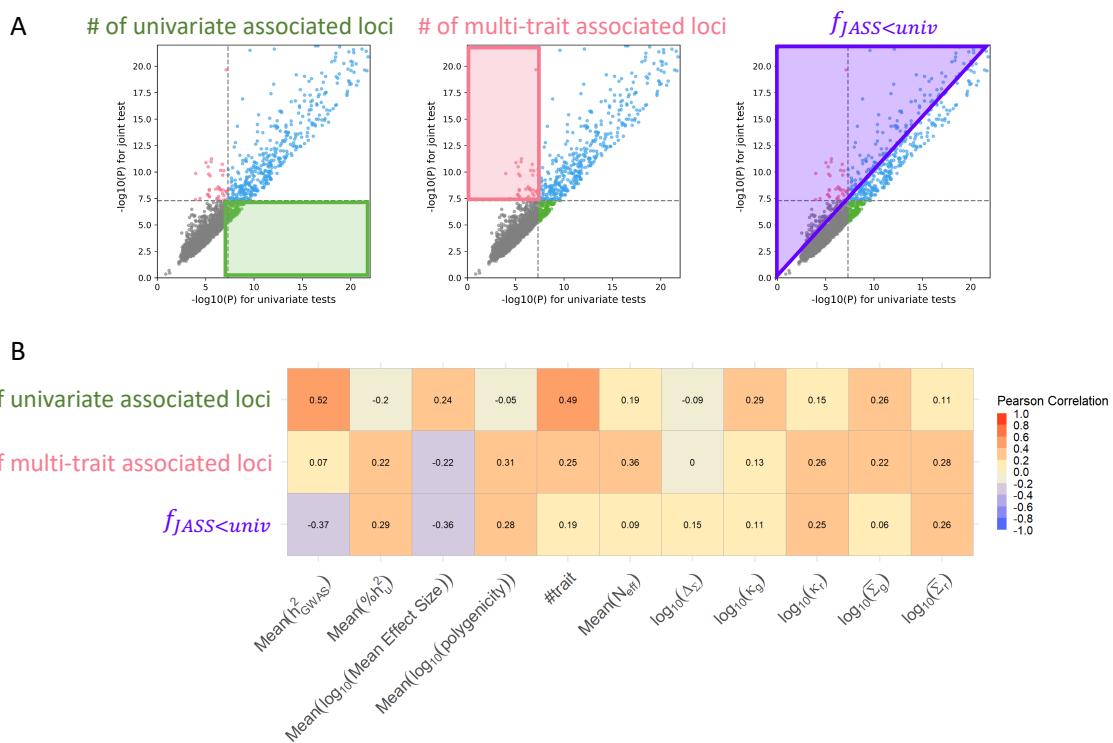
689 **sets.**

690 Visualization of the investigated features and their relation at the level of individual trait (panels A, C,
691 and E) and at the level of set of traits (panels B, D, and F). A) Schematic of a genetic feature derived at
692 the level of an individual trait. B) Schematic of a genetic feature derived at the level of a set of traits.
693 C) Violin plots representing the distribution across the 72 summary statistics of heritability,
694 polygenicity, MES and sample size of the study. D) Violin plots representing the distribution across the
695 19,266 sets of traits of the 11 genetic features derived for each trait set. E) Pearson correlation among
696 polygenicity, MES, h^2_{GWAS} , $\%h^2_u$, and sample size across 72 traits. F) Pearson correlation among the 11
697 features across 19,266 trait sets. (q-value annotation: *** $< 10^{-3}$, ** $< 10^{-2}$, * $< 5 \times 10^{-2}$)



699 **Figure 3. Determinant of JASS gain across trait sets.**

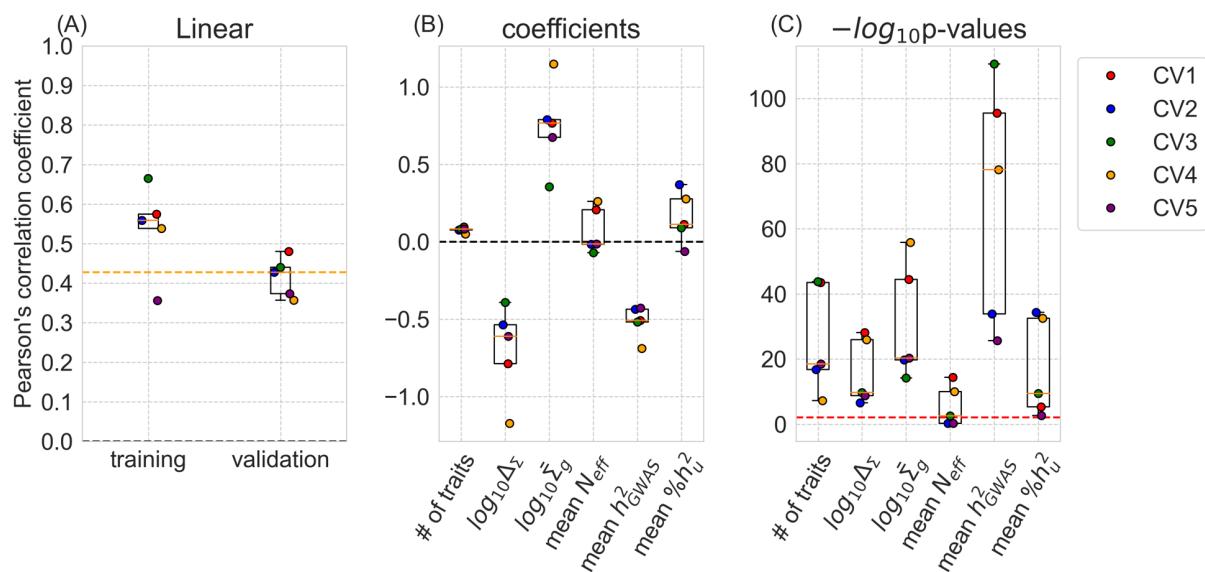
700 A) Illustration of the three metrics to assess univariate and multi-trait GWAS outcomes. On a quadrant
 701 plot representing the P -value of the multi-trait test with respect to the P -value of the univariate test,
 702 the following areas represent regions where: (green) only the univariate test is significant, (pink) only
 703 the multi-trait test is significant, (purple) the multi-trait test is more significant than the univariate
 704 test. B) Heatmap of the Pearson correlation between the number of univariate association loci, the
 705 number of new association loci detected by JASS, the association gain of JASS ($f_{JASS<univ}$) and the 11
 706 genetic features across 19,266 trait sets.



707

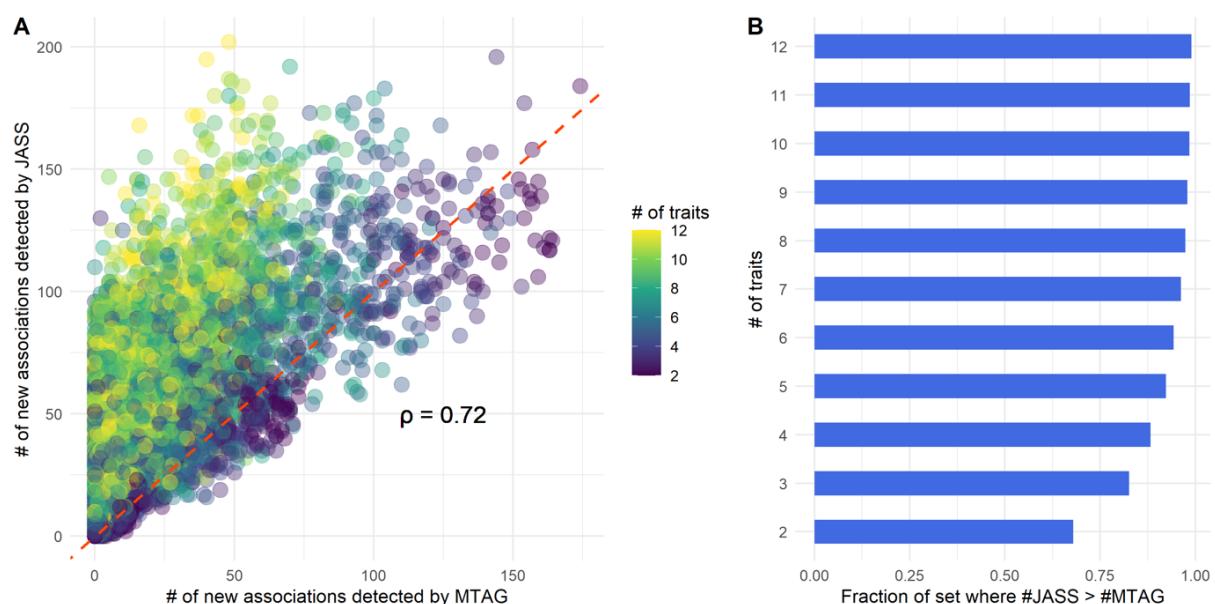
708 **Figure 4. Model prediction power and feature contributions.** (A) Boxplots of the prediction
709 power across the five-fold cross validations (CV) of the multivariate linear regression model
710 measured as the Pearson's correlation coefficient between the predicted and observed gain.
711 The performance of each CV is represented as a coloured dot. Orange dashed line: median
712 correlation coefficient between the predicted and observed gain in the validation data. (B, C)
713 The boxplots show the coefficients and $-\log_{10}(P\text{-values})$ of the six features in the regression
714 model across five-fold cross validations using each corresponding training data. Red dashed
715 line: Bonferroni corrected nominal significance threshold.

716



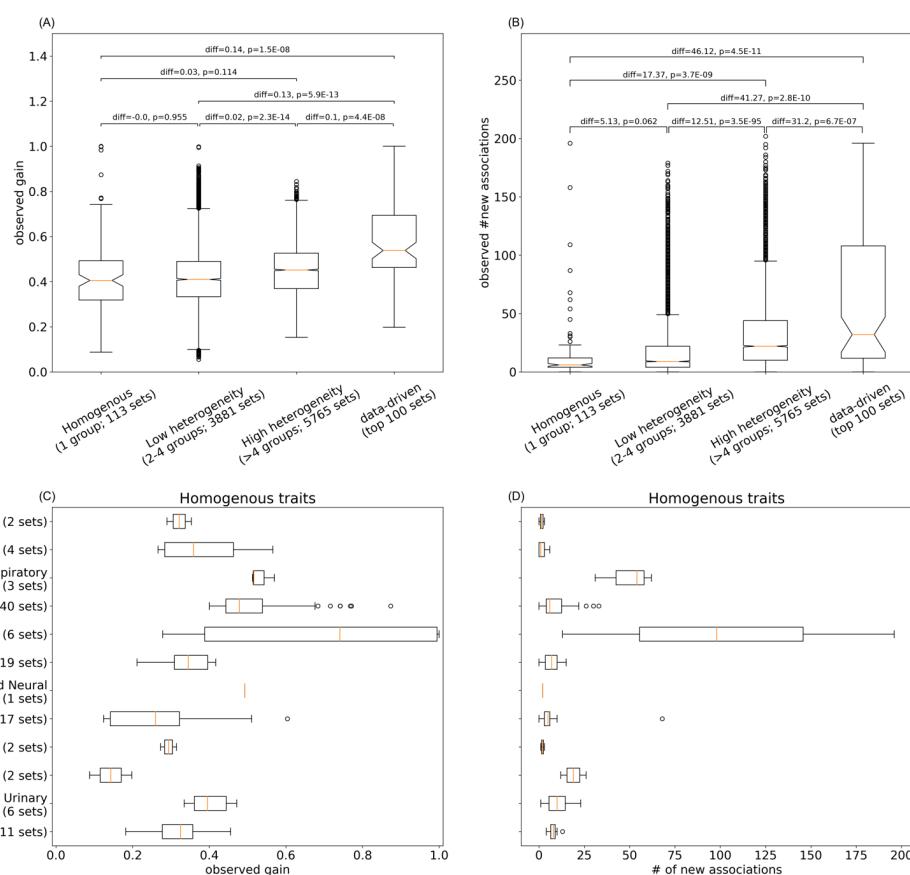
717

718 **Figure 5. Comparison of MTAG with JASS.** A) The number of new association loci found by
719 JASS with respect to the number of new association loci found MTAG across all the trait sets.
720 Each dot represents a set of traits. Dot colours represent the number of traits in the set. B)
721 Fraction of sets where the number of new association loci detected by JASS was superior to
722 the number of associations detected by MTAG stratified by the number of traits in the set.



723

724 **Figure 6. Comparison between clinical and data-driven trait sampling methods. (A, B)**
725 Distribution of the gain and the number of new association loci for trait sets selected by four
726 trait selection strategies from the validation data. *P*-values are from the two-sided Welch's *t*-
727 test. Differences in mean values in each pair compared in the test (right – left categories in
728 the order shown on the x-axis) are also shown. Note we used a pair of a training data and a
729 validation data to ensure independence for the test. The numbers under the labels on the x-
730 axis indicate the number of trait sets from each strategy. The observed JASS gain and the
731 number of new association loci are shown on the y-axis. (C,D) The observed gain and number
732 of new association loci detected for trait sets of the 'homogenous' category, visualised per
733 clinical grouping.



734