

The SumRank method efficiently detects valid pleiotropic variants using GWAS summary level data

Sander Lamballais^{1*}; Gennady V. Roshchupkin^{2,3*}; Raymond A. Poot⁴; Steven A. Kushner⁵; M. Arfan Ikram³; Hieab H. H. Adams^{1,3,6†}; Henning Tiemeier^{7,8†}.

¹ Department of Clinical Genetics, Erasmus MC University Medical Center, Rotterdam, the Netherlands

² Department of Epidemiology, Erasmus MC University Medical Center, Rotterdam, the Netherlands

³ Department of Radiology and Nuclear Medicine, Erasmus MC University Medical Center, Rotterdam, the Netherlands

⁴ Department of Cell Biology, Erasmus MC University Medical Center, Rotterdam, the Netherlands

⁵ Department of Psychiatry, Erasmus MC University Medical Center, Rotterdam, the Netherlands

⁶ Latin American Brain Health (BrainLat), Universidad Adolfo Ibáñez, Santiago, Chile

⁷ Department of Social and Behavioral Science, Harvard TH Chan School of Public Health, Boston, MA, USA

⁸ Department of Child and Adolescent Psychiatry, Erasmus MC-Sophia Children's Hospital, Rotterdam, the Netherlands

* Equal contribution

† Equal contribution

CORRESPONDENCE

Hieab H. H. Adams, MD, PhD. Departments of Clinical Genetics and of Radiology and Nuclear Medicine. Erasmus MC University Medical Center, Wytemaweg 80, 3015 CE, Rotterdam, the Netherlands. Telephone number: +31 10 70 33559. Fax Number: +31 10 70 43489. E-mail address: h.adams@erasmusmc.nl.

1 **ABSTRACT**

2 Genome-wide association studies (GWASs) have illustrated that pleiotropy is the norm rather
3 than the exception in the genetic architecture of complex traits. Previous methods have been
4 proposed to investigate cross-trait associations; however, these typically suffer from
5 substantially inflated false positive rates, are limited to associations between only two traits, or
6 examine only a subset of genome-wide variants. Here, we introduce SumRank, a novel method
7 for discovering cross-trait associations based on p-values from GWAS summary statistics.
8 Using numerical and genotype simulations, we demonstrated that SumRank had a near-zero
9 false positive rate, in contrast to contemporary meta-analytical approaches that yielded severely
10 inflated rates. We applied SumRank to GWAS summary data of eight psychiatric disorders and
11 identified 515 independent pleiotropic variants across 175 loci, of which 92 were novel and 71
12 had previously been associated with only one disorder. Overall, SumRank provides a valid and
13 efficient method for genome-wide discovery of pleiotropic variants that underlie a scalable
14 number of traits.

15 INTRODUCTION

16 Horizontal pleiotropy occurs when genetic variants are associated with multiple traits.
17 Pleiotropy has received increasing attention because it explains how a limited number of genes
18 map to numerous complex traits, allowing researchers to uncover novel genetic mechanisms
19 and therapeutic targets of disease. Recent studies have identified pleiotropic variants in loci
20 covering approximately 57% of the genome¹, thereby emphasizing the need for robust
21 frameworks and tools that further understanding. However, the relevance of pleiotropy to
22 human health and disease is still poorly understood.

23 The estimation of pleiotropic effects requires information on multiple traits and their genetic
24 architectures. This problem can be addressed through genome-wide association studies
25 (GWASs), which have been instrumental in uncovering the genetic architecture of numerous
26 traits². GWAS has been particularly insightful for traits that are highly polygenic. Multivariate
27 extensions of GWAS, such as vbBIMBAM³ and TATES⁴, can examine traits jointly within the
28 same population. However, these methods require individual-level data with information on all
29 relevant traits. In addition, sharing individual-level data is subject to privacy and informed
30 consent constraints, which limits accessibility to individual-level data by third parties. An
31 alternative approach is to study horizontal pleiotropy using GWAS summary statistics. Over
32 the past decade, summary-level results of thousands of GWASs have become widely available.
33 Summary-level methods have been proposed to estimate the probability that a genetic variant
34 is pleiotropic by performing calculations on the effect sizes or p -values across different
35 GWASs. Examples include Fisher's combination test (FCT), Cross-Phenotype Meta-Analysis
36 (CPMA)⁵, Association analysis based on subSETs (ASSET)⁶, and conjunctive FDR
37 (conjFDR)⁷.

38 However, summary-level methods have several limitations. First, horizontal pleiotropy is
39 defined as a genetic variant that has a non-zero effect on two or more traits under study.
40 However, summary-level methods based on meta-analytic approaches can falsely detect
41 variants as pleiotropic when they are associated only with a single trait, particularly when the
42 association is strong. This leads to a high false positive rate and thus an overestimation of the
43 number of pleiotropic variants. Second, GWASs provide results at the variant level for the
44 whole genome; however, summary-level methods may not have the same level of detail. For
45 example, the regional Bayes factor explores pleiotropy per independent genomic region rather
46 than per variant⁸, even though traits may be associated with the same region, but not with the

47 same causal variant. Third, genetic variants may be pleiotropic across any number of traits;
48 however, methods are typically designed to examine pleiotropy across two traits at a time.
49 Consequently, they can only detect pleiotropy at the pairwise level.

50 Here, we introduce SumRank, a method utilizing GWAS summary statistics that identifies
51 variant-level genome-wide pleiotropy without limitation for the number of traits being
52 examined simultaneously. Through numerical and genotype simulations, we showed that
53 SumRank detects pleiotropic variants with higher specificity than other methods. We also
54 showed that these findings hold under widely varying conditions. Finally, we demonstrated the
55 discovery power of SumRank by exploring pleiotropy across eight different psychiatric
56 disorders.

57 **RESULTS**

58 **Overview of the method**

59 SumRank is a combinatorial method for assessing genetic pleiotropy based on p -values of
60 GWAS summary statistics. The proposed method has several key features. First, it explicitly
61 tests the composite hypothesis that a variant is associated with all traits under study. Second,
62 the method transforms the p -values to ranks, thus mitigating the effect of extremely small p -
63 values in only one phenotype driving the detected effect. Third, SumRank can be applied to an
64 unlimited number of traits. Finally, applying SumRank across the genome requires very modest
65 computational power and memory and, more importantly, can easily be implemented in any
66 programming or scripting language that allows basic arithmetic operations. Below, we address
67 all these advantages.

68 The association between genetic variant g and trait Y_j can be modelled using the following
69 regression equation:

$$70 \mathbf{Y}_j = \beta_0 + \beta_{1,j}g + \gamma w + \varepsilon \quad (1)$$

71 where γw represents a non-genetic component and ε is an error term. A pleiotropic variant is
72 defined as one in which $\beta_{1,j} \neq 0$ for all j in $1 \dots m$ traits under consideration.

73 Given n variants and m traits, \mathbf{P} is an $n \times m$ matrix that contains the p -values from the
74 corresponding GWASs. SumRank uses a combinatorial approach to assess the pleiotropy of
75 variants based on \mathbf{P} . First, the p -values within each GWAS are sorted and ranked from smallest
76 to largest. If a variant is associated with a trait, its rank should be low. Second, the ranks for

77 each variant across the traits under study are summed, resulting in \mathbf{R} , a vector of length n . Under
78 the null hypothesis, low summed ranks are uncommon. Because pleiotropic variants are more
79 likely to yield small p-values in a GWAS, their summed ranks should be low. Therefore, the
80 third step is to determine the probability of \mathbf{R}_i or smaller for a given variant i (**Methods**):

$$81 \quad p_{sum,i} = \frac{m}{m!n^m} \sum_{k=1}^{R_i} \binom{k-1}{m-1} \quad (2)$$

82 Notably, psum can be approximated by summing the p-values directly within a given variant
83 (**Methods section**):

$$84 \quad p_{sum,i} = \frac{(\sum_{j=1}^m P_{i,j})^m}{m!} \quad (3)$$

85 The implications of SumRank are best understood when comparing the rejection regions with
86 FCT and CPMA (Fig. 1a). When a variant has a small p-value in both GWASs (bottom left,
87 Fig. 1a), the three methods yield correspondingly small p-values for pleiotropy. However, when
88 the variant is associated with only one of the two traits, FCT and CPMA can still return a small
89 p-value for pleiotropy regardless of the value of the other p-value. This will likely represent
90 genetic loci that are not truly pleiotropic, that is, false positive findings. Given that SumRank
91 relies on ranks rather than absolute p-values, it is robust to the influence of extremely small p-
92 values. Both p-values must be small for SumRank to return a small p-value for pleiotropy.

93 We further tested the implications of the rejection regions by simulating pairs of p-values. First,
94 we created 1 million pairs of p-values drawn from a uniform distribution, that is, under null
95 conditions. In this scenario, none of the methods inflated the false positive rates for the two
96 traits (Supplementary Fig. 1). Second, we created 1 million pairs where the first p-value was
97 always set to $5 \cdot 10^{-8}$ (i.e., genome-wide significance), but the second p-value was again drawn
98 from a uniform distribution. This represents a scenario where a variant is only associated with
99 the first trait, and therefore should not be classified as pleiotropic. We tested the methods
100 against a pleiotropy p-value threshold of $5 \cdot 10^{-8}$. Under these conditions, FCT and CPMA had
101 false positive rates of 4.81% and 0.80%, respectively, compared to 0.03% for SumRank (Fig.
102 1b). Crucially, the false positive rates for FCT and CPMA reached 100% when the first p-value
103 was set to $2.34 \cdot 10^{-9}$ and $4.07 \cdot 10^{-10}$, respectively (Fig. 1c). In contrast, the false positive rate of
104 SumRank remained 0.03% even when the first p-value was set to $5 \cdot 10^{-100}$.

105 **Performance in genotype simulations of two traits**

106 We conducted a series of population-based genotype simulations with varying genetic
107 architectures to evaluate the ability of SumRank and similar methods to detect pleiotropic
108 variants. To this end, we simulated correlated traits with a known genetic architecture using
109 data from the UK Biobank. From the UK Biobank genetic data, we selected independent
110 variants with a minor allele frequency above 5% using PLINK 1.9. We then randomly assigned
111 non-pleiotropic and pleiotropic causal effects to a subset of the independent variants, to
112 effectively create hypothetical traits. For the primary simulations, the causal effects were drawn
113 from a multivariate normal distribution. The heritability of the traits and the degree of genetic
114 correlation between traits were set manually. We obtained summary statistics for the simulated
115 traits through genome-wide analysis, which were used as input for SumRank and other
116 methods. The summary statistics resulting from these methods were clumped against genome-
117 wide significance ($5 \cdot 10^{-8}$) and $R > 0.2$, and the resulting lead and correlated variants were
118 assessed for the presence of pleiotropic and non-pleiotropic variants. As the primary test we
119 quantified the true pleiotropy rate, that is, the proportion of the variants marked as pleiotropic
120 where the true effect was indeed pleiotropic. An extensive description of the procedure is
121 provided in the Methods section.

122 For each simulation, we generated two traits for two independent cohorts of 100,000 individual
123 samples from the UK Biobank. As shown in [Fig. 2a](#), each trait had 1000 non-pleiotropic causal
124 variants that were unique to the traits (red) and 1000 pleiotropic causal variants that were shared
125 between them (green). Across 100 simulations, SumRank detected 21.1% (SD = 6.0%) of the
126 pleiotropic causal variants, which was lower than that of FCT (34.2%, SD = 13.1%) and CPMA
127 (32.4%, SD = 12.2%). However, SumRank labeled only 0.6% (SD = 0.2%) of the non-
128 pleiotropic causal variants as pleiotropic, whereas FCT and CPMA labeled 18.9% (SD = 1.0%)
129 and 17.8% (SD = 1.1%), respectively. In other words, SumRank had a true pleiotropy rate of
130 94.6% (SD = 1.6%) among the identified hits, compared with 47.6% (SD = 2.6%) for both FCT
131 and CPMA.

132 True pleiotropy rates depend on the underlying genetic architectures, that is, the number of
133 causal variants per trait and the overlap between them. Therefore, we reran the simulations
134 where each trait had 2000 causal variants in total but varied the overlap between the traits.
135 When the traits had a 25% overlap (i.e., 500 pleiotropic variants in total), the performance of
136 FCT and CPMA worsened, with the true pleiotropy rate dropping to 22.1% (SD = 1.1%) and

137 22.3% (SD = 1.4%), respectively (Fig 2b). In contrast, SumRank yielded a true pleiotropy rate
138 of 86.8% (SD = 4.6%). When the traits shared 75% of their variants (i.e., 1500 pleiotropic
139 variants in total), the relative performance of FCT and CPMA improved, with true pleiotropy
140 rates of 72.0% (SD = 1.8%) and 72.2% (SD = 1.8%), respectively (Fig 2c). Although SumRank
141 still reached a true pleiotropy rate of 98.3% (SD = 1.0%), its advantage over the other methods
142 was attenuated as SumRank identified fewer variants overall. Thus, the lower the ratio of
143 pleiotropic to non-pleiotropic causal variants, the better SumRank will perform relative to the
144 other methods.

145 As SumRank implicates fewer variants overall, it is possible that the methods prioritize the
146 same variants, but that SumRank is simply more conservative. To examine this post-hoc
147 hypothesis, we sorted the p-values for each method from lowest to highest and iteratively
148 calculated for the most significant variants what the true pleiotropy rate was, up to 1000 variants
149 (Fig. 3). SumRank had a 100% (SD = 0%) true pleiotropy rate when considering the top 1, 10,
150 and 100 variants. In contrast, FCT had a true pleiotropy rate of 79.5% (SD = 44.7%) when
151 considering the top variant, 57.5% (SD = 14.8%) in the top 10, and 54.3% (SD = 6.4%) in the
152 top 100. CPMA followed a similar pattern as FCT (top 1: 80.0% (44.7%); top 10: 58.0% (17.9
153 %); top 100: 54.8% (6.7%)). Thus, our post-hoc hypothesis can be rejected: the methods
154 prioritize vastly different variants, with SumRank prioritizing more true pleiotropic variants
155 than FCT and CPMA.

156 Next, we tested another post-hoc hypothesis: SumRank's apparent advantages over existing
157 methods might depend largely on the parameters used for the simulation. Therefore, we retested
158 the simulations by varying the values of different parameters. We first set the sample size to
159 50,000, 100,000, and 150,000 individuals (Supplementary Fig. 2). Larger sample sizes led to a
160 higher number of pleiotropic variants detected across all methods. The true pleiotropy rate for
161 SumRank decreased to 91.2% compared to 46.1% for FCT and 46.5% for CPMA. We then
162 varied the heritability of the traits to 0.25, 0.50, and 0.75 (Supplementary Fig. 3), where a higher
163 heritability led to more identified variants, and relatively poorer performance by FCT and
164 CPMA. We also changed the correlation between the effect sizes of the pleiotropic causal
165 variants to 0, 0.5, or 1 (Supplementary Fig. 4). Counterintuitively, as the correlation of the effect
166 sizes increased, the true pleiotropy rate for FCT decreased from 48.0% to 44.4%, and for CPMA
167 from 48.5% to 45.0%. In contrast, the true pleiotropy rate for SumRank increased from 94.3%
168 to 95.8%. Thus, we must also reject this post-hoc hypothesis.

169 **Performance in genotype simulations with more than two traits**

170 Our second hypothesis posited that SumRank can be simultaneously applied to more than two
171 traits. Thus, SumRank can explicitly test the composite hypothesis across any number of m
172 traits. To validate its use for multiple phenotypes, we tested the performance of SumRank using
173 different simulations. First, we considered the distribution of SumRank p -values when
174 performed on p -values drawn from a uniform distribution for 3, 5, 7, 10, 20, and 50 phenotypes,
175 where a pleiotropic variant is defined as affecting all phenotypes simultaneously
176 (Supplementary Fig 5). When the p -values for pleiotropy were tested against a threshold of
177 0.05, FCT and CPMA showed a false positive rate of approximately 5% across all conditions.
178 In contrast, the false positive rate of SumRank declined with the number of phenotypes, and it
179 was effectively 0% for 50 phenotypes. This is a direct consequence of the hypothesis being
180 tested; it is extremely unlikely that 50 randomly sampled p -values all suggest an association
181 between the variant and each trait.

182 As SumRank works with more than two traits, we simulated four traits to test whether the true
183 pleiotropy rate for SumRank was also better than that for FCT and CPMA. Note that pleiotropy
184 among the four traits can indicate any variant affecting two, three, or four traits. We sought to
185 explicitly test how well the methods detected variants that only affect all four traits
186 simultaneously. Therefore, we created three scenarios with varying numbers of pleiotropic
187 variants for each combination of traits. The first scenario is shown in Fig 4a, where we
188 simulated 1000 unique variants per trait (dark red), and 1000 pleiotropic variants that affected
189 all four traits. In this scenario, SumRank reached a true pleiotropy rate of 99.2% (SD = 0.4%),
190 with FCT and CPMA reaching 73.2% (SD = 1.0%) and 73.7% (SD = 1.0%), respectively. The
191 second scenario is shown in Fig 4b, where we simulated 1000 variants unique to each trait (dark
192 red), and 1000 pleiotropic variants for every combination of two (red), three (light red), or four
193 traits (green). Under such conditions, SumRank still primarily identified causal variants that
194 were pleiotropic for all four traits (79.6%, SD = 2.25%), and flagged few variants that were
195 related to three traits (17.5%), two traits (2.4%), or even one trait (0.5%). In contrast, FCT and
196 CPMA reached a true pleiotropy rates of 40.1% (SD = 1.3%) and 40.9% (SD = 1.3%),
197 respectively. Crucially, 9.0% and 8.4% of the variants that FCT and CPMA identified as related
198 to all four traits were simulated to be non-pleiotropic. This pattern of findings was most evident
199 in the scenario where most variants were not pleiotropic and only 100 variants were pleiotropic
200 for all four traits (Fig 4c). Of the loci that the methods found to be related to all traits, SumRank

201 was correct for 56% of the loci (SD = 7.8%), compared to 10.9% (SD = 1.3%) and 11.1% (SD
202 = 1.3%) for FCT and CPMA, respectively. We argue that these results support the notion that
203 SumRank is more robust to false positives when studying more than two traits.

204 **Example study: pleiotropy across eight psychiatric disorders**

205 To test the utility of SumRank with non-simulated data, we mirrored the design of the largest
206 pleiotropy study to date across eight prominent psychiatric disorders ([Supplementary Table 1](#))⁹:
207 anorexia nervosa, attention-deficit/hyperactivity disorder, autism spectrum disorder, bipolar
208 disorder, major depression, obsessive-compulsive disorder, schizophrenia, and Tourette
209 syndrome. Every possible combination of at least two traits (= 247) was tested against a
210 significance threshold of $(5 \cdot 10^{-8} / 247) = 2.02 \cdot 10^{-10}$. All calculations took approximately two
211 minutes on a single core.

212 Overall, we identified 515 independent variants across 175 pleiotropic loci ([Fig 5a](#)). Compared
213 to the underlying GWASs, 92 of these loci had previously not been associated with any trait,
214 71 loci with just one of the traits, and 12 loci were genome-wide significant in two or three of
215 the GWASs. The independent variants were primarily located in the introns of protein-coding
216 genes (47.7%), intergenic sequences (32.3%), or introns of non-coding genes (13.3%) ([Fig 5b](#)).
217 Most independent variants were associated with two traits (n = 198); however, surprisingly we
218 also identified large groups of independent variants related to three traits (n = 144), four traits
219 (n = 101), and five or more traits (n = 28) ([Fig. 5c](#)). The largest group was related to bipolar
220 disorder and schizophrenia (n = 117), followed by variants related to bipolar disorder,
221 schizophrenia, and major depressive disorder (n = 81) ([Fig. 6](#)).

222 **DISCUSSION**

223 SumRank overcomes several shortcomings of the existing pleiotropy methods. Most
224 importantly, as it assesses the composite hypothesis of whether individual variants associate
225 with all traits, SumRank is not affected by a strong association with only one or a subset of the
226 traits – which is reflected in its superior rate of true pleiotropic findings compared to other
227 methods. In addition, it is robust, can be applied to any number of variants across the genome,
228 and is readily computed as it relies on a simple algebraic formula.

229 The findings illustrate that the study of pleiotropy requires a delicate balancing act. Pleiotropy
230 can be studied by only considering genome-wide significant hits¹, but this approach fails to

231 identify loci that did not reach genome-wide significance in all GWASs involved. Summary-
232 level approaches can reveal such loci, but as is clear from our simulations this can come at the
233 cost of an inflated false positive rate and thus having a higher rate of irrelevant loci, particularly
234 when considering more than two traits. SumRank is a conservative approach that may yield
235 fewer pleiotropic loci than other methods, but the loci that it does return have a much higher
236 likelihood of being truly pleiotropic. Thus, SumRank strikes a balance that is more suitable for
237 functional genomic and mechanistic biological follow-up studies of pleiotropic loci.

238 Composite hypotheses have been used before in the study of cross-trait associations, for
239 example in conjFDR⁷, the regional Bayes factor⁸, and PLACO¹⁰. However, these approaches
240 typically focus on pairs of traits. Through SumRank, we showed that among psychiatric
241 disorders there is likely a significant amount of pleiotropy that extends beyond two traits, with
242 most of the identified loci relating to three or more traits. This is in line with previous studies
243 that focused only on genome-wide significant variants¹. As more GWASs are conducted and
244 their sample sizes increase, the relevance of multi-trait loci will become more apparent.
245 Methods like SumRank are needed to properly model and disentangle such loci.

246 We tested SumRank extensively under various simulation conditions and demonstrated that it
247 is a robust and efficient method. However, this method also has several shortcomings and areas
248 for improvement. SumRank relies purely on GWAS p-values and cannot incorporate any
249 information such as effect sizes, genetic correlations^{11,12}, omics data, or other external
250 information. Nevertheless, the identification of loci can be sufficient to uncover relevant genes
251 and accompanying gene pathways and physiological mechanisms. Furthermore, although
252 untested in the current study, SumRank may be able to incorporate functional data such as eQTL
253 data if such data also have accompanying p-values. Another limitation of SumRank is that like
254 most summary-level methods, it does not distinguish between the different types of pleiotropy,
255 that is, vertical and horizontal pleiotropy. Although variants may be marked as pleiotropic by
256 any of the GWAS-based methods, we emphasize that further work is needed to elucidate the
257 underlying mechanisms.

258 In conclusion, SumRank efficiently and robustly assesses the pleiotropy of genetic variants
259 based on p-values from GWAS. Compared to other summary-level methods, it provides
260 stringent control for false positives by using a composite hypothesis and allows for an unlimited
261 number of traits to be studied with high computational efficiency. We believe that SumRank

262 can play a role in identifying multi-trait pleiotropic loci with high confidence, which in turn
263 will prove invaluable for further elucidating the relevance of pleiotropy in health and disease.

264 METHODS

265 Theory

266 The SumRank method is based on the idea that pleiotropy can be assessed from GWAS
267 summary statistics by ranking the p-values from the smallest (first rank) to the largest
268 (maximum rank), and subsequently summing the ranks per variant. Considering m traits with
269 N variants, \mathbf{P} represents an $N \times m$ matrix, where each column contains the ranks of the p-values
270 within each trait. If the null hypothesis of no association is based on a uniform distribution, then
271 the probability that rank $P_{i,j}$ (variant i and trait j) is equal to any rank is $1/N$. The ranks can
272 subsequently be summed per variant to create \mathbf{R} , which is a vector of length N with $m \leq R_i \leq$
273 mN .

274 SumRank tests the likelihood of \mathbf{R}_i being equal to rank r or more extreme under a null
275 hypothesis of randomly assigned ranks. Consider a set of vectors x_1, \dots, x_m such that $x_1 + \dots +$
276 $x_m = \mathbf{R}_{sum,i}$ with $1 \leq x_i \leq N$. Let \mathcal{S} be a set of solutions with $x_i \geq 1$ and let \mathcal{S}_i be the set of solutions
277 with $x_i < N$ for exactly that i , then:

$$278 \quad |T| = |\overline{\mathcal{S}}_1 \cap \overline{\mathcal{S}}_2 \dots \overline{\mathcal{S}}_m| \quad (4)$$

279 Based on the inclusion/exclusion criteria we can rewrite eq. (3) as:

$$280 \quad |T| = |S| - \sum_{1 \leq i \leq m} |S_i| + \sum_{1 \leq i < j \leq m} |S_i \cap S_j| + \dots + (-1)^m |S_i \cap S_j \dots S_m| \quad (5)$$

281 where $|S|$ is a combinatorial composition which equals $\binom{r-1}{m-1}$, or in other words, how many
282 combinations of $m-1$ positive integer numbers can give us a sum equal to $r-1$. Therefore, $|S_i| =$
283 $\binom{r-N-1}{m-1}$, and $|S_i \cap S_j| = \binom{r-2N-1}{m-1}$, etc., and:

$$284 \quad \sum_{1 \leq i \leq m} |S_i| = \binom{r-N-1}{m-1} \binom{m}{1} \quad (6)$$

$$285 \quad \sum_{1 \leq i < j \leq m} |S_i \cap S_j| = \binom{r-N-1}{m-1} \binom{m}{2} \quad (7)$$

286 etc. It is now possible to simplify eq. (4):

$$287 \quad |T| = \sum_{k=0}^d (-1)^k \binom{r-kN-1}{m-1} \binom{m}{k} \quad (8)$$

288 where:

$$289 \quad d = \left\lfloor \frac{(r-m)}{N} \right\rfloor \quad (9)$$

290 We can further divide $|T|$ in eq. (7) by the total number of vectors of length m – as they arise
291 with equal probability under the null hypothesis – to derive:

$$292 \quad P(R_i = r) = \frac{1}{N^m} \sum_{k=0}^d (-1)^k \binom{r-kN-1}{m-1} \binom{m}{k} \quad (10)$$

293 Finally, we compute the p-value for SumRank:

$$294 \quad P(R_i \leq r) = \frac{1}{N^m} \sum_{i=1}^r \sum_{k=0}^d (-1)^k \binom{r-kN-1}{m-1} \binom{m}{k} \quad (11)$$

295 In practice, we want to test the null hypothesis when R_i is much smaller than N , and thus d will
296 always be approximately zero. We can then simplify eq. (11):

$$297 \quad P(R_i \leq r) = \frac{1}{N^m} \sum_{i=1}^r \binom{i-1}{m-1} \quad (12)$$

$$298 \quad = \frac{m}{m!N^m} \sum_{i=1}^r \frac{1}{i} \frac{i!}{(i-m)!} \quad (13)$$

$$299 \quad = \frac{m}{m!N^m} \sum_{i=1}^r (1-m+1) \dots (i-1) \quad (14)$$

300 Further simplification yields eq. (3), shown in the Results section:

$$301 \quad p_{sum,i} = \frac{(\sum_{j=1}^m P_{i,j})^m}{m!}$$

302 Genotypic data

303 We used release version 3 for the imputed genotype datasets from the UK Biobank. The
304 procedures for genotyping, imputation, and quality control have been described elsewhere¹¹. In
305 brief, blood samples from 488,377 participants were assayed with two different genotyping
306 arrays: the Applied Biosystems UK BiLEVE Axiom Array by Affymetrix and the Applied
307 BioSystems UKB Axiom Array. After quality control of the markers and samples, the
308 autosomes were phased with SHAPEIT3¹³, using the 1000 Genomes phase 3 dataset as a
309 reference¹⁴, and subsequently imputed against the Haplotype Reference Consortium (HRC)
310 dataset¹⁵. The data were also separately imputed against a combination of the 1000 Genomes
311 phase 3 and the UK10K datasets¹⁶, and all data were combined – using the HRC imputation
312 when a variant was available in both – into a dataset consisting of 93,095,623 autosomal
313 variants in 478,442 individuals.

314 **Genotype simulations**

315 The genotype simulations were conducted in a series of steps. First, we identified independent
316 variants based on the full genetic dataset of the UK Biobank. We excluded individuals with a
317 call rate of < 0.99 ; and we excluded variants with any missingness, a violation of the Hardy-
318 Weinberg equilibrium ($p < 10^{-6}$), and less or more than two alleles. To obtain the independent
319 variants we used PLINK2¹⁷ with `--indep-pairwise 50 5 0.05`.

320 Second, we created 20 simulations where we randomly selected a subset of the independent
321 variants that we assigned effect sizes for the traits that we aimed to simulate. In the case of two
322 traits, we selected 1000 causal variants that only affected the first trait, 1000 causal variants that
323 only affected the second trait, and 1000 causal variants that were pleiotropic and thus affected
324 both traits. The effect sizes were drawn from a normal distribution $N(0, 1)$, and we set the
325 Pearson's correlation between the effect sizes within a pleiotropic variant to 0.5.

326 Third, for each simulation we selected a subset of 100,000 individuals per trait, and we
327 calculated the simulated traits based on the causal variants. The assigned trait value was set as
328 the sum of the genetic component scaled by the heritability, and a complementary non-genetic
329 component:

$$330 \quad \mathbf{Y} = \sqrt{h^2}\mathbf{G} + N(0, 1 - h^2) \quad (15)$$

331 The Pearson's correlation between the traits was forced to 0.5, and the h^2 for both traits was set
332 to 0.5.

333 Fourth, we conducted GWASs for the simulated traits and applied the pleiotropy methods. The
334 GWASs were conducted with fastGWA¹⁸. Each GWAS was run on 100,000 individuals. The
335 samples were non-overlapping, as overlap between the samples can lead to bias in the
336 estimates¹⁹, and correcting for this bias is beyond the scope of the current study. For the
337 pleiotropy assessments we focused on SumRank, FCT, and CPMA.

338 Finally, to see whether the methods identified the non-pleiotropic or pleiotropic causal variants,
339 we clumped the results using PLINK2¹⁷ and assessed whether any causal variants were in
340 linkage disequilibrium (LD) with the lead variants. We based the LD structure for clumping on
341 the total sample that partook in the simulated GWASs. We set independent variants to be those
342 that were genome-wide significant ($p < 5 \cdot 10^{-8}$) that were not in LD after clumping ($r^2 < 0.6$). A
343 second round of clumping was performed against $r^2 < 0.1$ to define the lead variants. For each
344 lead variant, we checked whether any causal variants in LD were pleiotropic or non-pleiotropic.

345 The simulations relied on several set parameters and modeling decisions. We performed
346 multiple sensitivity analyses to understand how the behavior of the methods changed. For the
347 sample size we reran the simulations with 50,000 and 150,000 individuals per GWAS. For
348 heritability we also simulated h^2 being equal to 0.25 and 0.75. Finally, we changed the Pearson's
349 correlation of the effect sizes within pleiotropic variants to 0.0 and 1.0.

350 **Application to psychiatric disorders**

351 For the psychiatric disorders we used data from the GWASs performed as part of the Psychiatric
352 Genomics Consortium (PGC) ([Supplementary Table 1](#)). The GWASs were focused on anorexia
353 nervosa²⁰, attention-deficit/hyperactivity disorder²¹, autism spectrum disorder²², bipolar
354 disorder²³, major depressive disorder²⁴, obsessive-compulsive disorder²⁵, schizophrenia²⁶, and
355 Tourette syndrome²⁷.

356 To assess pleiotropy, we applied SumRank to all possible combinations of the GWASs, i.e.
357 247. For each variant we identified the combination with the lowest p-value. The genome-wide
358 significant threshold was set to $5 \cdot 10^{-8} / 247 = 2.02 \cdot 10^{-10}$. The results were further processed in
359 FUMA to identify the loci and obtain information on the functional consequences²⁸.

360 **ACKNOWLEDGEMENTS**

361 This research has been conducted using the UK Biobank Resource under application number
362 23509. This work was supported by the Netherlands Organization for Scientific Research
363 (NWO/ZonMW grant 016.VICI.170.200 [SL, HT], NWO VENI 916.19.151 [SL, HA]). We
364 thank SURFsara (www.surfsara.nl) for their support in using the Dutch national supercomputer
365 Cartesius.

366 **CONFLICT OF INTEREST**

367 None declared.

368 **DATA AVAILABILITY**

369 All summary statistics from the PGC can be obtained from
370 <https://www.med.unc.edu/pgc/download-results/>. The raw UK Biobank data can be obtained
371 through a data access application available at <https://www.ukbiobank.ac.uk>.

372 **CODE AVAILABILITY**

373 All code is freely available. The code for SumRank as well as FCT and CPMA are available
374 through the “sumrank” R package (<https://github.com/slamballais/sumrank>). The code used to
375 download the summary statistics, run the simulations, and perform the analyses can be found
376 in a separate repository (https://github.com/slamballais/sumrank_paper).

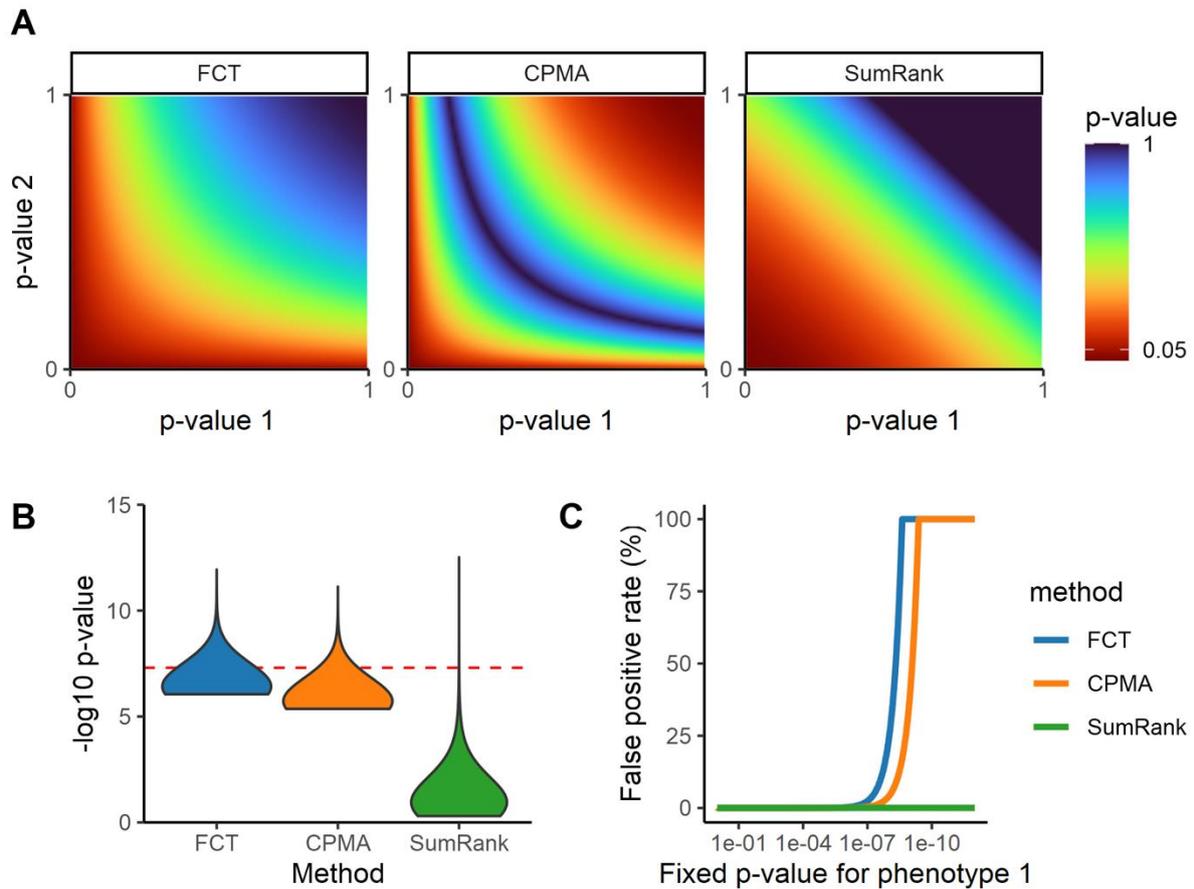
377 REFERENCES

- 378 1. Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex
379 traits. *Nat Genet* **51**, 1339-1348 (2019).
- 380 2. Visscher, P.M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation.
381 *Am J Hum Genet* **101**, 5-22 (2017).
- 382 3. Stephens, M. A unified framework for association analysis with multiple related
383 phenotypes. *PLoS One* **8**, e65245 (2013).
- 384 4. van der Sluis, S., Posthuma, D. & Dolan, C.V. TATES: efficient multivariate genotype-
385 phenotype analysis for genome-wide association studies. *PLoS Genet* **9**, e1003235
386 (2013).
- 387 5. Cotsapas, C. *et al.* Pervasive sharing of genetic effects in autoimmune disease. *PLoS*
388 *Genet* **7**, e1002254 (2011).
- 389 6. Bhattacharjee, S. *et al.* A subset-based approach improves power and interpretation for
390 the combined analysis of genetic association studies of heterogeneous traits. *Am J Hum*
391 *Genet* **90**, 821-35 (2012).
- 392 7. Andreassen, O.A., Thompson, W.K. & Dale, A.M. Boosting the power of schizophrenia
393 genetics by leveraging new statistical tools. *Schizophr Bull* **40**, 13-7 (2014).
- 394 8. Pickrell, J.K. *et al.* Detection and interpretation of shared genetic influences on 42
395 human traits. *Nat Genet* **48**, 709-17 (2016).
- 396 9. Cross-Disorder Group of the Psychiatric Genomics Consortium. Electronic address,
397 p.m.h.e. & Cross-Disorder Group of the Psychiatric Genomics, C. Genomic
398 Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric
399 Disorders. *Cell* **179**, 1469-1482 e11 (2019).
- 400 10. Ray, D. & Chatterjee, N. A powerful method for pleiotropic analysis under composite
401 null hypothesis identifies novel shared loci between Type 2 Diabetes and Prostate
402 Cancer. *PLoS Genet* **16**, e1009218 (2020).
- 403 11. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data.
404 *Nature* **562**, 203-209 (2018).
- 405 12. Lee, C.H., Shi, H., Pasaniuc, B., Eskin, E. & Han, B. PLEIO: a method to map and
406 interpret pleiotropic loci with GWAS summary statistics. *Am J Hum Genet* **108**, 36-48
407 (2021).
- 408 13. O'Connell, J. *et al.* Haplotype estimation for biobank-scale data sets. *Nat Genet* **48**, 817-
409 20 (2016).
- 410 14. 1000 Genomes Project Consortium *et al.* A global reference for human genetic
411 variation. *Nature* **526**, 68-74 (2015).
- 412 15. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation.
413 *Nat Genet* **48**, 1279-83 (2016).
- 414 16. Huang, J. *et al.* Improved imputation of low-frequency and rare variants using the
415 UK10K haplotype reference panel. *Nat Commun* **6**, 8111 (2015).
- 416 17. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer
417 datasets. *Gigascience* **4**, 7 (2015).
- 418 18. Jiang, L. *et al.* A resource-efficient tool for mixed model association analysis of large-
419 scale data. *Nat Genet* **51**, 1749-1755 (2019).
- 420 19. LeBlanc, M. *et al.* A correction for sample overlap in genome-wide association studies
421 in a polygenic pleiotropy-informed framework. *BMC Genomics* **19**, 494 (2018).
- 422 20. Watson, H.J. *et al.* Genome-wide association study identifies eight risk loci and
423 implicates metabo-psychiatric origins for anorexia nervosa. *Nat Genet* **51**, 1207-1214
424 (2019).

- 425 21. Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for attention
426 deficit/hyperactivity disorder. *Nat Genet* **51**, 63-75 (2019).
- 427 22. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum
428 disorder. *Nat Genet* **51**, 431-444 (2019).
- 429 23. Mullins, N. *et al.* Genome-wide association study of more than 40,000 bipolar disorder
430 cases provides new insights into the underlying biology. *Nat Genet* **53**, 817-829 (2021).
- 431 24. Howard, D.M. *et al.* Genome-wide meta-analysis of depression identifies 102
432 independent variants and highlights the importance of the prefrontal brain regions. *Nat*
433 *Neurosci* **22**, 343-352 (2019).
- 434 25. International Obsessive Compulsive Disorder Foundation Genetics, C. & Studies,
435 O.C.D.C.G.A. Revealing the complex genetic architecture of obsessive-compulsive
436 disorder using meta-analysis. *Mol Psychiatry* **23**, 1181-1188 (2018).
- 437 26. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in
438 schizophrenia. *Nature* (2022).
- 439 27. Yu, D. *et al.* Interrogating the Genetic Determinants of Tourette's Syndrome and Other
440 Tic Disorders Through Genome-Wide Association Studies. *Am J Psychiatry* **176**, 217-
441 227 (2019).
- 442 28. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping
443 and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826 (2017).

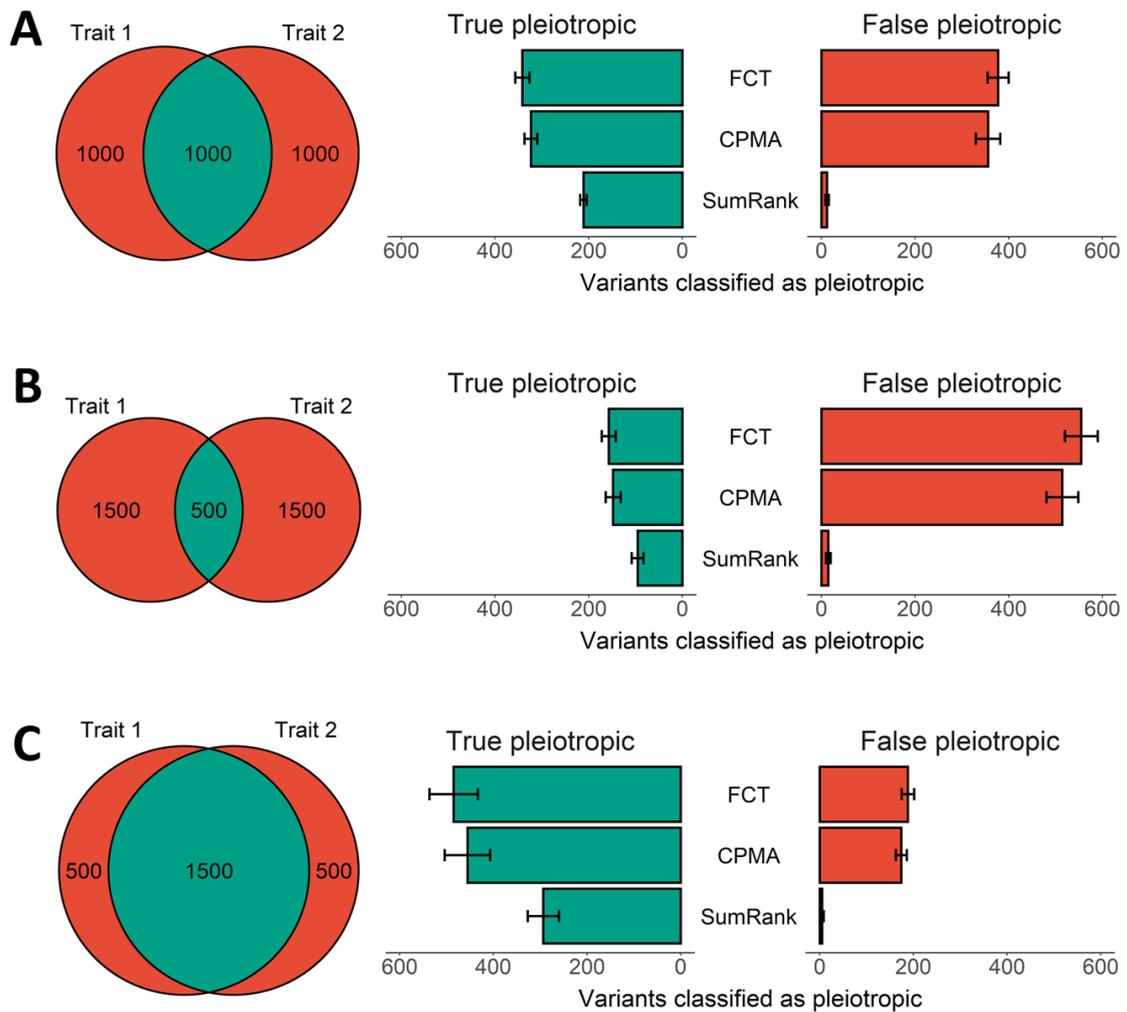
444

445 **Figure 1. Performance of FCT, CPMA, and SumRank under theoretical conditions.** (A)
446 Rejection region plots for the different methods given two p-values. (B) Violin plot for the
447 different methods under the condition where the methods are applied to a p-value of $5 \cdot 10^{-8}$ and
448 uniformly sampled p-values. The y-axis indicates the p-value derived from the methods. (C)
449 Similar to panel B, but the fixed p-value is varied across a range (x-axis) and plotted against
450 the proportion of tests that yield statistically significant results (y-axis). CPMA = Cross-
451 Phenotype Meta-Analysis; FCT = Fisher's combination test.



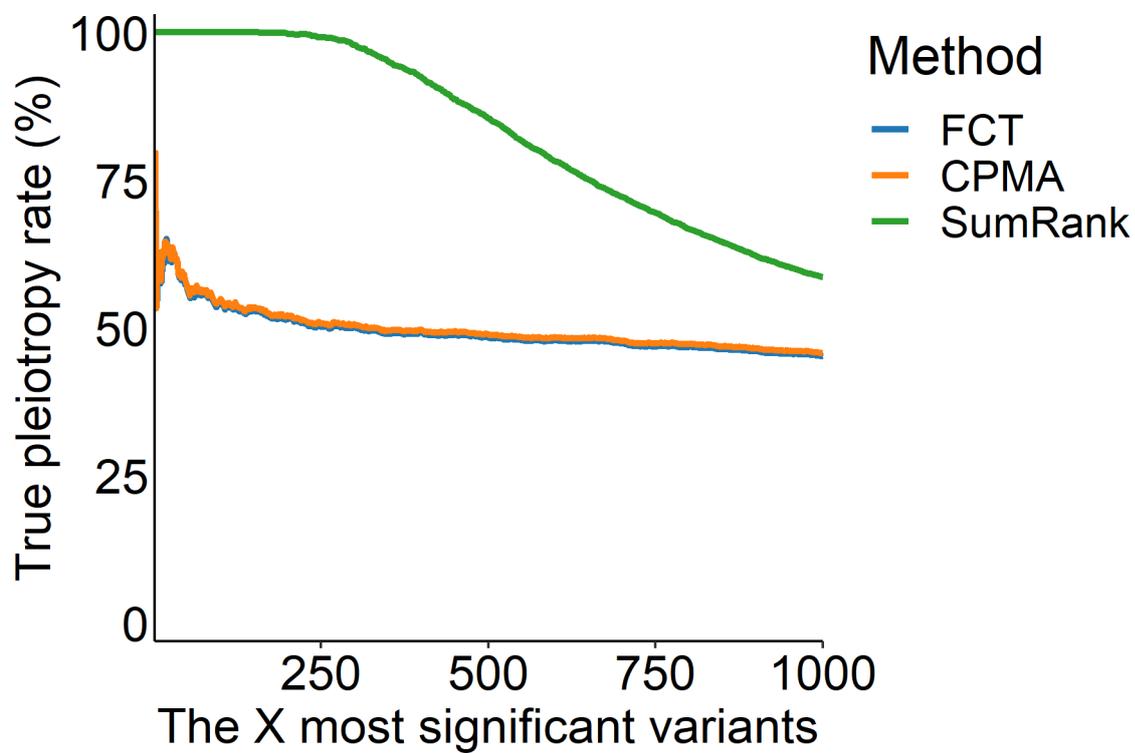
452

453 **Figure 2. Results for FCT, CPMA, and SumRank for the genotype simulations of two**
 454 **traits, across 100 simulations, for various genetic architectures.** Each trait was simulated to
 455 have 2000 causal variants, with each condition varying in the degree of overlap: (A) 50%, (B)
 456 25%, (C) 75%. The left columns display the number of shared causal variants (green) and
 457 unique causal variants (red) for various simulation conditions. The right column contains bar
 458 plots of the variants that each method classified as pleiotropic, and therefore an overview of the
 459 number of correct and incorrect findings.



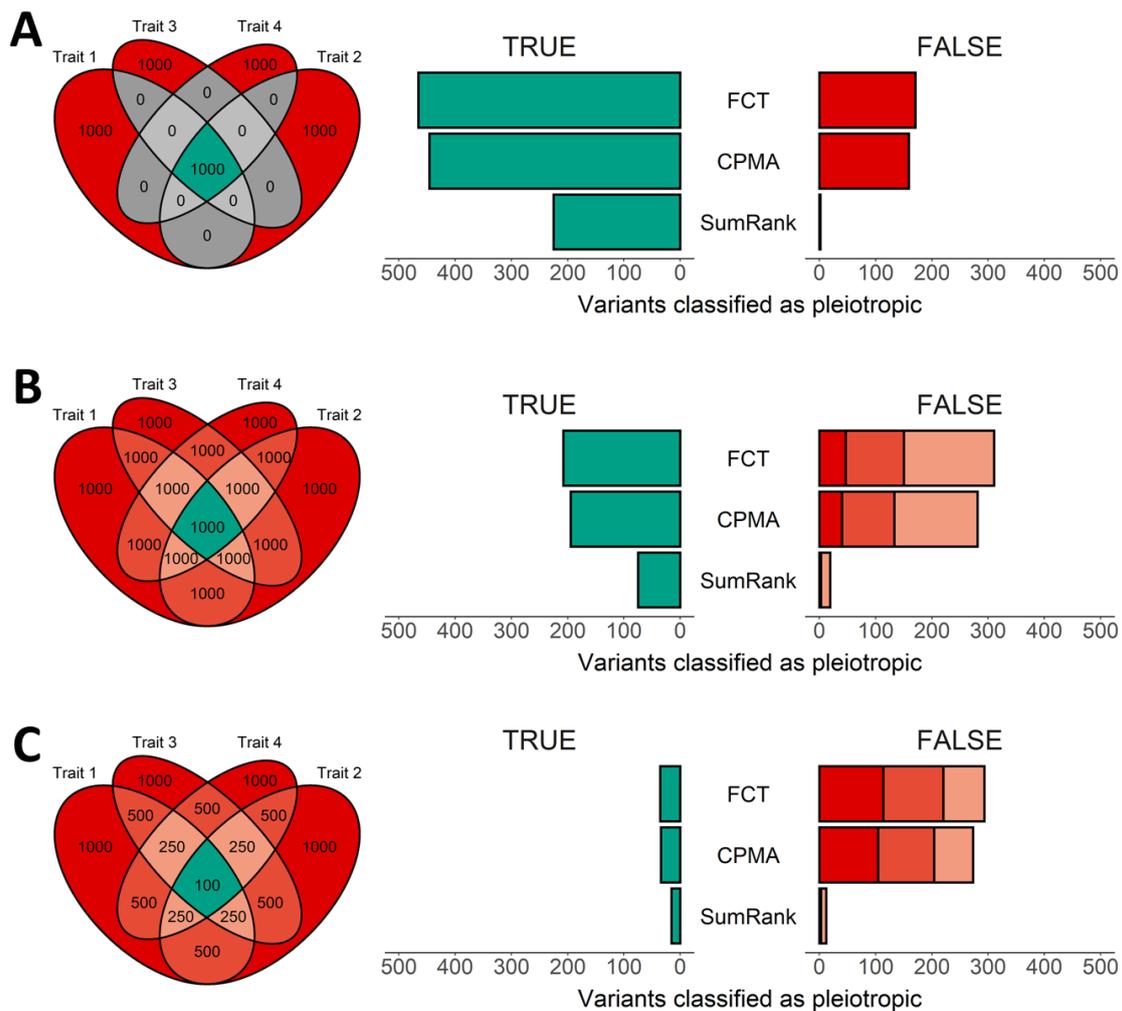
460

461 **Figure 3. True pleiotropy rate for the most significant hits per method.** The x-axis
462 represents the p-values, sorted from lowest to highest. The y-axis represents the true pleiotropy
463 rate for the top variants up to that point.



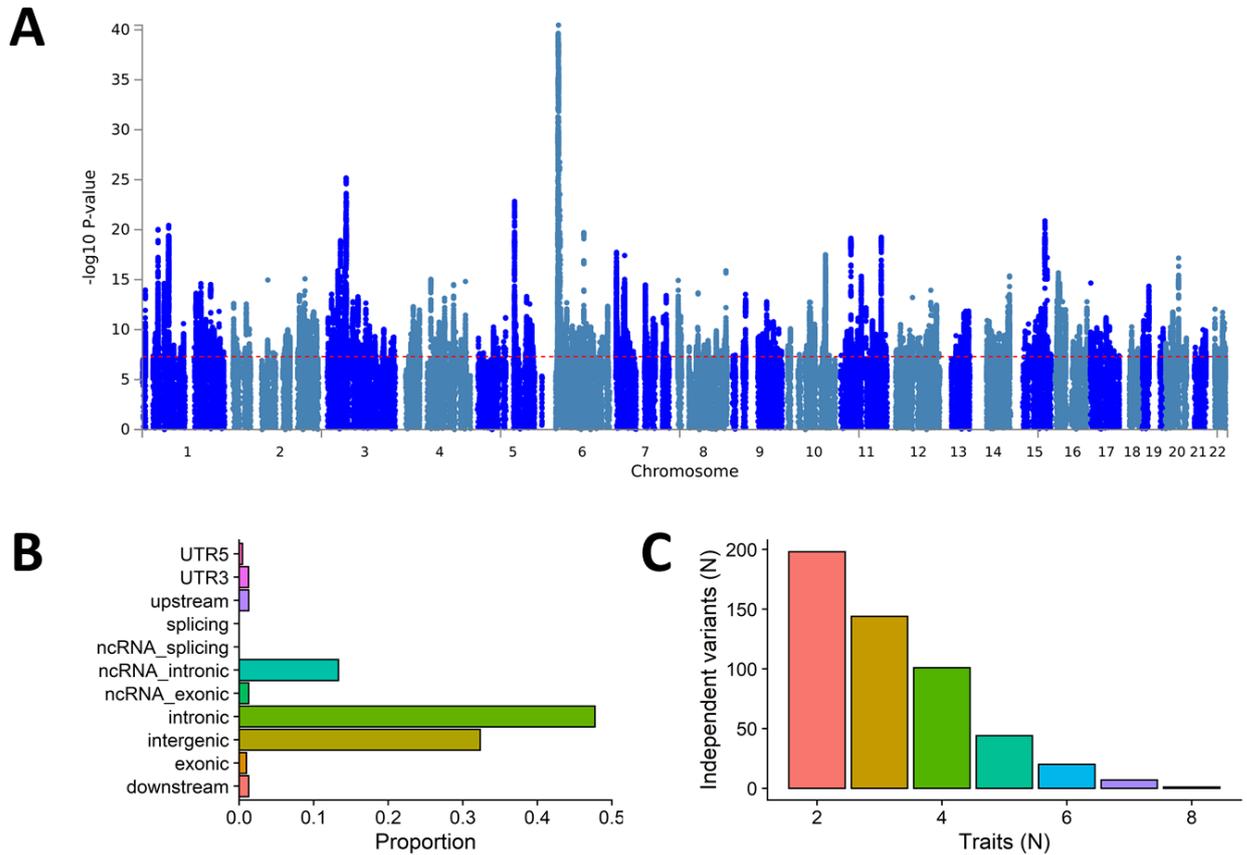
464

465 **Figure 4. Results for FCT, CPMA, and SumRank for the genotype simulations of four**
 466 **traits, across 100 simulations, for various genetic architectures.** The left column displays
 467 the genetic architectures. The right column contains bar plots of the variants that each method
 468 classified as pleiotropic for all four traits, and therefore an overview of the number of correct
 469 and incorrect findings. The conditions mimic settings where (A) each possible combination of
 470 traits has an equal number of causal variants, (B) variants either only influence one or four
 471 traits, (C) the number of causal variants that influence more traits decreases.



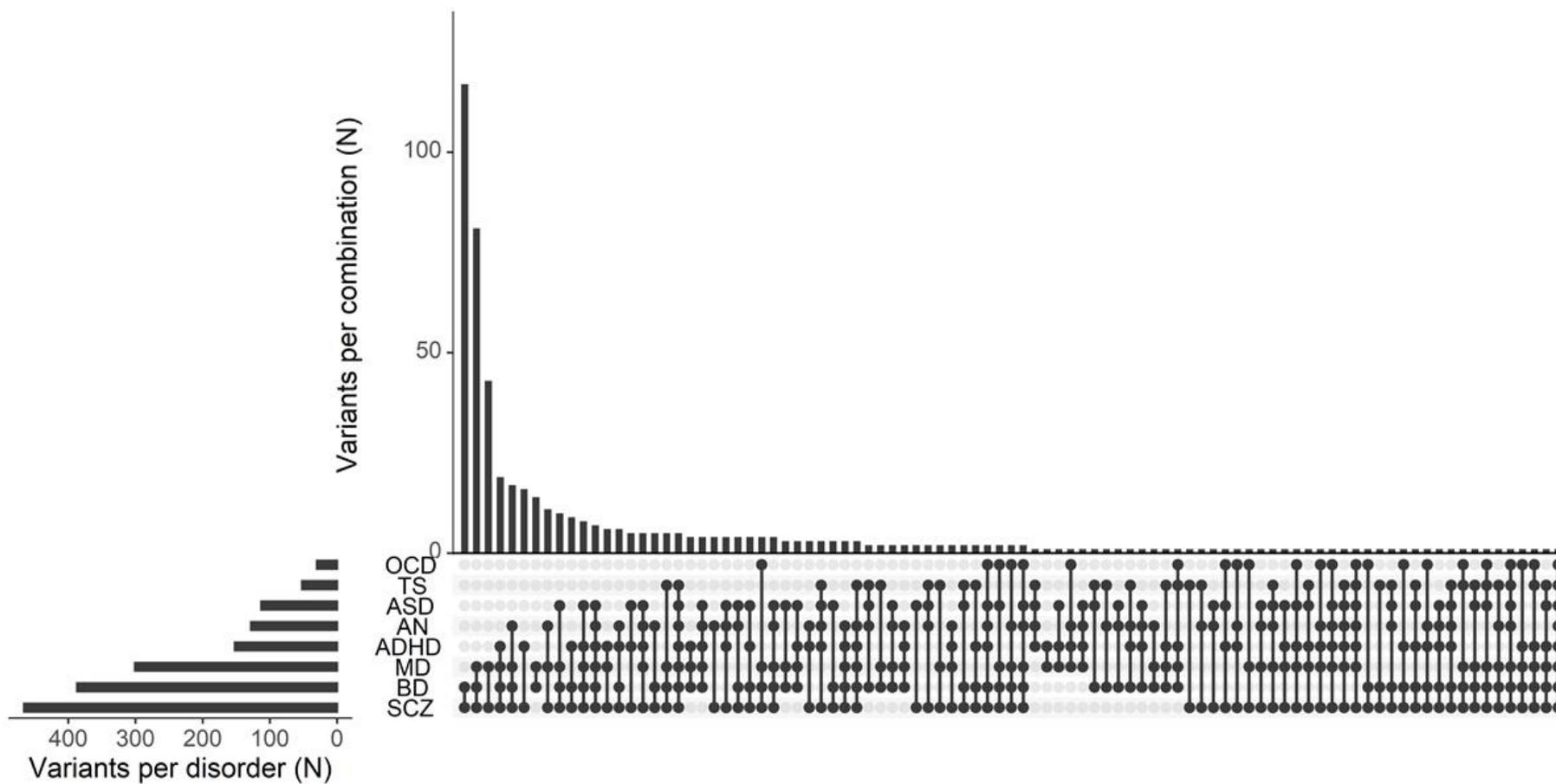
472

473 **Figure 5. Results from SumRank on pleiotropy across eight psychiatric disorders.** All
474 combinations between the eight disorders were tested, totaling 247 combinations. (A) A
475 Manhattan plot where each variant is represented by the lowest p-value across the 247
476 combinations. (B) Functional consequences of the pleiotropic variants. (C) The number of traits
477 that independent variants associated with.



478

479 **Figure 6. Upset plot of the SumRank pleiotropy results.** ADHD = attention deficit hyperactivity disorder; AN = anorexia nervosa; ASD =
 480 autism spectrum disorder; BD = bipolar disorder; MD = Major depressive disorder; OCD = obsessive-compulsive disorder; SCZ =
 481 schizophrenia; TS = Tourette syndrome.



482