

Roelfs et al. | Genetics of profiles relevant to mental health

1 **Phenotypically independent profiles relevant to mental health are genetically correlated**

2 Daniel Roelfs^{1,*}, MSc, Dag Alnæs^{1,2}, PhD, Oleksandr Frei¹, PhD, Dennis van der Meer^{1,3}, PhD, Olav B.
3 Smeland¹, PhD, Ole A. Andreassen¹, PhD, Lars T. Westlye^{1,4}, PhD, Tobias Kaufmann^{1,5*}, PhD

4
5 ¹ *NORMENT, KG Jebsen Centre for Neurodevelopmental Disorders, Division of Mental Health and
6 Addiction, Oslo University Hospital & Institute of Clinical Medicine, University of Oslo, Oslo, Norway*

7 ² *Bjørknes College, Oslo, Norway*

8 ³ *School of Mental Health and Neuroscience, Faculty of Health, Medicine and Life Sciences, Maastricht
9 University, Maastricht, The Netherlands*

10 ⁴ *Department of Psychology, University of Oslo, Oslo, Norway*

11 ⁵ *Department of Psychiatry and Psychotherapy, University of Tübingen, Tübingen, Germany*

12

13 * Correspondence: Daniel Roelfs & Tobias Kaufmann, Ph.D.

14 Email: daniel.roelfs@medisin.uio.no, tobias.kaufmann@medisin.uio.no

15 Postal address: OUS, PO Box 4956 Nydalen, 0424 Oslo, Norway

16 Telephone: +47 23 02 73 50, Fax: +47 23 02 73 33

17

18 Counts: Main: 4088 words | Figures: 4 | References: 47

19 **Abstract**

20 Genome-wide association studies (GWAS) and family-based studies have revealed partly overlapping
21 genetic architectures between various psychiatric disorders. Given clinical overlap between disorders, our
22 knowledge of the genetic architectures underlying specific symptom profiles and risk factors is limited.
23 Here, we aimed to derive distinct profiles relevant to mental health in healthy individuals and to study
24 how these genetically relate to each other and to common psychiatric disorders. Using independent
25 component analysis, we decomposed self-report mental health questionnaires from 136,678 healthy
26 individuals of the UK Biobank, excluding data from individuals with a diagnosed neurological or
27 psychiatric disorder, into thirteen distinct profiles relevant to mental health, capturing different symptoms
28 as well as social and risk factors underlying reduced mental health. Utilizing genotypes from 117,611 of
29 those individuals with White English ancestry, we performed GWAS for each mental health profile and
30 assessed genetic correlations between these profiles, and between the profiles and common psychiatric
31 disorders and cognitive traits. We found that mental health profiles were genetically correlated with a
32 wide range of psychiatric disorders and cognitive traits, with strongest effects typically observed between
33 a given mental health profile and a disorder for which the profile is common (e.g. depression symptoms
34 and major depressive disorder, psychosis and schizophrenia). Strikingly, although the profiles were
35 phenotypically uncorrelated, many of them were genetically correlated with each other. This study
36 provides evidence that statistically independent mental health profiles partly share genetic underpinnings
37 and show genetic overlaps with psychiatric disorders, suggesting that shared genetics across psychiatric
38 disorders cannot be exclusively attributed to the known overlapping symptomatology between the
39 disorders.

40 **Introduction**

41 Psychiatric disorders are highly polygenic, exhibiting a multitude of significantly associated genetic
42 variants with small effect sizes. Recent large-scale genome-wide association studies (GWAS) have
43 identified a large number of single-nucleotide polymorphisms (SNP) associated with psychiatric disorders
44 such as schizophrenia (SCZ)¹, bipolar disorder (BD)², major depression (MD)³, attention deficit
45 hyperactivity disorder (ADHD)⁴, autism spectrum disorders (ASD)⁵, post-traumatic stress disorder
46 (PTSD)⁶, and anxiety (ANX)⁷. In addition to substantial polygenicity, previous findings have documented
47 genetic overlap between disorders⁸⁻¹¹, even in the absence of genetic correlations expressed as additive
48 genetic effects for two traits, as recently demonstrated for schizophrenia and educational attainment^{12,13}.
49 Adding to the complexity, psychiatric disorders also overlap with multiple complex traits, such as BMI¹⁴,
50 cardio-metabolic diseases¹⁵ and a number of psychosocial and other risk factors for reduced mental
51 health¹⁶⁻¹⁸. The latter are particularly challenging in the context of genetics, since genetic overlap may not
52 necessarily point to causative effects but rather point at common underlying factors^{19,20}. Taken together,
53 the landscape of current psychiatric genetics suggests highly complex patterns of associations and unclear
54 specificity for many common psychiatric disorders.

55
56 While GWAS studies have allowed to disentangle parts of the genetic architecture of psychiatric
57 disorders, these methods alone are not sufficient to answer some of the challenges posed in psychiatric
58 genetics. One of those challenges is the lack of clinical demarcation between psychiatric disorders. For
59 example, patients with the same diagnosis may not necessarily exhibit common symptoms²¹ and patients
60 with different diagnoses may show highly overlapping clinical phenotypes²². The notion that mental
61 disorders like schizophrenia and bipolar disorders reflect biologically heterogeneous categories is also
62 supported by neuroimaging studies^{23,24}. Nonetheless, a majority of large-scale genetic studies use a
63 classical case-control design based on a categorical operationalization of disease without stratifying other
64 measures such as symptoms, functioning or symptom severity. Likewise, control groups are rarely
65 screened for subthreshold symptoms. For example in the case of psychosis, approximately 6% of the
66 general population are reported to have a psychotic experience in their lifetime, and only a minority of that
67 group will develop a diagnosed psychiatric illness such as schizophrenia or bipolar disorder²⁵. Finally, the
68 likelihood of inducing selection bias when drawing cases and controls from different populations are high
69 and may impose confounds in case-control designs²⁶. Thus, whereas studies using the classical case-
70 control design have been instrumental and produced a strong body of discoveries in psychiatric genetics,
71 these designs have limitations that may prevent us from discovering signal more closely related to clinical
72 characteristics of the disorder. In addition, case-control designs require immense effort and resources
73 given that the high polygenicity of common psychiatric disorders requires vast sample sizes to detect
74 effects²⁷.

75
76 Recent large-scale population level efforts such as the UK Biobank²⁸ now provide alternatives for the
77 study of psychiatric disorders. The mental health data available in UK Biobank includes data from more
78 than 150,000 individuals and covers questions on current and previous symptoms in different psychiatric
79 domains²⁹. For example, a recent study revealed genetic associations with psychotic experiences in the UK
80 Biobank and reported genetic correlations between psychotic experience and common psychiatric
81 disorders³⁰. While this study formed two groups of subjects (with and without psychotic experience)
82 others have suggested continuous measures of psychopathology obtained from questionnaire data, such as
83 the p-factor³¹. While bundling variance of psychopathology in a single common factor can be a useful
84 proxy of mental health vulnerability, the specificity of the p-factor to disorder-specific mechanisms is
85 limited³¹. Independent component analysis^{32,33} provides a complementary approach to decompose the
86 variance from mental health questionnaires into independent latent variables. For example, using
87 independent component analysis on mental health questionnaires of children and adolescents, Alnæs and
88 colleagues have identified a set of independent components reflecting symptoms of attention deficit,
89 psychosis, depression, anxiety, and more³⁴. Independent components obtained from mental health
90 questionnaires may each capture either global (e.g. joint symptoms of depression, stress and anxiety) or
91 specific aspects (e.g. pure psychosis symptoms) of mental health in a data-driven fashion, thereby yielding
92 multiple distinct profiles relevant to mental health symptoms beyond a common p-factor.

93
94 Here, in order to disentangle the genetic architecture underlying psychiatric symptoms and traits as well as
95 psychosocial and other risk factor for reduced mental health we investigated structures of
96 psychopathology and corresponding genetics using independent component analysis in the UK Biobank
97 mental health data. This allowed us to study the genetic relationships between statistically independent
98 profiles relevant to mental health, and between these profiles and psychiatric disorders as well as cognitive
99 traits. We focused our analysis on data from individuals who had no previous diagnosis with a
100 neurological or psychiatric disorder, yielding novel insights into variation in mental health in a healthy
101 (undiagnosed) population. Given that preclinical symptoms in healthy individuals may share biological
102 mechanisms with symptoms in diagnosed individuals, we hypothesized that the genetic architecture of
103 specific variations in mental health in *healthy* (undiagnosed) individuals overlaps with specific major
104 psychiatric disorders. However, we did not have an a-priori hypothesis for the degree of specificity. The
105 known pleiotropy between major psychiatric disorders (reproduced in Suppl. Fig. 1) might reflect similar
106 symptoms or risk factors occurring in different disorders or similar mechanisms underlying different
107 symptoms or risks. We therefore investigated if statistically independent mental health profiles are also
108 genetically independent or if they share a common genetic architecture, which may yield insights into the
109 sources of pleiotropy in psychiatric genetics.

110
111 **Methods and Materials**
112 *Sample and exclusion criteria*
113 We accessed data from the UK Biobank²⁸ with permission no. 27412, and included data from individuals
114 who had participated in an online follow-up questionnaire on mental health (UK Biobank category 136).
115 All participants provided signed informed consent before inclusion in the study. UK Biobank was
116 approved by the National Health Service National Research Ethics Service (ref. 11/NW/0382).
117 Participants with a diagnosed psychiatric or neurological disorder (F or G ICD10 diagnosis) were
118 excluded from the analysis except for those with a nerve, nerve root and plexus disorders (categories G50
119 to G59). In addition, we excluded participants with more than 10% missing answers in the mental health
120 questionnaires. This resulted in mental health data from 136,678 individuals, which was used in an
121 independent component analysis. For the genetic analysis, we selected data from all White English
122 individuals with available genotypes, yielding a set of 117,611 participants aged 47-80 years (mean: 64,
123 SD: 7.66, age at mental health assessment) and comprised 56.2% females.

124
125 *Processing of mental health data*
126 Fig. 1A depicts the analysis workflow. The UK Biobank database contains about 140 questions on mental
127 health and risk factors related to reduced mental health. The questions on the mental health risk factors are
128 retrospective (e.g. “have you ever ...”, “at any point in your lifetime ...” etc.). This warrants some caution
129 in the interpretation, since retrospective analysis yields lower power than a prospective design. We
130 selected only primary questions that were answered by all participants, excluding follow-up questions. Of
131 the 60 resulting questions that were available, we removed questions that asked specifically about
132 symptoms occurring in the past two weeks to remove potential short-term temporal effects. Furthermore,
133 we excluded questions where more than 10% of the responses were missing (1 question excluded). In the
134 resulting set of 43 questions (Suppl. Table 1), we imputed missing data using k-nearest neighbor
135 imputation with k=3 with the *bnstruct* package³⁵ in R³⁶ and z-standardized the data (Suppl. Fig. 2).

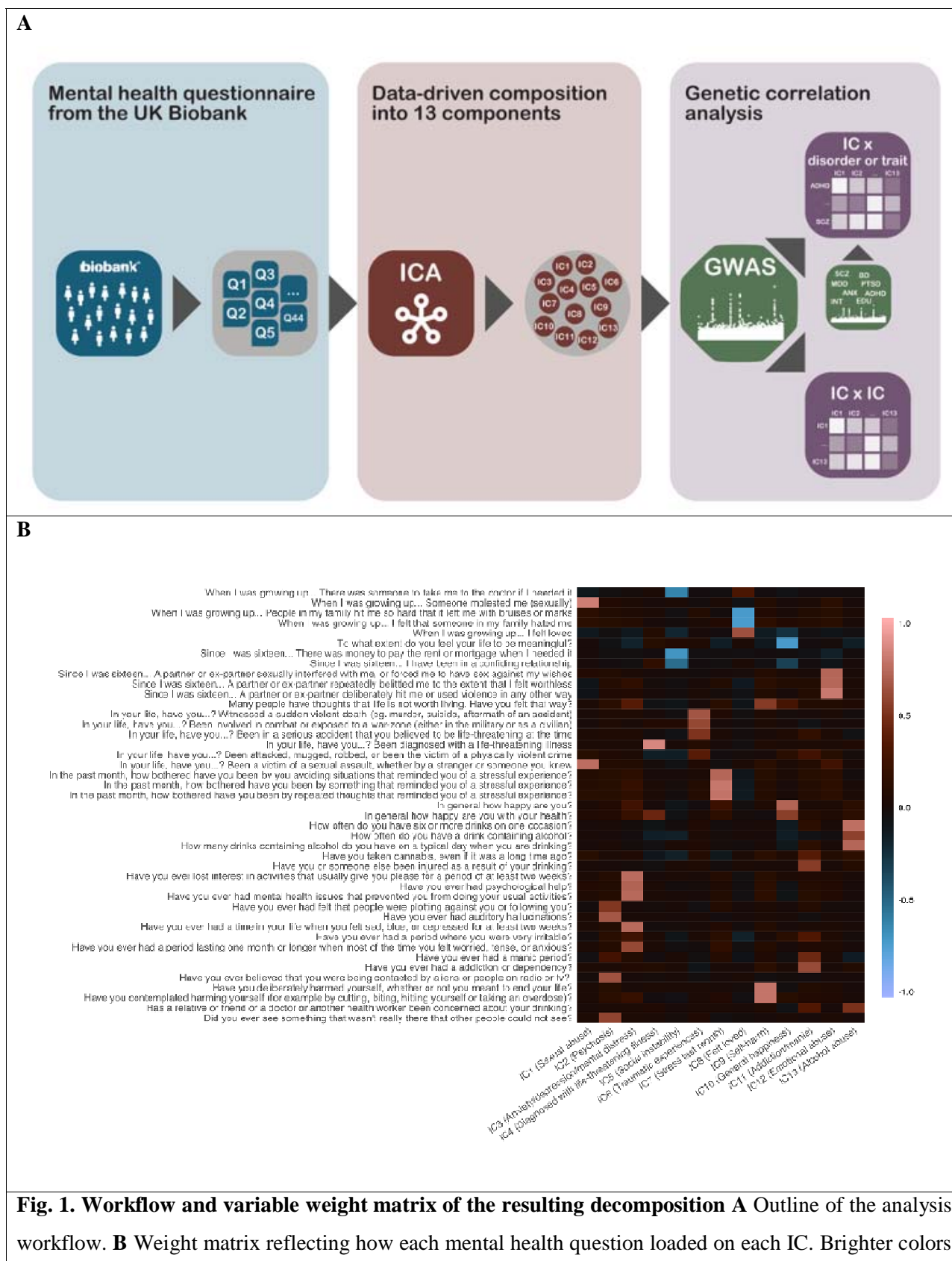


Fig. 1. Workflow and variable weight matrix of the resulting decomposition **A** Outline of the analysis workflow. **B** Weight matrix reflecting how each mental health question loaded on each IC. Brighter colors

indicate higher loading, darker colors indicate lower loading. All 43 questions were captured by at least one of the 13 independent components. To facilitate interpretation, loadings of IC1, IC2, IC5, IC9, IC10, IC11, and IC12 were inverted so that all components showed the same direction of effect (higher component score indicating increased scoring on the items).

137
138 The resulting data covering 43 questions from 136,678 individuals was decomposed using independent
139 component analysis (ICA). ICA is a statistical clustering method that decomposes multiple related
140 variables into statistically independent components. The resulting components show a high degree of
141 within-cluster correlation, but no correlation between the clusters. Of note, the number of components
142 needs to be pre-specified, and this selection of appropriate model order is to a certain degree a subjective
143 task where depending on the stringency of the criteria defined for model order selection it is possible to
144 obtain several different solutions that meet the requirements for an appropriate threshold. Here, we used
145 *icasso*³⁷ in MATLAB in combination with visual inspection of the loadings of the questions on the
146 components. The PCA identified 13 components with an eigenvalue larger than 1, and stability (Iq) was
147 effectively 1. A model order lower than 13 would group together questions into components which we
148 preferred to keep separate. A model order larger than 13 was not reasonable as it would yield components
149 that largely reflect single items. We therefore concluded that a model order of 13 independent components
150 yields the best clustering solution where the resulting components are stable and highly interpretable. The
151 individual scores for each of the 43 questions were subsequently residualized for age (both linear and
152 quadratic term), sex, and the first 20 genetic principal components. Next, we decomposed the items into
153 13 independent components using the fastICA algorithm as implemented in R³⁸. Fig. 1B depicts how each
154 of the 43 items loaded on the components, indicating independent components (ICs) that captured
155 questions on sexual abuse (IC1), psychosis (IC2), anxiety, depression and mental distress (IC3), a
156 diagnosis with a life-threatening illness (IC4), social instability (IC5), traumatic experiences (IC6), stress
157 in the past month (IC7), experiences of feeling loved (IC8), thoughts around self-harm behavior (IC9),
158 general happiness (IC10), addiction behavior and manic experiences (IC11), experiences of emotional
159 abuse (IC12), and alcohol abuse (IC13). Of note, we here introduced this labeling of the ICs only to
160 improve legibility of the results yet caution is warranted as the label is not necessarily encompassing all
161 facets of a given component. The labels only highlight some of the core domains of questions weighing
162 strongly on a given components, yet all interpretations need to be made in the light of the ICA framework
163 (Fig. 1B)

164
165 The distribution of IC2 indicated very few non-zero scores (Suppl. Fig. 3). This component loaded mostly
166 on psychosis questions (Fig. 1B), indicating that only few of the included healthy individuals had
167 symptoms in this domain. We therefore conducted an additional supplemental analysis in which we

168 dichotomized IC2 such that loadings lower than 1 were labeled as “no/few symptoms”, and loadings equal
169 to or higher than 1 were labeled as “with symptoms”.

170

171 *Processing of genetic data*

172 From the UK Biobank v3 imputed genetic data, we removed SNPs with an imputation quality score below
173 0.5, with a minor allele frequency below 0.001, missing in more than 5% of individuals, and that failed the
174 Hardy-Weinberg equilibrium test at $p < 1e-9$. We removed also individuals with more than 10% missing
175 data. We performed a genome-wide association analysis (GWAS) on each of the 13 independent
176 components in PLINK 2^{39,40}. Using a publicly available conversion toolbox for GWAS summary statistics
177 (github.com/precimed/python_convert), we removed the MHC region and calculated a z-score for every
178 SNP (8,165,726 SNPs after QC). We utilized linkage-disequilibrium score regression^{10,41} to estimate
179 genetic correlations between each of the independent components, and between the components and
180 publicly available GWAS summary statistics for SCZ¹, BD², MD⁴², ADHD⁴, ASD⁵, PTSD⁶, ANX⁷, as
181 well as intelligence⁴³, and educational attainment⁴⁴ (Suppl. Table 2). For all aforementioned GWASs, we
182 used those versions that did not have UK Biobank participants included. From the MD GWAS, we
183 removed participants from the 23andMe dataset as well, leaving only cases with a diagnosed major
184 depressive disorder (MDD). Prior to estimating genetic correlations, we set a threshold that only ICs with
185 a heritability 1.96 times larger than its standard error should be included in the analysis and only those
186 where visual quality control of corresponding Q-Q plots indicated genetic signal. These quality control
187 steps were implemented to ensure that we did not make inferences on data that did not provide sufficient
188 variance explained by genetics. Partitioned heritability⁴⁵ was estimated using the LDSC toolbox⁴¹ and Q-Q
189 plots were generated using custom scripts in R. Finally, we processed the GWAS summary statistics of
190 each independent component through the Functional Mapping and Annotation toolbox (FUMA) to map
191 lead SNPs onto genes⁴⁶. FUMA parameters were kept as default, and we used the FUMA default
192 European ancestry reference panel.

193

194 *Code and data availability*

195 Code and GWAS summary statistics will be made publicly available via GitHub (github.com/norment)
196 upon acceptance of the manuscript. Furthermore, the derived independent components (individual level
197 data) will be made available to the UK Biobank upon acceptance (derived variable return) to allow its use
198 in future UK Biobank studies.

199

200 **Results**

201 Fig. 2 depicts SNP-based heritability (h^2) for the 13 ICs (Suppl. Table 3 for additional statistics).
202 Heritability was generally low, yet all components yielded a heritability that was higher than 1.96 times
203 the standard error. IC13, capturing questions on alcohol abuse had the highest heritability ($h^2 = 0.0763$,
204 $SE = 0.0055$), closely followed by IC3, capturing anxiety, depression, and mental distress ($h^2 = 0.0744$,
205 $SE = 0.0052$). The lowest heritability among the components was for IC2, reflecting psychosis questions
206 ($h^2 = 0.0089$, $SE = 0.0043$), likely owing to the low number of individuals with psychosis symptoms
207 (Suppl. Fig. 3). We therefore performed a supplemental analysis to investigate if dichotomization of this
208 IC would benefit the analysis (Suppl. Fig. 4). In brief, as dichotomization only slightly improved
209 heritability estimates, we kept IC2 as a continuous component for the main analysis to stay consistent with
210 the other components, yet we provide results with the dichotomized component in Suppl. Fig. 4. In
211 addition to passing the heritability criterion of 1.96 times the standard error, the Q-Q plots of all ICs
212 passed visual quality control (Suppl. Fig. 5) warranting inclusion of all components into subsequent
213 genetic correlation analyses.

214

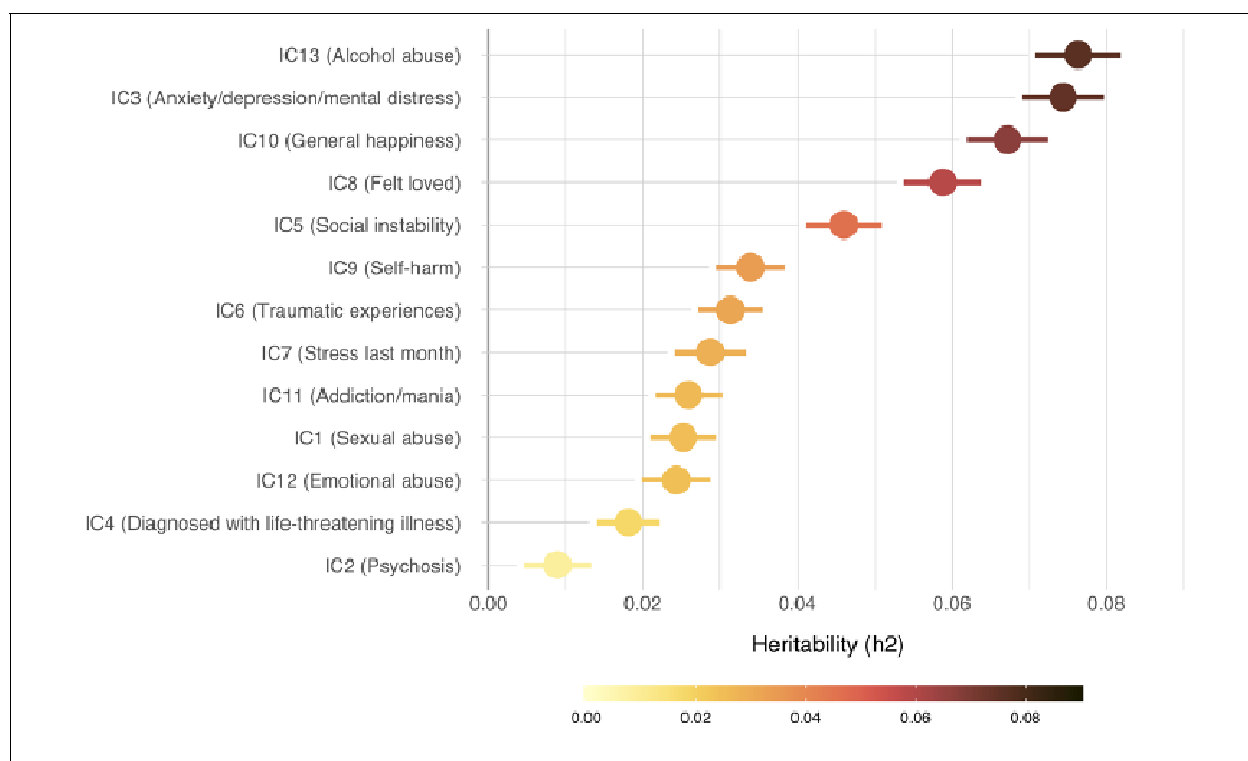


Fig. 2. Heritability estimates of the independent components SNP-based heritability for each IC sorted by decreasing heritability (h^2). Heritability calculated using the LDSC toolbox⁴¹. Error bars reflect standard errors.

215

216 Except for IC4, all ICs showed genome wide significant SNPs at a threshold of $5e-8$ (Suppl. Fig. 6). Using
217 FUMA, we discovered 7 independent loci for IC13, 2 for IC2, IC7, and IC8, 1 locus for IC1, IC3, IC5,
218 IC10, IC11 and IC12, and IC4, IC6, and IC9 had no significant genetic risk loci. Suppl. Table 4 provides a
219 list of mapped genes for all ICs, illustrating that IC13 had the most mapped genes among all ICs (74
220 mapped genes).

221
222 We assessed genetic correlations between each of the 13 ICs and a set of psychiatric disorders as well as
223 cognitive traits. Out of 117 comparisons, 70 were significant after FDR correction ($\alpha = 0.05$), which
224 amounts to 59%. Fig. 3 depicts all genetic correlations with ICs, sorted separately for each disorder or
225 cognitive trait (sorted by absolute genetic correlation). Suppl. Fig. 7 shows the same genetic correlations
226 separated by IC. We found that in most cases the strongest genetic correlation was with the IC most
227 closely related to that disorder or trait. For example, anxiety most strongly correlated with IC3, which
228 reflects anxiety, depression, and mental distress (genetic correlation $r_g = 0.70$, $p_{FDR} < .00027$). SCZ was
229 most highly correlated with IC2, which represents psychosis questions ($r_g = 0.54$, $p_{FDR} = .001$). The
230 highest genetic correlation of BD was with IC11, which represents addiction and mania ($r_g = 0.5$, $p_{FDR} =$
231 $6.5e-12$). For PTSD, the component reflecting traumatic experience (IC6) only ranked sixth among the
232 sorted associations, yet the two ICs showing strongest association with PTSD reflected anxiety,
233 depression, and mental distress (IC3; $r_g = 0.53$, $p_{FDR} = .0017$) and diagnosed with life-threatening illness
234 (IC4; $r_g = 0.51$, $p_{FDR} = .080$), both of which are closely related to PTSD. ASD correlated strongest with
235 IC2 (reflecting psychosis; $r_g = 0.40$, $p_{FDR} = .031$) and ADHD correlated strongest with IC8 (Felt loved; r_g
236 $= -0.51$, $p_{FDR} = 4.7e-21$). Educational attainment and intelligence were both strongest negatively correlated
237 with the IC reflecting social instability (IC5, $r_g = -0.74$ and $r_g = -0.76$, respectively; both $p_{FDR} < 2.5e-74$).
238 In general, the strongest associations among all ICs, either positive or negative were with MDD while the
239 weakest associations were with educational attainment.

240

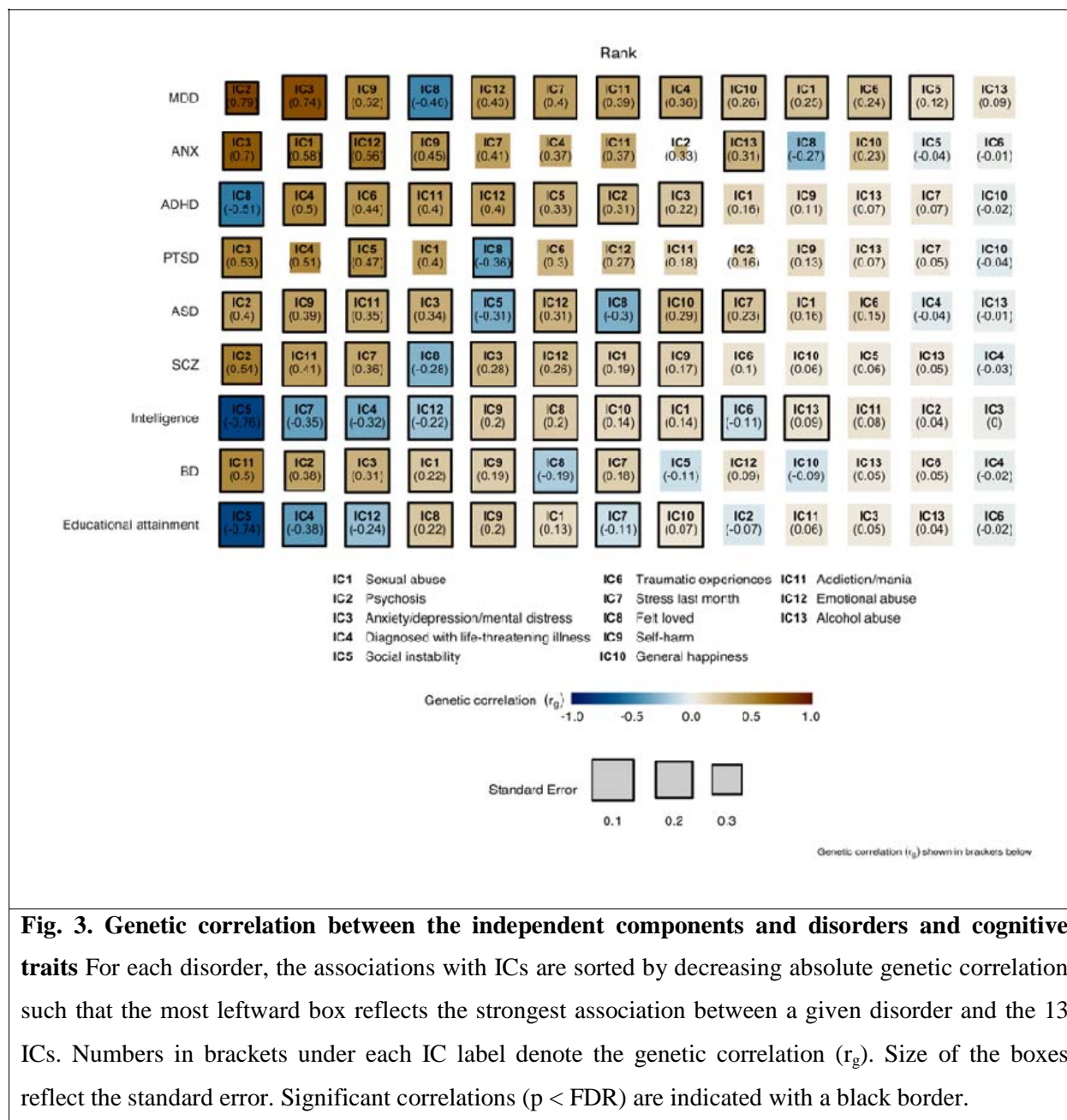


Fig. 3. Genetic correlation between the independent components and disorders and cognitive traits For each disorder, the associations with ICs are sorted by decreasing absolute genetic correlation such that the most leftward box reflects the strongest association between a given disorder and the 13 ICs. Numbers in brackets under each IC label denote the genetic correlation (r_g). Size of the boxes reflect the standard error. Significant correlations ($p < \text{FDR}$) are indicated with a black border.

241
 242 Next, we assessed the genetic correlations between the ICs. Independent components are statistically
 243 independent by design, and thus on the phenotype level the ICs were not correlated with each other (Fig.
 244 4, lower half; correlations essentially zero). However, almost half of the IC pairs were nonetheless
 245 significantly genetically correlated with each other (45%, $p < \text{FDR}$). IC3 (anxiety, depression, mental
 246 illness) was genetically correlated with 10 other ICs. IC9 (self-harm) was correlated with 9 other ICs and
 247 IC6 (traumatic experiences) and IC8 (felt loved) were each genetically correlated with eight other ICs.
 248 IC11 (addiction/mania) and IC12 (emotional abuse) were each genetically correlated with seven other ICs.
 249 IC1 (sexual abuse) and IC5 (social instability) were both genetically correlated with six other ICs. IC2

250 (psychosis) was correlated with 5 other ICs. IC4 (diagnosed with life-threatening illness) and IC13
 251 (alcohol abuse) were both genetically correlated with 4 other ICs. And IC7 (stress last month) and IC10
 252 (general happiness) were both genetically correlated with 3 other ICs. No IC had no significant genetic
 253 correlations with other ICs. The analysis therefore revealed a large amount of genetic correlations despite
 254 statistical (phenotypic) independence of the symptom profiles.
 255

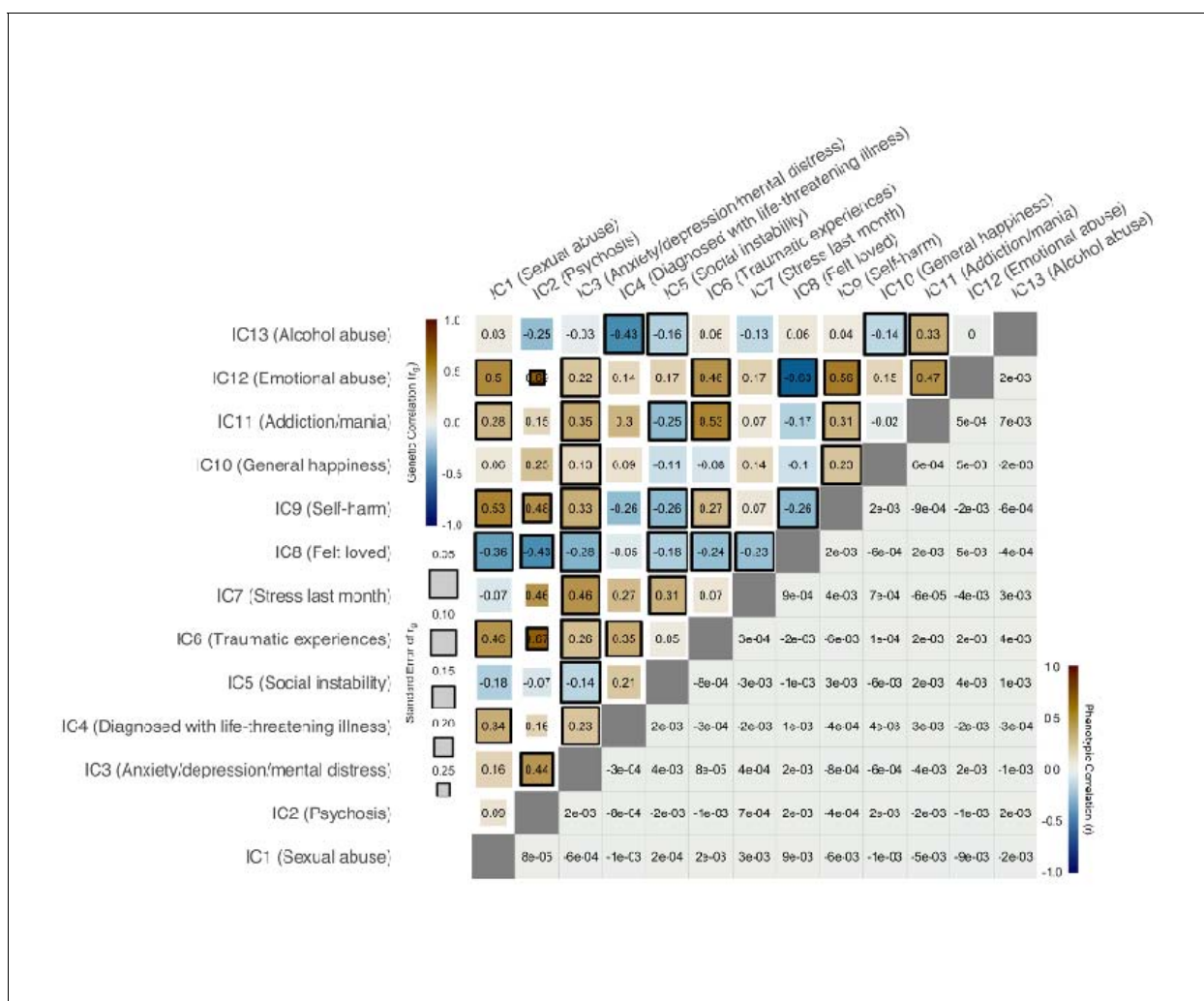


Fig. 4. Phenotypic and genetic correlation between the ICs The lower half of the IC by IC matrix depicts phenotypic correlations, reflecting the Pearson correlation of subject level component scores between independent components. As expected by ICA design, correlations were close to zero. The upper half of the matrix depicts the genetic correlations (r_g), indicating significant genetic correlations in 40 of 78 tests. Size of the boxes indicate standard error and significant correlations ($p < FDR$) are indicated with a black border.

257 **Discussion**

258 In the present study, we decomposed mental health questionnaire data from more than 130,000 individuals
259 into phenotypically distinct profiles relevant to mental health (independent components) that reflected
260 compositions of symptoms, psychosocial and other risk factors for reduced mental health. We found that
261 variations in these profiles in healthy individuals (without a neurological or psychiatric diagnosis) were
262 genetically correlated with psychiatric disorders and cognitive traits. Strongest correlations were observed
263 between components and disorders with known symptoms in a similar domain (e.g. psychosis symptoms
264 with schizophrenia), but the large amount of significant correlations between disorders and mental health
265 profiles suggested limited specificity. Indeed, we found a large proportion of significant genetic
266 correlations between the phenotypically uncorrelated profiles, suggesting overlapping genetic
267 architectures underlying distinct symptoms and risk factors. A number of the questions included in the
268 analyses revolved around risk factors for mental health, such as a history of childhood abuse, sexual
269 abuse, and an unstable home situation. Caution is warranted in the interpretation of these effects. The
270 genetic correlation with the independent components capturing these items do not suggest that there is a
271 genetic component to high-risk environments but rather are likely to capture second order effects. In other
272 words, a GWAS on such these risk factors is more likely to reflect other factors underlying these risk
273 factors than the risk factor directly.

274
275 The implications of our findings are twofold. First, our results support pleiotropy in psychiatric disorders
276 beyond overlapping symptoms (e.g. BD and MDD both involving depressive episodes), suggesting that
277 even distinct psychiatric symptoms are genetically overlapping. Second, our findings support that normal
278 variability in mental health within healthy individuals may inform the study of the biology of psychiatric
279 disorders.

280
281 While pleiotropy between major psychiatric disorders has been widely established⁹⁻¹¹ (reproduced in
282 Suppl. Fig. 1), the sources underlying pleiotropy remain largely unknown. Specifically, disorders
283 oftentimes overlap in symptomatology and therefore the degree to which the observed genetic correlations
284 between disorders reflect phenotypic overlap between disorders remains to be investigated. Our approach
285 of decomposing mental health data into distinct profiles allowed us to study genetic correlations in a
286 sample with known phenotypic correlations and to assess how these profiles correlate with the genetics of
287 different diagnoses. We observed that most disorders correlated strongest with the independent
288 components capturing a related phenotype. For example, the strongest association with IC3, which reflects
289 variance in anxiety, depression, and mental distress, was with ANX, the strongest association with IC2
290 (psychosis) was with SCZ. Therefore, the ranking of association strengths suggested a certain degree of
291 specificity. However, that degree was strongly limited as most of the disorders and components were

292 significantly genetically correlated. For example, MDD showed significant correlations between all but
293 one component, ASD correlated with all but four components, and ANX and ADHD were correlated with
294 all but 5 components, though correlation strengths were overall lower than with MDD, possibly due to
295 lower sample size. There were also significant associations between components and cognitive traits
296 although overall weaker associations compared to those with disorders. About half of the genetic
297 correlations with intelligence and educational attainment pointed in the opposite direction, considerably
298 more than for the psychiatric disorders, reflecting higher cognitive ability with fewer psychiatric
299 symptoms. Importantly, when looking at the correlations between mental health profiles, we found that
300 almost half of the genetic correlation matrix between ICs yielded significant genetic correlations despite a
301 lack of phenotypic correlations (independence of the components). This suggests that some of the same
302 genes are involved in the genetics of distinct profiles relevant to mental health and may indirectly support
303 pleiotropy independent of phenotypic overlap in psychiatric disorders. Whereas more research is needed
304 before conclusions on the sources underlying the observed pleiotropy can be drawn, one possible
305 explanation for the significant correlations in the ICs could be that, since all independent components each
306 capture a facet of mental health, there may be a number of SNPs that are involved across mental health
307 symptoms. These SNPs may be involved in overall mental health, from psychological well-being to
308 psychosis symptoms. Our analysis of significant SNPs in FUMA did not identify overlapping SNPs
309 between ICs, however, this may be attributed to the relatively low number of significant loci discovered in
310 the ICs. Advanced statistical tools and further increasing sample sizes may help pinpoint specific genes
311 involved with different symptoms. Furthermore, it is also plausible that environmental effects may factor
312 into the explanation of the significant genetic correlations despite phenotypic independence if the
313 environmental factors differ markedly between the ICs.

314

315 *Limitations*

316 Notable strengths of the present study include the use of data-driven decomposition of mental health data
317 in a large sample of healthy individuals and its application to study pleiotropy in psychiatric genetics. Its
318 main limitations include the low heritability of the resulting independent components, and the limited
319 number of individuals with psychosis symptoms yielding suboptimal distribution in IC2 (Suppl. Fig. 3).
320 First, it is important to note that all ICs passed quality control. Heritability of all ICs exceeded our pre-
321 defined heritability threshold of 1.96 times its standard error, and all Q-Q plots indicated genetic signal
322 (Suppl. Fig. 5). Furthermore, low heritability can still produce good genetic signal as a result from a low
323 number of genetic variants involved but where each has large effects¹³. For example, while IC2 had the
324 lowest heritability among the ICs, it showed one of the strongest genetic signals and together with IC7 and
325 IC8 it ranked second in terms of the number of loci discovered in FUMA, following IC13 (alcohol abuse)
326 that showed the highest heritability, strongest genetic signal on the Q-Q plot and the largest number of

327 significant loci and mapped genes. Second, although sample size and symptom distributions factored into
328 the results, these are mostly reflected in the standard error of genetic associations, not in a lack of effect.
329 For example, ANX⁷ (n = 21,761) and PTSD⁶ (n = 9,537) GWASs have relatively little power as reflected
330 in the larger standard errors in genetic correlations with these disorders, but nonetheless the strongest
331 associations with these disorders were with components that match symptoms of the disorders (both
332 correlated strongest with IC3, reflecting anxiety/depression/mental distress). Likewise, the suboptimal
333 symptom distributions in IC2 and corresponding low heritability is reflected in large standard errors of the
334 resulting genetic correlations but nonetheless IC2, reflecting psychosis, was most strongly associated with
335 SCZ. Supplemental analysis with dichotomized IC2 also confirmed that the distribution alone is unlikely
336 to explain the observed associations (Suppl. Fig. 4).

337
338 Furthermore, it is important to note that while we excluded individuals with psychiatric disorders based on
339 ICD codes we may still include individuals with (sub-threshold) psychiatric disorders that have not been
340 diagnosed. For example, most patients with depressive symptoms in the UK will be treated by first-line
341 care, which may not be registered in the UK Biobank⁴⁷. However, distributions of z-scores on the
342 individual questions appear quite similar between individuals without and individuals with a diagnosis
343 (Suppl. Fig. 2). While we cannot rule out subtle contributions on IC decomposition by the potential
344 remaining inclusion of a subset of patients in primary care, these results support that should such a
345 confound be present, it is unlikely to have introduced enough structural variance to diminish our main
346 findings.

347
348 *Conclusion*
349 In the present study, we revealed genetic overlap between statistically independent profiles relevant to
350 mental health capturing compositions of symptoms, psychosocial and other risk factors for reduced mental
351 health and provide evidence that variations in mental health in healthy individuals relate genetically to
352 psychiatric disorders and cognitive traits. These findings support that pleiotropy between psychiatric
353 disorders cannot simply be explained by overlapping symptoms or risks but may rather point to similar
354 biological underpinnings of distinct symptoms or risks. Our results underscore the potential of data-driven
355 approaches to the study of mental health, and suggests that supplementing the classic case-control design
356 with a dimensional approach may improve our understanding of the genetic underpinnings of complex
357 disorders of the mind.

358

359 **Acknowledgements**

360 The authors were funded by the Research Council of Norway (#276082 LifespanHealth, #223273
361 NORMENT, #283798 ERA-NET Neuron SYNSCHIZ, #249795), the South-East Norway Regional
362 Health Authority (2019101, 2019107, and 2020086), and the European Research Council under the
363 European Union's Horizon2020 Research and Innovation program (ERC Starting Grant #802998), as well
364 as the Horizon2020 Research and Innovation Action Grant CoMorMent (#847776). This research has
365 been conducted using the UK Biobank Resource (access code 27412, <https://www.ukbiobank.ac.uk/>). This
366 work was performed on the TSD (Tjenester for Sensitive Data) facilities, owned by the University of
367 Oslo, operated and developed by the TSD service group at the University of Oslo, IT-Department (USIT).
368 Computations were also performed on resources provided by UNINETT Sigma2 - the National
369 Infrastructure for High Performance Computing and Data Storage in Norway.

371 **Conflicts of interest**

372 D.R., D.A., O.F., D.vd.M., O.B.S., L.T.W. and T.K. declare no conflicts of interest. O.A.A. is a consultant
373 to HealthLytix and received speakers honorarium from Lundbeck.

375 **Author contributions**

376 D.R. and T.K. conceived the study; D.R. analyzed the data with contributions from T.K.; All authors
377 contributed with conceptual input on methods and/or interpretation of results; D.R. and T.K. wrote the
378 first draft of the paper and all authors contributed to the final manuscript.

380 **References**

- 381 1. Pardiñas AF, Holmans P, Pocklington AJ, et al. Common schizophrenia alleles are enriched in
382 mutation-intolerant genes and in regions under strong background selection. *Nat Genet.* 2018;50(3):381-
383 389. doi:10.1038/s41588-018-0059-2
- 384 2. Stahl EA, Breen G, Forstner AJ, et al. Genome-wide association study identifies 30 loci
385 associated with bipolar disorder. *Nat Genet.* 2019;51(5):793-803. doi:10.1038/s41588-019-0397-8
- 386 3. Howard DM, Adams MJ, Clarke T-K, et al. Genome-wide meta-analysis of depression identifies
387 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci.*
388 2019;22(3):343-352. doi:10.1038/s41593-018-0326-7
- 389 4. Demontis D, Walters RK, Martin J, et al. Discovery of the first genome-wide significant risk loci
390 for attention deficit/hyperactivity disorder. *Nat Genet.* 2019;51(1):63-75. doi:10.1038/s41588-018-0269-7
- 391 5. Grove J, Ripke S, Als TD, et al. Identification of common genetic risk variants for autism
392 spectrum disorder. *Nat Genet.* 2019;51(3):431-444. doi:10.1038/s41588-019-0344-8

- 393 6. Duncan LE, Ratanatharathorn A, Aiello AE, et al. Largest GWAS of PTSD (N=20 070) yields
394 genetic overlap with schizophrenia and sex differences in heritability. *Mol Psychiatry*. 2018;23(3):666-
395 673. doi:10.1038/mp.2017.77
- 396 7. Otowa T, Hek K, Lee M, et al. Meta-analysis of genome-wide association studies of anxiety
397 disorders. *Mol Psychiatry*. 2016;21(10):1391-1399. doi:10.1038/mp.2015.197
- 398 8. Andreassen OA, Thompson WK, Schork AJ, et al. Improved detection of common variants
399 associated with schizophrenia and bipolar disorder using pleiotropy-informed conditional false discovery
400 rate. *PLoS Genet*. 2013;9(4):e1003455. doi:10.1371/journal.pgen.1003455
- 401 9. Anttila V, Bulik-Sullivan B, Finucane HK, et al. Analysis of shared heritability in common
402 disorders of the brain. *Science*. 2018;360(6395):eaap8757. doi:10.1126/science.aap8757
- 403 10. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human
404 diseases and traits. *Nat Genet*. 2015;47(11):1236-1241. doi:10.1038/ng.3406
- 405 11. Smeland OB, Bahrami S, Frei O, et al. Genome-wide analysis reveals extensive genetic overlap
406 between schizophrenia, bipolar disorder, and intelligence. *Mol Psychiatry*. Published online January 4,
407 2019;10.1038/s41380-018-0332-x. doi:10.1038/s41380-018-0332-x
- 408 12. Bansal V, Mitjans M, Burik CAP, et al. Genome-wide association study results for educational
409 attainment aid in identifying genetic heterogeneity of schizophrenia. *Nat Commun*. 2018;9(1):3078.
410 doi:10.1038/s41467-018-05510-z
- 411 13. Frei O, Holland D, Smeland OB, et al. Bivariate causal mixture model quantifies polygenic
412 overlap between complex traits beyond genetic correlation. *Nat Commun*. 2019;10(1):2417-2417.
413 doi:10.1038/s41467-019-10310-0
- 414 14. Bahrami S, Steen NE, Shadrin A, et al. Shared Genetic Loci Between Body Mass Index and
415 Major Psychiatric Disorders: A Genome-wide Association Study. *JAMA Psychiatry*. Published online
416 January 8, 2020;10.1001/jamapsychiatry.2019.4188. doi:10.1001/jamapsychiatry.2019.4188
- 417 15. Amare AT, Schubert KO, Klingler-Hoffmann M, Cohen-Woods S, Baune BT. The genetic
418 overlap between mood disorders and cardiometabolic diseases: a systematic review of genome wide and
419 candidate gene studies. *Transl Psychiatry*. 2017;7(1):e1007-e1007. doi:10.1038/tp.2016.261
- 420 16. Straatmann VS, Lai E, Lange T, et al. How do early-life factors explain social inequalities in
421 adolescent mental health? Findings from the UK Millennium Cohort Study. *J Epidemiol Community
422 Health*. 2019;73(11):1049-1060. doi:10.1136/jech-2019-212367
- 423 17. Gunnell D, Kidger J, Elvidge H. Adolescent mental health in crisis. *BMJ*. 2018;361:k2608.
424 doi:10.1136/bmj.k2608
- 425 18. Chen LP, Murad MH, Paras ML, et al. Sexual abuse and lifetime diagnosis of psychiatric
426 disorders: systematic review and meta-analysis. *Mayo Clin Proc*. 2010;85(7):618-629.
427 doi:10.4065/mcp.2009.0583

- 428 19. Anderson JS, Shade J, DiBlasi E, Shabalin AA, Docherty AR. Polygenic risk scoring and
429 prediction of mental health outcomes. *Curr Opin Psychol.* 2019;27:77-81.
430 doi:10.1016/j.copsyc.2018.09.002
- 431 20. Selzam S, Krapohl E, von Stumm S, et al. Predicting educational achievement from DNA. *Mol*
432 *Psychiatry.* 2017;22(2):267-272. doi:10.1038/mp.2016.107
- 433 21. Andreasen NC. A unitary model of schizophrenia: Bleuler's "fragmented phrene" as
434 schizencephaly. *Arch Gen Psychiatry.* 1999;56(9):781-787. doi:10.1001/archpsyc.56.9.781
- 435 22. Craddock N, Owen MJ. The Kraepelinian dichotomy - going, going... but still not gone. *Br J*
436 *Psychiatry J Ment Sci.* 2010;196(2):92-95. doi:10.1192/bjp.bp.109.073429
- 437 23. Alnaes D, Kaufmann T, van der Meer D, et al. Brain Heterogeneity in Schizophrenia and Its
438 Association With Polygenic Risk. *JAMA Psychiatry.* 2019;76(7):739-748.
439 doi:10.1001/jamapsychiatry.2019.0257
- 440 24. Hibar DP, Westlye LT, Doan NT, et al. Cortical abnormalities in bipolar disorder: an MRI
441 analysis of 6503 individuals from the ENIGMA Bipolar Disorder Working Group. *Mol Psychiatry.*
442 2018;23(4):932-942. doi:10.1038/mp.2017.73
- 443 25. McGrath JJ, Saha S, Al-Hamzawi A, et al. Psychotic Experiences in the General Population: A
444 Cross-National Analysis Based on 31,261 Respondents From 18 Countries. *JAMA Psychiatry.*
445 2015;72(7):697-705. doi:10.1001/jamapsychiatry.2015.0575
- 446 26. Pirastu N, Cordioli M, Nandakumar P, et al. Genetic analyses identify widespread sex-differential
447 participation bias. *bioRxiv.* Published online January 1, 2020:2020.03.22.001453.
448 doi:10.1101/2020.03.22.001453
- 449 27. Holland D, Frei O, Desikan R, et al. Beyond SNP Heritability: Polygenicity and Discoverability
450 of Phenotypes Estimated with a Univariate Gaussian Mixture Model. *bioRxiv.* Published online January 1,
451 2019:133132. doi:10.1101/133132
- 452 28. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the
453 causes of a wide range of complex diseases of middle and old age. *PLoS Med.* 2015;12(3):e1001779.
454 doi:10.1371/journal.pmed.1001779
- 455 29. Davis KAS, Coleman JRI, Adams M, et al. Mental health in UK Biobank – development,
456 implementation and results from an online questionnaire completed by 157 366 participants: a reanalysis.
457 *BJPpsych Open.* 2020;6(2):e18. doi:10.1192/bjo.2019.100
- 458 30. Legge SE, Jones HJ, Kendall KM, et al. Association of Genetic Liability to Psychotic Experiences
459 With Neuropsychotic Disorders and Traits. *JAMA Psychiatry.* 2019;76(12):1256-1265.
460 doi:10.1001/jamapsychiatry.2019.2508

- 461 31. Caspi A, Houts RM, Belsky DW, et al. The p Factor: One General Psychopathology Factor in the
462 Structure of Psychiatric Disorders? *Clin Psychol Sci J Assoc Psychol Sci*. 2014;2(2):119-137.
463 doi:10.1177/2167702613497473
- 464 32. Jutten C, Héroult J. Détection de grandeurs primitives dans un message composite par une
465 architecture de calcul neuromimétique en apprentissage non supervisé. *10° Colloq Sur Trait Signal Images*
466 *FRA 1985*. Published online 1985:1017-1022.
- 467 33. Comon P. Independent component analysis, A new concept? *High Order Stat*. 1994;36(3):287-
468 314. doi:10.1016/0165-1684(94)90029-9
- 469 34. Alnæs D, Kaufmann T, Doan NT, et al. Association of Heritable Cognitive Ability and
470 Psychopathology With White Matter Properties in Children and Adolescents. *JAMA Psychiatry*.
471 2018;75(3):287-295. doi:10.1001/jamapsychiatry.2017.4277
- 472 35. Alberto Franzin, Francesco Sambo, Barbara di Camillo. bnstruct: an R package for Bayesian
473 Network structure learning in the presence of missing data. *Bioinformatics*. 2017;33(8):1250-1252.
474 doi:10.1093/bioinformatics/btw807
- 475 36. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for
476 Statistical Computing; 2017. www.R-project.org
- 477 37. Himberg J, Hyvarinen A. Icasto: software for investigating the reliability of ICA estimates by
478 clustering and visualization. *2003 IEEE XIII Workshop Neural Netw Signal Process IEEE Cat*
479 *No03TH8718*. Published online 2003:259-268. doi:10.1109/NNSP.2003.1318025
- 480 38. Marchini J, Heaton C, Ripley B. *FastICA: FastICA Algorithms to Perform ICA and Projection*
481 *Pursuit.*; 2019. <https://CRAN.R-project.org/package=fastICA>
- 482 39. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK:
483 rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7. doi:10.1186/s13742-015-0047-
484 8
- 485 40. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and
486 population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-575. doi:10.1086/519795
- 487 41. Bulik-Sullivan BK, Loh PR, Finucane HK, et al. LD Score regression distinguishes confounding
488 from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47(3):291-295.
489 doi:10.1038/ng.3211
- 490 42. Wray NR, Ripke S, Mattheisen M, et al. Genome-wide association analyses identify 44 risk
491 variants and refine the genetic architecture of major depression. *Nat Genet*. 2018;50(5):668-681.
492 doi:10.1038/s41588-018-0090-3
- 493 43. Savage JE, Jansen PR, Stringer S, et al. Genome-wide association meta-analysis in 269,867
494 individuals identifies new genetic and functional links to intelligence. *Nat Genet*. 2018;50(7):912-919.
495 doi:10.1038/s41588-018-0152-6

Roelfs et al. | Genetics of profiles relevant to mental health

- 496 44. Lee JJ, Wedow R, Okbay A, et al. Gene discovery and polygenic prediction from a genome-wide
497 association study of educational attainment in 1.1 million individuals. *Nat Genet.* 2018;50(8):1112-1121.
498 doi:10.1038/s41588-018-0147-3
- 499 45. Finucane HK, Bulik-Sullivan B, Gusev A, et al. Partitioning heritability by functional annotation
500 using genome-wide association summary statistics. *Nat Genet.* 2015;47(11):1228-1235.
501 doi:10.1038/ng.3404
- 502 46. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of
503 genetic associations with FUMA. *Nat Commun.* 2017;8(1):1826. doi:10.1038/s41467-017-01261-5
- 504 47. Glanville KP, Coleman JRI, Howard DM, et al. Multiple measures of depression to enhance
505 validity of Major Depressive Disorder in the UK Biobank. *medRxiv.* Published online January 1,
506 2020:2020.09.18.20196451. doi:10.1101/2020.09.18.20196451
507