Phenotypically independent mental health profiles are genetically related

Daniel Roelfs[1,*], MSc, Dag Alnæs[1], PhD, Oleksandr Frei[1], PhD, Dennis van der Meer[1,2], PhD, Olav B. Smeland[1], PhD, Ole A. Andreassen[1], PhD, Lars T. Westlye[1,3], PhD, Tobias Kaufmann[1,*], PhD

[1] NORMENT, KG Jebsen Centre for Neurodevelopmental Disorders, Division of Mental Health and Addiction, Oslo University Hospital & Institute of Clinical Medicine, University of Oslo, Oslo, Norway

[2] School of Mental Health and Neuroscience, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands

[3] Department of Psychology, University of Oslo, Oslo, Norway

* Correspondence: Daniel Roelfs & Tobias Kaufmann, Ph.D.
Email: daniel.roelfs@medisin.uio.no, tobias.kaufmann@medisin.uio.no
Postal address: OUS, PO Box 4956 Nydalen, 0424 Oslo, Norway
Telephone: +47 23 02 73 50, Fax: +47 23 02 73 33

Counts: Main: 3545 words | Abstract: 244 words | Tables: 0 | Figures: 4 | Supplementary Tables: 4 | Supplementary Figures: 6

Keywords: Mental health profiles, Pleiotropy, Psychiatric disorders, Cognitive traits, Independent component analysis, Psychiatric genetics

Abstract

Background: Genome-wide association studies (GWAS) and family-based studies have revealed partly overlapping genetic architectures between various psychiatric disorders. Given clinical overlap between disorders our knowledge of the genetic architectures underlying specific symptom profiles is limited. Here, we aimed to derive distinct profiles of mental health in healthy individuals and to study how these genetically relate to each other and to common psychiatric disorders.

Methods: We decomposed self-report mental health questionnaires from 136,678 healthy individuals of the UK Biobank, excluding data from individuals with a diagnosed neurological or psychiatric disorder, into thirteen distinct mental health profiles using independent component analysis. Utilizing genotypes from 117,611 of those individuals with Caucasian ancestry, we performed GWAS for each mental health profile and assessed genetic correlations between these profiles, and between the profiles and common psychiatric disorders and cognitive traits.

Results: We found that mental health profiles were genetically correlated with a wide range of psychiatric disorders and cognitive traits, with strongest effects typically observed between a given mental health profile and a disorder for which the profile is common (e.g. depression symptoms and major depressive disorder, psychosis and schizophrenia). Strikingly, although the profiles were phenotypically uncorrelated, many of them were genetically correlated with each other.

Conclusions: This study provides evidence that statistically independent mental health profiles in healthy individuals partly share genetic underpinnings and show genetic overlaps with psychiatric disorders, suggesting that shared genetics across psychiatric disorders cannot be exclusively attributed to the overlapping symptomatology between the disorders.

Introduction

Psychiatric disorders are highly polygenic, exhibiting a multitude of significantly associated genetic variants with small effect sizes. Recent large-scale genome-wide association studies (GWAS) have identified a large number of single-nucleotide polymorphisms (SNP) associated with psychiatric disorders such as schizophrenia (SCZ)[1], bipolar disorder (BD)[2], major depression (MD)[3], attention deficit hyperactivity disorder (ADHD)[4], autism spectrum disorders (ASD)[5], post-traumatic stress disorder (PTSD)[6], and anxiety (ANX)[7]. In addition to substantial polygenicity, previous findings have documented genetic overlap between disorders[8–11], even in the absence of genetic correlations as recently demonstrated for schizophrenia and educational attainment[12,13]. Adding to the complexity, psychiatric disorders also overlap with multiple complex traits, such as BMI[14] and cardio-metabolic diseases[15]. Taken together, the landscape of current psychiatric genetics suggests highly complex patterns of associations and unclear specificity for many common psychiatric disorders.

While GWAS studies have allowed to disentangle parts of the genetic architecture of psychiatric disorders, these methods alone are not sufficient to answer some of the challenges posed in psychiatric genetics. One of those challenges is the lack of clinical demarcation between psychiatric disorders. For example, patients with the same diagnosis may not necessarily exhibit common symptoms[16] and patients with different diagnoses may show highly overlapping clinical phenotypes[17]. The notion that mental disorders like schizophrenia and bipolar disorders reflect biologically heterogeneous categories is also supported by neuroimaging studies[18,19]. Nonetheless, a majority of large-scale genetic studies use a classical case-control design based on a categorical operationalization of disease without stratifying other measures such as symptoms, functioning or symptom severity. Likewise, control groups are rarely screened for subthreshold symptoms. For example in the case of psychosis, approximately 6% of the general population are reported to have a psychotic experience in their lifetime, and only a minority of that group will develop a diagnosed psychiatric illness such as schizophrenia or bipolar disorder[20]. Finally, the likelihood of inducing selection bias when drawing cases and controls from different populations are high and may impose confounds in case-control designs[21]. Thus, whereas studies using the classical case-control design have been instrumental and produced a strong body of discoveries in psychiatric genetics, these designs have limitations that may prevent us from discovering signal more closely related to clinical characteristics of the disorder. In addition, case-control designs require immense effort and resources given that the high polygenicity of common psychiatric disorders requires vast sample sizes to detect effects[22].

Recent large-scale population level efforts such as the UK Biobank[23] now provide alternatives for the study of psychiatric disorders. The mental health data available in UK Biobank includes data from more than 150,000 individuals and covers questions on current and previous symptoms in different psychiatric

domains[24]. For example, a recent study revealed genetic associations with psychotic experiences in the UK Biobank and reported genetic correlations between psychotic experience and common psychiatric disorders[25]. While this study formed two groups of subjects (with and without psychotic experience) others have suggested continuous measures of psychopathology obtained from questionnaire data, such as the p-factor[26]. While bundling variance of psychopathology in a single common factor can be a useful proxy of mental health vulnerability, the specificity of the p-factor to disorder-specific mechanisms is limited[26]. Independent component analysis[27,28] provides a complementary approach to decompose the variance from mental health questionnaires into independent latent variables. For example, using independent component analysis on mental health questionnaires of children and adolescents, Alnæs and colleagues have identified a set of independent components reflecting symptoms of attention deficit, psychosis, depression, anxiety, and more[29]. Independent components obtained from mental health questionnaires may each capture either global (e.g. joint symptoms of depression, stress and anxiety) or specific aspects (e.g. pure psychosis symptoms) of mental health in a data-driven fashion, thereby yielding multiple distinct profiles of mental health symptoms beyond a common p-factor.

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

Methods and Materials

Sample and exclusion criteria

We accessed data from the UK Biobank[23] with permission no. 27412, and included data from individuals who had participated in an online follow-up interview on mental health (UK Biobank category 136). All participants provided signed informed consent before inclusion in the study. UK Biobank was approved by

112 the National Health Service National Research Ethics Service (ref. 11/NW/0382). Participants with a
113 diagnosed psychiatric or neurological disorder (F or G ICD10 diagnosis) were excluded from the analysis
114 except for those with a nerve, nerve root and plexus disorders (categories G50 to G59). In addition, we
115 excluded participants with more than 10% missing answers in the mental health questionnaires. This
116 resulted in mental health data from 136,678 individuals, which was used in an independent component
117 analysis. For the genetic analysis, we selected data from all Caucasian individuals with available genotypes,
118 yielding a set of 117,611 participants aged 47-80 years (mean: 64, SD: 7.66, age at mental health assessment)
119 and comprised 56.2% females.
120
121 *Processing of mental health data*
122 Fig. 1A depicts the analysis workflow. Of the 60 primary questions that were available as part of the mental
123 health online assessment, we removed questions that asked specifically about symptoms occurring in the
124 past two weeks to remove potential short-term temporal effects. Furthermore, we excluded questions where
125 more than 10% of the responses were missing (1 question excluded). In the resulting set of 43 questions
126 (Suppl. Table 1), we imputed missing data using k-nearest neighbor imputation with k = 3 with the bnstruct
127 package[30] in R[31] and z-standardized the data (Suppl. Fig. 2).
128
129 The resulting data covering 43 questions from 136,678 individuals was decomposed using independent
130 component analysis (ICA). Using icasso[32] in MATLAB and by visually inspecting the loadings of the
131 questions on the components, we estimated that a model order of 13 independent components yields the
132 best clustering solution where the resulting components are stable and highly interpretable. The PCA
133 identified 13 components with an eigenvalue larger than 1, and stability (Iq) was effectively 1. A model
134 order lower than 13 would group together questions into components which we preferred to keep separate.
135 A model order larger than 13 was not reasonable as it would yield components that largely reflect single
136 items. The individual scores for each of the 43 questions were subsequently residualized for age (both linear
137 and quadratic term), sex, and the first 20 genetic principal components. Next, we decomposed the items into
138 13 independent components using the fastICA algorithm as implemented in R[33]. Fig. 1B depicts how each
139 of the 43 items loaded on the components, indicating independent components (ICs) that captured questions
140 on sexual abuse (IC1), psychosis (IC2), anxiety, depression and mental distress (IC3), a diagnosis with a
141 life-threatening illness (IC4), social instability (IC5), traumatic experiences (IC6), stress in the past month
142 (IC7), experiences of feeling loved (IC8), thoughts around self-harm behavior (IC9), general happiness
143 (IC10), addiction behavior and manic experiences (IC11), experiences of emotional abuse (IC12), and
144 alcohol abuse (IC13).

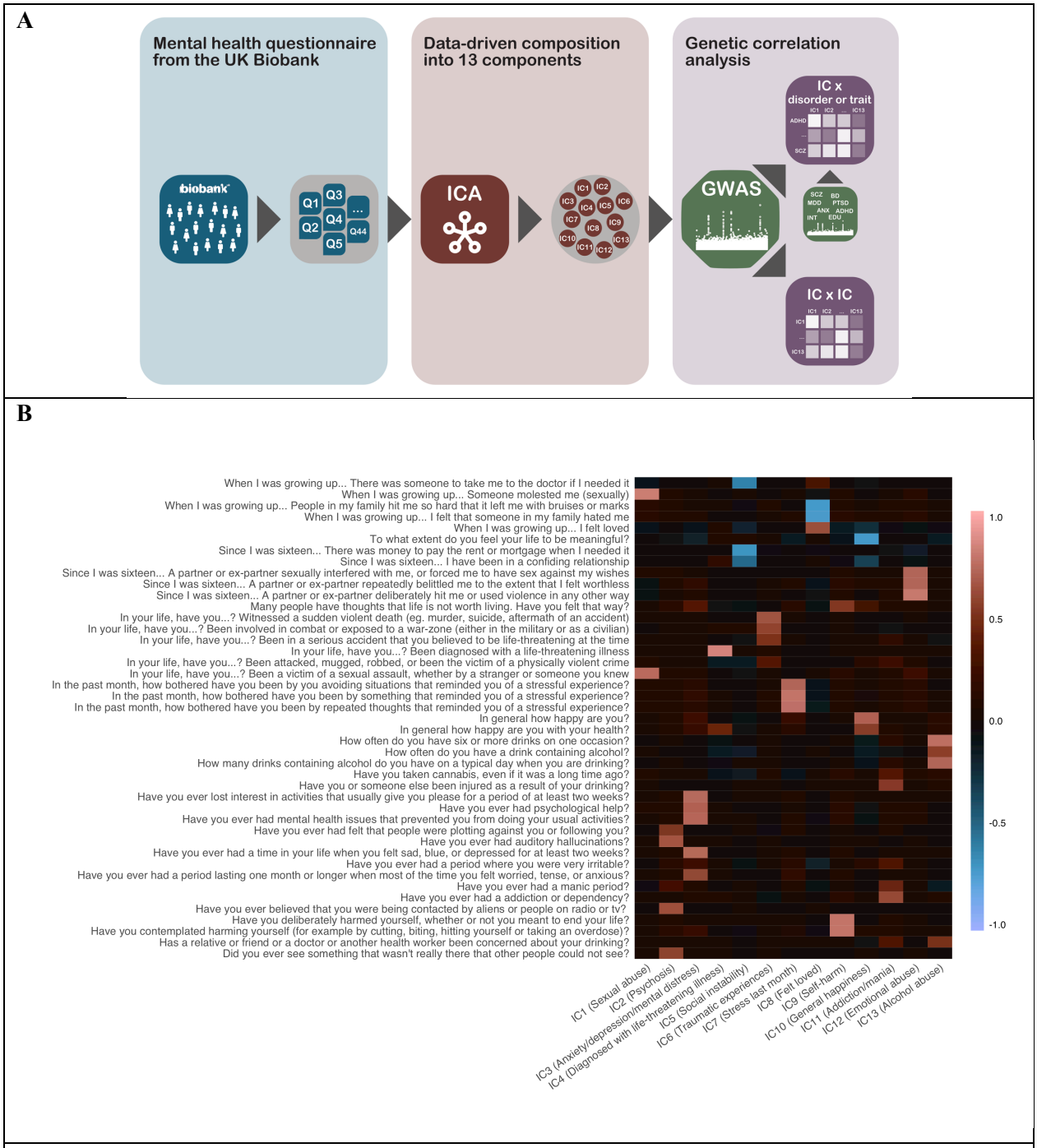


Fig. 1. Workflow and variable weight matrix of the resulting decomposition A Outline of the analysis workflow. **B** Weight matrix reflecting how each mental health question loaded on each IC. Brighter colors indicate higher loading, darker colors indicate lower loading. All 43 questions were captured 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).

146 The distribution of IC2 indicated very few non-zero scores (Suppl. Fig. 3). This component loaded mostly
147 on psychosis questions (Fig. 1B), indicating that only few of the included healthy individuals had symptoms
148 in this domain. We therefore conducted an additional supplemental analysis in which we dichotomized IC2
149 such that loadings lower than 1 were labeled as "no/few symptoms", and loadings equal to or higher than 1
150 were labeled as "with symptoms".
151
152 *Processing of genetic data*
153 From the UK Biobank v3 imputed genetic data, we removed SNPs with an imputation quality score below
154 0.5, with a minor allele frequency below 0.001, missing in more than 5% of individuals, and that failed the
155 Hardy-Weinberg equilibrium test at $p < 1e-9$. We removed also individuals with more than 10% missing
156 data. We performed a genome-wide association analysis (GWAS) on each of the 13 independent
157 components in PLINK 2[34,35]. Using a publicly available conversion toolbox for GWAS summary statistics
158 (github.com/precimed/python_convert), we removed the MHC region and calculated a z-score for every
159 SNP (8,165,726 SNPs after QC). We utilized linkage-disequilibrium score regression[10,36] to estimate genetic
160 correlations between each of the independent components, and between the components and publicly
161 available GWAS summary statistics for SCZ[1], BD[2], MD[37], ADHD[4], ASD[5], PTSD[6], ANX[7], as well as
162 intelligence[38], and educational attainment[39] (Suppl. Table 2). For all aforementioned GWASs, we used those
163 versions that did not have UK Biobank participants included. From the MD GWAS, we removed
164 participants from the 23andMe dataset as well, leaving only cases with a diagnosed major depressive
165 disorder (MDD). Prior to estimating genetic correlations, we set a threshold that only ICs with a heritability
166 1.96 times larger than its standard error should be included in the analysis and only those where visual
167 quality control of corresponding Q-Q plots indicated genetic signal. These quality control steps were
168 implemented to ensure that we did not make inferences on data that did not provide sufficient variance
169 explained by genetics. Partitioned heritability[40] was estimated using the LDSC toolbox[36] and Q-Q plots were
170 generated using custom scripts in R. Finally, we processed the GWAS summary statistics of each
171 independent component through the Functional Mapping and Annotation toolbox (FUMA) to map lead
172 SNPs onto genes[41]. FUMA parameters were kept as default, and we used the FUMA default European
173 ancestry reference panel.
174
175 *Code and data availability*
176 Code and GWAS summary statistics will be made publicly available via GitHub (github.com/norment) upon
177 acceptance of the manuscript. Furthermore, the derived independent components (individual level data) will
178 be made available to the UK Biobank upon acceptance (derived variable return) to allow its use in future
179 UK Biobank studies.
180

Results

Fig. 2 depicts SNP-based heritability (h2) for the 13 ICs (Suppl. Table 3 for additional statistics). Heritability was generally low, yet all components yielded a heritability that was higher than 1.96 times the standard error. IC13, capturing questions on alcohol abuse had the highest heritability (h2 = 0.0763, SE = 0.0055), closely followed by IC3, capturing anxiety, depression, and mental distress (h2 = 0.0744, SE = 0.0052). The lowest heritability among the components was for IC2, reflecting psychosis questions (h2 = 0.0089, SE = 0.0043), likely owing to the low number of individuals with psychosis symptoms (Suppl. Fig. 3). We therefore performed a supplemental analysis to investigate if dichotomization of this IC would benefit the analysis (Suppl. Fig. 4). In brief, as dichotomization only slightly improved heritability estimates, we kept IC2 as a continuous component for the main analysis to stay consistent with the other components, yet we provide results with the dichotomized component in Suppl. Fig. 4. In addition to passing the heritability criterion of 1.96 times the standard error, the Q-Q plots of all ICs passed visual quality control (Suppl. Fig. 5) warranting inclusion of all components into subsequent genetic correlation analyses.

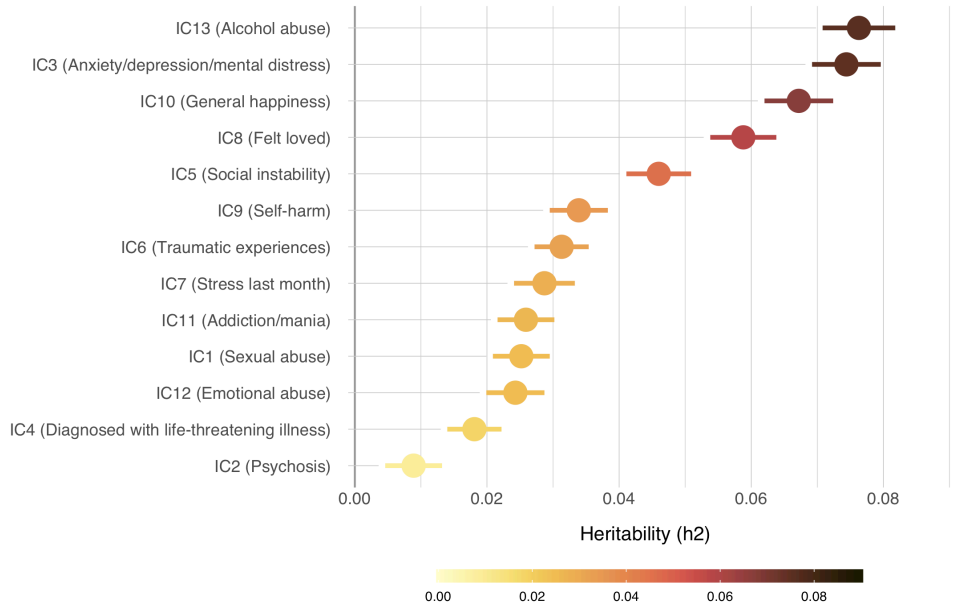


Fig. 2. Heritability estimates of the independent components SNP-based heritability for each IC sorted by decreasing heritability (h2). Heritability calculated using the LDSC toolbox[36]. Error bars reflect standard errors.

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

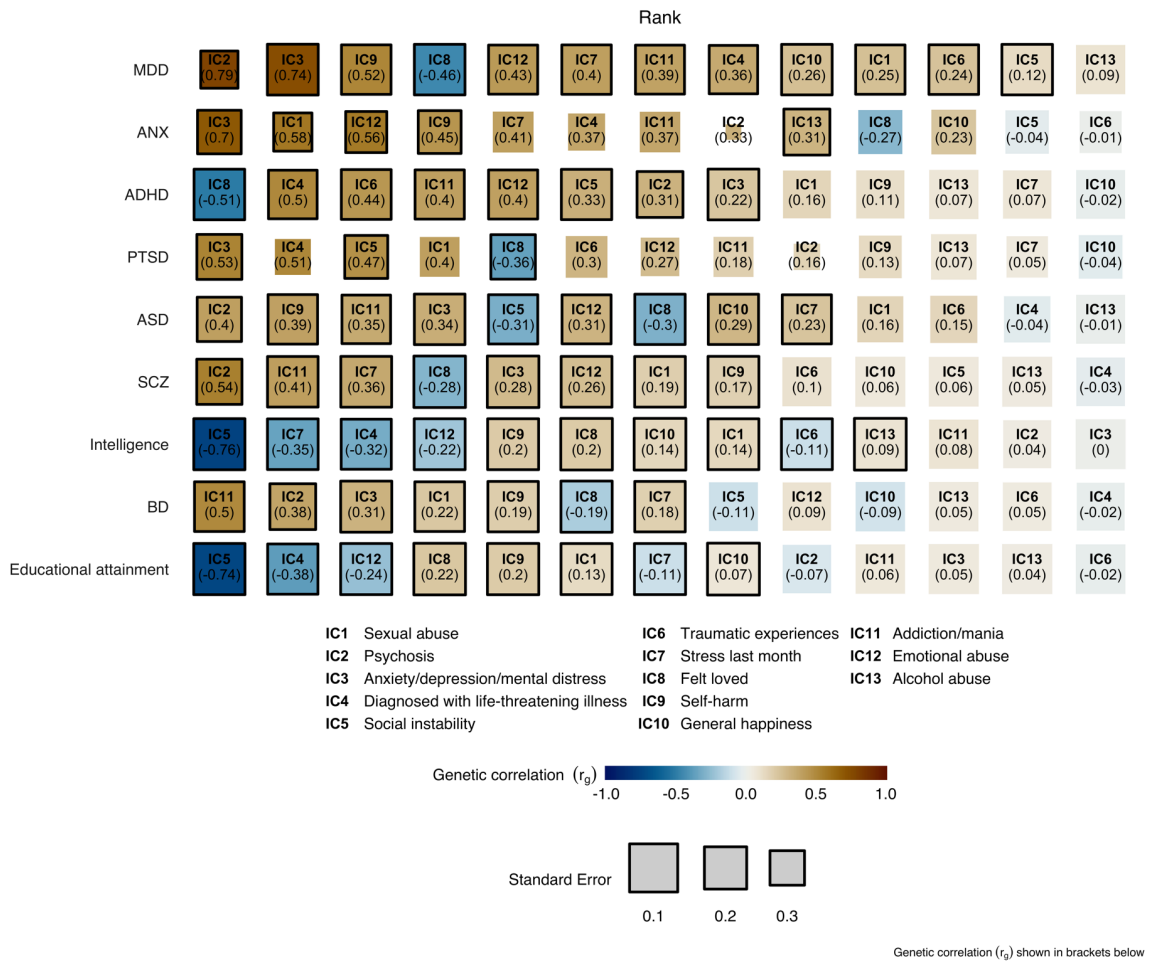


Fig. 3. Genetic correlation between the independent components and disorders and cognitive traits For each disorder, the associations with ICs are sorted by decreasing absolute genetic correlation such that the most leftward box reflects the strongest association between a given disorder and the 13 ICs.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

We assessed genetic correlations between each of the 13 ICs and a set of psychiatric disorders as well as cognitive traits. Out of 117 comparisons, 70 were significant after FDR correction, which amounts to 60%. Fig. 3 depicts all genetic correlations with ICs, sorted separately for each disorder or cognitive trait (sorted by absolute genetic correlation). We found that in most cases the strongest genetic correlation was with the IC most closely related to that disorder or trait. For example, anxiety most strongly correlated with IC3, which reflects anxiety, depression, and mental distress (genetic correlation $r_g = 0.70$, $p_{FDR} < .00027$). SCZ was most highly correlated with IC2, which represents psychosis questions ($r_g = 0.54$, $p_{FDR} = .001$). The highest genetic correlation of BD was with IC11, which represents addiction and mania ($r_g = 0.5$, $p_{FDR} = 6.5e\text{-}12$). For PTSD, the component reflecting traumatic experience (IC6) only ranked sixth among the sorted associations, yet the two ICs showing strongest association with PTSD reflected anxiety, depression, and mental distress (IC3; $r_g = 0.53$, $p_{FDR} = .0017$) and diagnosed with life-threatening illness (IC4; $r_g = 0.51$, $p_{FDR} = .080$), both of which are closely related to PTSD. ASD correlated strongest with IC2 (reflecting psychosis; $r_g = 0.40$, $p_{FDR} = .031$) and ADHD correlated strongest with IC8 (Felt loved; $r_g = -0.51$, $p_{FDR} = 4.7e\text{-}21$). Educational attainment and intelligence were both strongest negatively correlated with the IC reflecting social instability (IC5, $r_g = -0.74$ and $r_g = -0.76$, respectively; both $p_{FDR} < 2.5e\text{-}74$). In general, the strongest associations among all ICs, either positive or negative were with MDD while the weakest associations were with educational attainment.

Next, we assessed the genetic correlations between the ICs. Independent components are statistically independent by design, and thus on the phenotype level the ICs were not correlated with each other (Fig. 4, lower half; correlations essentially zero). However, approximately half of the IC pairs were nonetheless significantly genetically correlated with each other (51%, $p <$ FDR). IC3 (anxiety, depression, mental illness) was genetically correlated with 10 other ICs. IC9 (self-harm) was correlated with 9 other ICs and IC6 (traumatic experiences) and IC8 (felt loved) were each genetically correlated with eight other ICs. IC11 (addiction/mania) and IC12 (emotional abuse) were each genetically correlated with seven other ICs. IC1 (sexual abuse) and IC5 (social instability) were both genetically correlated with six other ICs. IC2 (psychosis) was correlated with 5 other ICs. IC4 (diagnosed with life-threatening illness) and IC13 (alcohol abuse) were both genetically correlated with 4 other ICs. And IC7 (stress last month) and IC10 (general happiness) were both genetically correlated with 3 other ICs. No IC had no significant genetic correlations with other ICs. The analysis therefore revealed a large amount of genetic correlations despite statistical (phenotypic) independence of the symptom profiles.

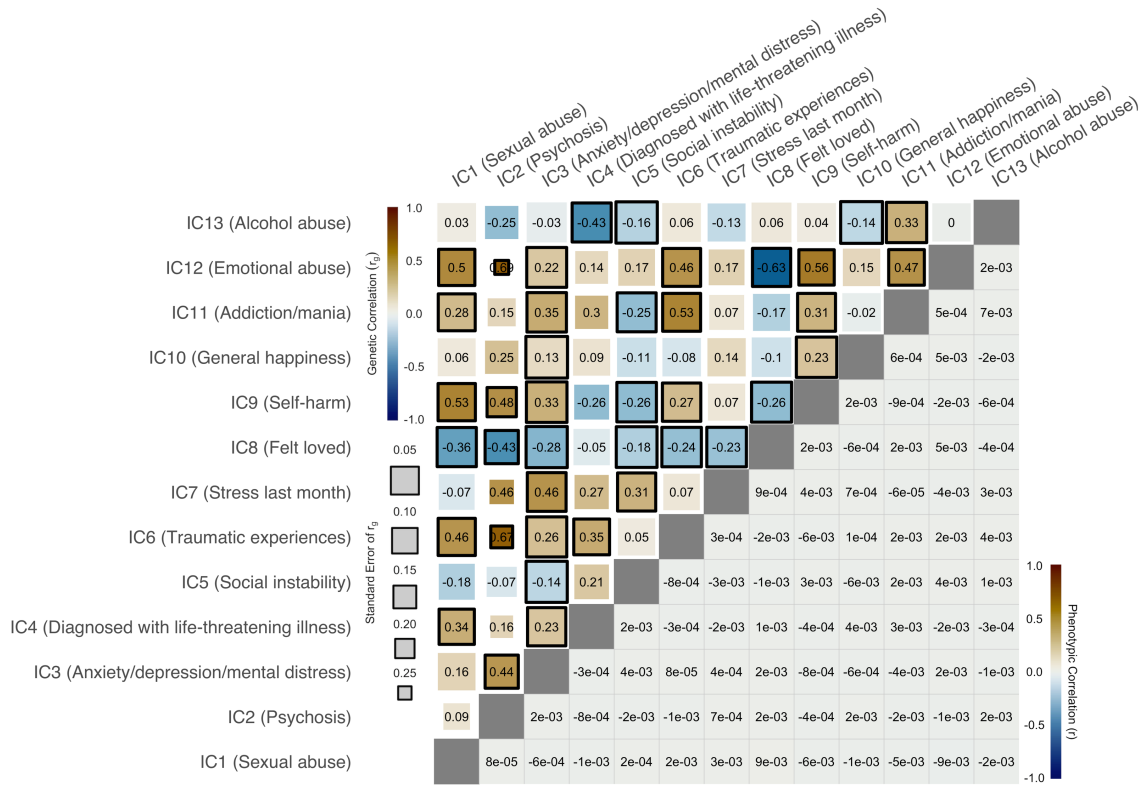


Fig. 4. Phenotypic and genetic correlation between the ICs The lower half of the IC by IC matrix depicts phenotypic correlations, reflecting the Pearson correlation of subject level component scores between independent components. As expected by ICA design, correlations were close to zero. The upper 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.

Discussion

In the present study, we decomposed mental health questionnaire data from more than 130,000 individuals into phenotypically distinct mental health profiles (independent components). We found that variations in mental health in healthy individuals (without a neurological or psychiatric diagnosis) were genetically correlated with psychiatric disorders and cognitive traits. Strongest correlations were observed between components and disorders with known symptoms in a similar domain (e.g. psychosis symptoms with schizophrenia), but the large amount of significant correlations between disorders and mental health profiles

suggested limited specificity. Indeed, we found a large proportion of significant genetic correlations between the phenotypically uncorrelated profiles, suggesting overlapping genetic architectures underlying distinct symptoms. The implications of our findings are twofold. First, our results support pleiotropy in psychiatric disorders beyond overlapping symptoms (e.g. BD and MDD both involving depressive episodes), suggesting that even distinct psychiatric symptoms are genetically overlapping. Second, our findings support that normal variability in mental health within healthy individuals may inform the study of the biology of psychiatric disorders.

While pleiotropy between major psychiatric disorders has been widely established[9–11] (reproduced in Suppl. Fig. 1), the sources underlying pleiotropy remain largely unknown. Specifically, disorders oftentimes overlap in symptomatology and therefore the degree to which the observed genetic correlations between disorders reflect phenotypic overlap between disorders remains to be investigated. Our approach of decomposing mental health data into distinct profiles allowed us to study genetic correlations in a sample with known phenotypic correlations and to assess how these profiles correlate with the genetics of different diagnoses. We observed that most disorders correlated strongest with the independent components capturing a related phenotype. For example, the strongest association with IC3, which reflects variance in anxiety, depression, and mental distress, was with ANX, the strongest association with IC2 (psychosis) was with SCZ. Therefore, the ranking of association strengths suggested a certain degree of specificity. However, that degree was strongly limited as most of the disorders and components were significantly genetically correlated. For example, MDD showed significant correlations between all but one component, ASD correlated with all but four components, and ANX and ADHD were correlated with all but 5 components, though correlation strengths were overall lower than with MDD, possibly due to lower sample size. There were also significant associations between components and cognitive traits although overall weaker associations compared to those with disorders. About half of the genetic correlations with intelligence and educational attainment pointed in the opposite direction, considerably more than for the psychiatric disorders, reflecting higher cognitive ability with fewer psychiatric symptoms. Importantly, when looking at the correlations between mental health profiles, we found that almost half of the genetic correlation matrix between ICs yielded significant genetic correlations despite a lack of phenotypic correlations (independence of the components). This suggests that some of the same genes are involved in the genetics of distinct mental health profiles and may indirectly support pleiotropy independent of phenotypic overlap in psychiatric disorders. Whereas more research is needed before conclusions on the sources underlying the observed pleiotropy can be drawn, one possible explanation for the significant correlations in the ICs could be that, since all independent components each capture a facet of mental health, there may be a number of SNPs that are involved across mental health symptoms. These SNPs may be involved in overall mental health, from psychological well-being to psychosis symptoms. Our analysis of significant SNPs in FUMA did not

identify overlapping SNPs between ICs, however, this may be attributed to the relatively low number of significant loci discovered in the ICs. Advanced statistical tools and further increasing sample sizes may help pinpoint specific genes involved with different symptoms. Furthermore, it is also plausible that environmental effects may factor into the explanation of the significant genetic correlations despite phenotypic independence if the environmental factors differ markedly between the ICs.

Limitations

Notable strengths of the present study include the use of data-driven decomposition of mental health data in a large sample of healthy individuals and its application to study pleiotropy in psychiatric genetics. Its main limitations include the low heritability of the resulting independent components, and the limited number of individuals with psychosis symptoms yielding suboptimal distribution in IC2 (Suppl. Fig. 3). First, it is important to note that all ICs passed quality control. Heritability of all ICs exceeded our pre-defined heritability threshold of 1.96 times its standard error, and all Q-Q plots indicated genetic signal (Suppl. Fig. 5). Furthermore, low heritability can still produce good genetic signal as a result from a low number of genetic variants involved but where each has large effects[13]. For example, while IC2 had the lowest heritability among the ICs, it showed one of the strongest genetic signal and together with IC7 and IC8 it ranked second in terms of the number of loci discovered in FUMA, following IC13 (alcohol abuse) that showed the highest heritability, strongest genetic signal on the Q-Q plot and the largest number of significant loci and mapped genes. Second, although sample size and symptom distributions factored into the results, these are mostly reflected in the standard error of genetic associations, not in a lack of effect. For example, ANX[7] (n = 21,761) and PTSD[6] (n = 9,537) GWASs have relatively little power as reflected in the larger standard errors in genetic correlations with these disorders, but nonetheless the strongest associations with these disorders were with components that match symptoms of the disorders (both correlated strongest with IC3, reflecting anxiety/depression/mental distress). Likewise, the suboptimal symptom distributions in IC2 and corresponding low heritability is reflected in large standard errors of the resulting genetic correlations but nonetheless IC2, reflecting psychosis, was most strongly associated with SCZ. Supplemental analysis with dichotomized IC2 also confirmed that the distribution alone is unlikely to explain the observed associations (Suppl. Fig. 4).

Conclusion

In the present study, we revealed genetic overlap between statistically independent mental health profiles and provide evidence that variations in mental health in healthy individuals relate genetically to psychiatric disorders and cognitive traits. These findings support that pleiotropy between psychiatric disorders cannot simply be explained by overlapping symptoms but may rather point to similar biological underpinnings of distinct symptoms. Our results underscore the potential of data-driven approaches to the study of mental

310 health, and suggests that supplementing the classic case-control design with a dimensional approach may
311 improve our understanding of the genetic underpinnings of complex disorders of the mind.
312

313 **Acknowledgements**
314 The authors were funded by the Research Council of Norway (#276082 LifespanHealth, #223273
315 NORMENT, #283798 ERA-NET Neuron SYNSCHIZ, #249795), the South-East Norway Regional Health
316 Authority (2019101), and the European Research Council under the European Union's Horizon2020
317 Research and Innovation program (ERC Starting Grant #802998), as well as the Horizon2020 Research and
318 Innovation Action Grant CoMorMent (#847776). This research has been conducted using the UK Biobank
319 Resource (access code 27412, https://www.ukbiobank.ac.uk/). This work was performed on the TSD
320 (Tjenester for Sensitive Data) facilities, owned by the University of Oslo, operated and developed by the
321 TSD service group at the University of Oslo, IT-Department (USIT). Computations were also performed on
322 resources provided by UNINETT Sigma2 - the National Infrastructure for High Performance Computing
323 and Data Storage in Norway.
324

325 **Conflicts of interest**
326 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
327 to HealthLytix and received speakers honorarium from Lundbeck.
328

329 **Author contributions**
330 D.R. and T.K. conceived the study; D.R. analyzed the data with contributions from T.K.; All authors
331 contributed with conceptual input on methods and/or interpretation of results; D.R. and T.K. wrote the first
332 draft of the paper and all authors contributed to the final manuscript.
333

334 **References**
335 1. Pardiñas AF, Holmans P, Pocklington AJ, et al. Common schizophrenia alleles are enriched in
336 mutation-intolerant genes and in regions under strong background selection. *Nat Genet*.
337 2018;50(3):381-389. doi:10.1038/s41588-018-0059-2
338 2. Stahl EA, Breen G, Forstner AJ, et al. Genome-wide association study identifies 30 loci associated
339 with bipolar disorder. *Nat Genet*. 2019;51(5):793-803. doi:10.1038/s41588-019-0397-8
340 3. Howard DM, Adams MJ, Clarke T-K, et al. Genome-wide meta-analysis of depression identifies
341 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat*
342 *Neurosci*. 2019;22(3):343-352. doi:10.1038/s41593-018-0326-7
343 4. Demontis D, Walters RK, Martin J, et al. Discovery of the first genome-wide significant risk loci
344 for attention deficit/hyperactivity disorder. *Nat Genet*. 2019;51(1):63-75. doi:10.1038/s41588-018-

0269-7

5. Grove J, Ripke S, Als TD, et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet*. 2019;51(3):431-444. doi:10.1038/s41588-019-0344-8
6. Duncan LE, Ratanatharathorn A, Aiello AE, et al. Largest GWAS of PTSD (N=20 070) yields genetic overlap with schizophrenia and sex differences in heritability. *Mol Psychiatry*. 2018;23(3):666-673. doi:10.1038/mp.2017.77
7. Otowa T, Hek K, Lee M, et al. Meta-analysis of genome-wide association studies of anxiety disorders. *Mol Psychiatry*. 2016;21(10):1391-1399. doi:10.1038/mp.2015.197
8. Andreassen OA, Thompson WK, Schork AJ, et al. Improved detection of common variants associated with schizophrenia and bipolar disorder using pleiotropy-informed conditional false discovery rate. *PLoS Genet*. 2013;9(4):e1003455. doi:10.1371/journal.pgen.1003455
9. Anttila V, Bulik-Sullivan B, Finucane HK, et al. Analysis of shared heritability in common disorders of the brain. *Science*. 2018;360(6395):eaap8757. doi:10.1126/science.aap8757
10. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet*. 2015;47(11):1236-1241. doi:10.1038/ng.3406
11. Smeland OB, Bahrami S, Frei O, et al. Genome-wide analysis reveals extensive genetic overlap between schizophrenia, bipolar disorder, and intelligence. *Mol Psychiatry*. January 2019:10.1038/s41380-018-0332-x. doi:10.1038/s41380-018-0332-x
12. Bansal V, Mitjans M, Burik CAP, et al. Genome-wide association study results for educational attainment aid in identifying genetic heterogeneity of schizophrenia. *Nat Commun*. 2018;9(1):3078. doi:10.1038/s41467-018-05510-z
13. Frei O, Holland D, Smeland OB, et al. Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation. *Nat Commun*. 2019;10(1):2417-2417. doi:10.1038/s41467-019-10310-0
14. Bahrami S, Steen NE, Shadrin A, et al. Shared Genetic Loci Between Body Mass Index and Major Psychiatric Disorders: A Genome-wide Association Study. *JAMA Psychiatry*. January 2020:10.1001/jamapsychiatry.2019.4188. doi:10.1001/jamapsychiatry.2019.4188
15. Amare AT, Schubert KO, Klingler-Hoffmann M, Cohen-Woods S, Baune BT. The genetic overlap between mood disorders and cardiometabolic diseases: a systematic review of genome wide and candidate gene studies. *Transl Psychiatry*. 2017;7(1):e1007-e1007. doi:10.1038/tp.2016.261
16. Andreasen NC. A unitary model of schizophrenia: Bleuler's "fragmented phrene" as schizencephaly. *Arch Gen Psychiatry*. 1999;56(9):781-787. doi:10.1001/archpsyc.56.9.781
17. Craddock N, Owen MJ. The Kraepelinian dichotomy - going, going... but still not gone. *Br J Psychiatry J Ment Sci*. 2010;196(2):92-95. doi:10.1192/bjp.bp.109.073429
18. Alnaes D, Kaufmann T, van der Meer D, et al. Brain Heterogeneity in Schizophrenia and Its

380 Association With Polygenic Risk. *JAMA Psychiatry*. 2019;76(7):739-748.
381 doi:10.1001/jamapsychiatry.2019.0257
382 19. Hibar DP, Westlye LT, Doan NT, et al. Cortical abnormalities in bipolar disorder: an MRI analysis
383 of 6503 individuals from the ENIGMA Bipolar Disorder Working Group. *Mol Psychiatry*.
384 2018;23(4):932-942. doi:10.1038/mp.2017.73
385 20. McGrath JJ, Saha S, Al-Hamzawi A, et al. Psychotic Experiences in the General Population: A
386 Cross-National Analysis Based on 31,261 Respondents From 18 Countries. *JAMA Psychiatry*.
387 2015;72(7):697-705. doi:10.1001/jamapsychiatry.2015.0575
388 21. Pirastu N, Cordioli M, Nandakumar P, et al. Genetic analyses identify widespread sex-differential
389 participation bias. *bioRxiv*. January 2020:2020.03.22.001453. doi:10.1101/2020.03.22.001453
390 22. Holland D, Frei O, Desikan R, et al. Beyond SNP Heritability: Polygenicity and Discoverability of
391 Phenotypes Estimated with a Univariate Gaussian Mixture Model. *bioRxiv*. January 2019:133132.
392 doi:10.1101/133132
393 23. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the
394 causes of a wide range of complex diseases of middle and old age. *PLoS Med*.
395 2015;12(3):e1001779. doi:10.1371/journal.pmed.1001779
396 24. Davis KAS, Coleman JRI, Adams M, et al. Mental health in UK Biobank – development,
397 implementation and results from an online questionnaire completed by 157 366 participants: a
398 reanalysis. *BJPsych Open*. 2020;6(2):e18. doi:10.1192/bjo.2019.100
399 25. Legge SE, Jones HJ, Kendall KM, et al. Association of Genetic Liability to Psychotic Experiences
400 With Neuropsychotic Disorders and Traits. *JAMA Psychiatry*. 2019;76(12):1256-1265.
401 doi:10.1001/jamapsychiatry.2019.2508
402 26. Caspi A, Houts RM, Belsky DW, et al. The p Factor: One General Psychopathology Factor in the
403 Structure of Psychiatric Disorders? *Clin Psychol Sci J Assoc Psychol Sci*. 2014;2(2):119-137.
404 doi:10.1177/2167702613497473
405 27. Jutten C, Hérault J. Détection de grandeurs primitives dans un message composite par une
406 architecture de calcul neuromimétique en apprentissage non supervisé. *10° Colloq Sur Trait Signal*
407 *Images FRA 1985*. 1985:1017-1022.
408 28. Comon P. Independent component analysis, A new concept? *High Order Stat*. 1994;36(3):287-314.
409 doi:10.1016/0165-1684(94)90029-9
410 29. Alnæs D, Kaufmann T, Doan NT, et al. Association of Heritable Cognitive Ability and
411 Psychopathology With White Matter Properties in Children and Adolescents. *JAMA Psychiatry*.
412 2018;75(3):287-295. doi:10.1001/jamapsychiatry.2017.4277
413 30. Alberto Franzin, Francesco Sambo, Barbara di Camillo. bnstruct: an R package for Bayesian
414 Network structure learning in the presence of missing data. *Bioinformatics*. 2017;33(8):1250-1252.

415 doi:10.1093/bioinformatics/btw807
416 31. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R
417 Foundation for Statistical Computing; 2017. www.R-project.org.
418 32. Himberg J, Hyvarinen A. Icasso: software for investigating the reliability of ICA estimates by
419 clustering and visualization. *2003 IEEE XIII Workshop Neural Netw Signal Process IEEE Cat*
420 *No03TH8718*. 2003:259-268. doi:10.1109/NNSP.2003.1318025
421 33. Marchini J, Heaton C, Ripley B. *FastICA: FastICA Algorithms to Perform ICA and Projection*
422 *Pursuit*.; 2019. https://CRAN.R-project.org/package=fastICA.
423 34. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK:
424 rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7. doi:10.1186/s13742-
425 015-0047-8
426 35. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and
427 population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-575. doi:10.1086/519795
428 36. Bulik-Sullivan BK, Loh PR, Finucane HK, et al. LD Score regression distinguishes confounding
429 from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47(3):291-295.
430 doi:10.1038/ng.3211
431 37. Wray NR, Ripke S, Mattheisen M, et al. Genome-wide association analyses identify 44 risk variants
432 and refine the genetic architecture of major depression. *Nat Genet*. 2018;50(5):668-681.
433 doi:10.1038/s41588-018-0090-3
434 38. Savage JE, Jansen PR, Stringer S, et al. Genome-wide association meta-analysis in 269,867
435 individuals identifies new genetic and functional links to intelligence. *Nat Genet*. 2018;50(7):912-
436 919. doi:10.1038/s41588-018-0152-6
437 39. Lee JJ, Wedow R, Okbay A, et al. Gene discovery and polygenic prediction from a genome-wide
438 association study of educational attainment in 1.1 million individuals. *Nat Genet*. 2018;50(8):1112-
439 1121. doi:10.1038/s41588-018-0147-3
440 40. Finucane HK, Bulik-Sullivan B, Gusev A, et al. Partitioning heritability by functional annotation
441 using genome-wide association summary statistics. *Nat Genet*. 2015;47(11):1228-1235.
442 doi:10.1038/ng.3404
443 41. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of
444 genetic associations with FUMA. *Nat Commun*. 2017;8(1):1826. doi:10.1038/s41467-017-01261-5
445
446 **Supplementary Tables**
447

no	question	coding
1	In your life, did you seek or receive help from a professional (medical doctor, psychologist, social worker, counsellor, nurse, clergy, or other helping professional) for mental distress, psychological problems or unusual experiences?	502
2	In your life, have you suffered from a period of mental distress that prevented you from doing your usual activities?	502
3	Have you ever had a time in your life lasting two weeks or more when you lost interest in most things like hobbies, work, or activities that usually give you pleasure?	503
4	Have you ever had a time in your life when you felt sad, blue, or depressed for two weeks or more in a row?	503
5	Have you ever had a period of time when you were feeling so good, high, excited, or hyper that other people thought you were not your normal self or you were so hyper that you got into trouble?	502
6	Have you ever had a period of time when you were so irritable that you found yourself shouting at people or starting fights or arguments?	502
7	Have you ever had a period lasting one month or longer when most of the time you felt worried, tense, or anxious?	502
8	Have you been addicted to or dependent on one or more things, including substances (not cigarettes/coffee) or behaviours (such as gambling)?	502
9	In the next two questions, a drink is defined as one unit of alcohol. How many drinks containing alcohol do you have on a typical day when you are drinking?	522
10	Has a relative or friend or a doctor or another health worker been concerned about your drinking or suggested you cut down?	524
11	Have you or someone else been injured as a result of your drinking?	524
12	How often do you have a drink containing alcohol?	521
13	In the next two questions, a drink is defined as one unit of alcohol. How often do you have six or more drinks on one occasion?	523
14	Have you taken CANNABIS (marijuana, grass, hash, ganja, blow, draw, skunk, weed, spliff, dope), even if it was a long time ago?	526
15	Did you ever hear things that other people said did not exist, like strange voices coming from inside your head talking to you or about you, or voices coming out of the air when there was no one around?	502

no	question	coding
16	Did you ever believe that there was an unjust plot going on to harm you or to have people follow you, and which your family and friends did not believe existed?	502
17	Did you ever see something that wasn't really there that other people could not see?	502
18	Did you ever believe that a strange force was trying to communicate directly with you by sending special signs or signals that you could understand but that no one else could understand (for example through the radio or television)?	502
19	When I was growing up... I felt that someone in my family hated me	532
20	When I was growing up... People in my family hit me so hard that it left me with bruises or marks	532
21	When I was growing up... I felt loved	532
22	When I was growing up... Someone molested me (sexually)	532
23	When I was growing up... There was someone to take me to the doctor if I needed it	532
24	Next is a list of problems and complaints that people sometimes have in response to such extremely stressful experiences. Please indicate how much you have been bothered by that problem in the past month: Avoiding activities or situations because they reminded you of a stressful experience?	534
25	Next is a list of problems and complaints that people sometimes have in response to such extremely stressful experiences. Please indicate how much you have been bothered by that problem in the past month: Repeated, disturbing memories, thoughts, or images of a stressful experience?	534
26	Next is a list of problems and complaints that people sometimes have in response to such extremely stressful experiences. Please indicate how much you have been bothered by that problem in the past month: Feeling very upset when something reminded you of a stressful experience?	534
27	Since I was sixteen... A partner or ex-partner repeatedly belittled me to the extent that I felt worthless	532
28	Since I was sixteen... I have been in a confiding relationship	532
29	Since I was sixteen... A partner or ex-partner deliberately hit me or used violence in any other way	532
30	Since I was sixteen... A partner or ex-partner sexually interfered with me, or forced me to have sex against my wishes	532
31	Since I was sixteen... There was money to pay the rent or mortgage when I needed it	532

no	question	coding
32	In your life, have you...? Been in a serious accident that you believed to be life-threatening at the time	533
33	In your life, have you...? Been involved in combat or exposed to a war-zone (either in the military or as a civilian)	533
34	In your life, have you...? Been diagnosed with a life-threatening illness	533
35	In your life, have you...? Been attacked, mugged, robbed, or been the victim of a physically violent crime	533
36	In your life, have you...? Witnessed a sudden violent death (eg. murder, suicide, aftermath of an accident)	533
37	In your life, have you...? Been a victim of a sexual assault, whether by a stranger or someone you knew	533
38	Many people have thoughts that life is not worth living. Have you felt that way?	535
39	Have you deliberately harmed yourself, whether or not you meant to end your life?	503
40	Have you contemplated harming yourself (for example by cutting, biting, hitting yourself or taking an overdose)?	535
41	In general how happy are you?	537
42	In general how happy are you with your HEALTH?	537
43	To what extent do you feel your life to be meaningful?	538

no	question	coding

Suppl. Table 1. Full list of questions This table lists the questions included in this study as presented to the participant. Data was available from 157,352 individuals, but through exclusion criteria 136.678 individuals remained. There are in total 43 questions. The third column indicates the coding. Possible answers for each coding are as follows:

502 – Prefer not to answer/Do not know/No/Yes

503 – Prefer not to answer/No/Yes

521 – Prefer not to answer/Never/Monthly or less/2 to 4 times a month/2 to 3 times a month/4 or more times a week

522 – Prefer not to answer/1 or 2/3 or 4/5 or 6/7,8 or 9/10 or more

523 – Prefer not to anwer/Never/Less than monthly/Monthly/Weekly/Daily or almost daily

524 – Prefer not to answer/Yes, but not in the last year/Yes, during the last year

526 – Prefer not to answer/No/Yes, 1-2 times/Yes, 3-10 times/Yes, 11-100 times/Yes, more than 100 times

532 – Prefer not to answer/Never true/Rarely true/Sometimes true/Often/Very often true

533 – Prefer not to answer/Never/Yes, but not in the last 12 months/Yes, within the last 12 months

534 – Prefer not to answer/Not at all/A little bit/Moderately/Quite a bit/Extremely

535 – Prefer not to answer/No/Yes, once/Yes, more than once

537 – Prefer not to answer/Do not know/Extremely happy/Very happy/Moderately happy/Moderately unhappy/Very unhappy/Extremely unhappy

538 – Prefer not to answer/Do not know/Not at all/A little/A moderate amount/Very much/An extreme amount

Phenotype	Consortium	Sample	Citation	n_{case}	$n_{control}$
SCZ	PGC	Meta	Pardiñas et al., 2018	40,675	64,643
BD	PGC	European	Stahl et al., 2019	20,352	31,358
MDD	PGC	European	Wray et al., 2018	69,576	161,613
ADHD	PGC	European	Demontis et al., 2019	19,099	34,194
ASD	PGC, iPSYCH	Meta	Grove et al., 2019	18,381	27,969
PTSD	PGC	European	Duncan et al., 2018	2,424	7,113
ANX	ANGST	Meta	Otowa et al., 2016	7,016	14,745
Intelligence	CTG	Meta	Savage et al., 2018	269,867	
Educational attainment	SSGAC	European	Lee et al., 2018	766,345	

Suppl. Table 2. Cohort Overview Overview of the cohorts including references and sample size.

450

IC	Component represents	h2 (SE)	Lambda GC	Intercept (SE)
IC1	Sexual abuse	0.0252 (0.0043)	1.0557	1.0023 (0.0063)
IC2	Psychosis	0.0089 (0.0043)	1.0225	1.0034 (0.007)
IC3	Anxiety/depression/mental distress	0.0744 (0.0052)	1.1523	0.9993 (0.0074)
IC4	Diagnosed with life-threatening illness	0.0181 (0.0041)	1.0557	1.0103 (0.0066)
IC5	Social instability	0.046 (0.0049)	1.0988	1.0138 (0.0067)
IC6	Traumatic experiences	0.0313 (0.0041)	1.0649	1.0021 (0.0065)
IC7	Stress last month	0.0287 (0.0046)	1.0557	0.9978 (0.0068)
IC8	Felt loved	0.0588 (0.005)	1.1301	1.0058 (0.0067)
IC9	Self-harm	0.0339 (0.0044)	1.0741	1.0017 (0.0059)
IC10	General happiness	0.0672 (0.0052)	1.1555	1.0239 (0.0074)
IC11	Addiction/mania	0.0259 (0.0043)	1.0710	1.0085 (0.0062)
IC12	Emotional abuse	0.0243 (0.0044)	1.0527	0.996 (0.0062)
IC13	Alcohol abuse	0.0763 (0.0055)	1.1587	1.0055 (0.0074)

Suppl. Table 3. Heritability statistics from LDSC Heritability estimates and additional statistics on the ICs

451

IC	Component represents	no	Gene	Chr	pmin	Individual Significant SNP
IC1	Sexual abuse	1	ADARB2	10	3.74966e-08	rs554213191
IC1	Sexual abuse	2	ADARB2	10	3.74966e-08	rs554213191
IC2	Psychosis	1	GAS2L3	12	4.49537e-08	rs540594149
IC2	Psychosis	2	ANO4	12	7.47504e-07	rs540594149
IC2	Psychosis	3	FAM109A	12	1.14947e-06	rs749014627
IC2	Psychosis	4	SH2B3	12	5.60159e-07	rs749014627
IC2	Psychosis	5	ATXN2	12	5.60159e-07	rs749014627
IC2	Psychosis	6	ACAD10	12	1.24816e-06	rs749014627
IC2	Psychosis	7	RP11-162P23.2	12	1.24816e-06	rs749014627
IC2	Psychosis	8	ALDH2	12	1.32971e-06	rs749014627
IC2	Psychosis	9	MAPKAPK5	12	5.77870e-07	rs749014627
IC2	Psychosis	10	TMEM116	12	1.19305e-05	rs749014627
IC2	Psychosis	11	HECTD4	12	4.04123e-06	rs749014627

IC	Component represents	no	Gene	Chr	pmin	Individual Significant SNP
IC2	Psychosis	12	GAS2L3	12	4.49537e-08	rs540594149
IC2	Psychosis	13	ANO4	12	7.47504e-07	rs540594149
IC2	Psychosis	14	FAM109A	12	1.14947e-06	rs749014627
IC2	Psychosis	15	SH2B3	12	5.60159e-07	rs749014627
IC2	Psychosis	16	ATXN2	12	5.60159e-07	rs749014627
IC2	Psychosis	17	ACAD10	12	1.24816e-06	rs749014627
IC2	Psychosis	18	RP11-162P23.2	12	1.24816e-06	rs749014627
IC2	Psychosis	19	ALDH2	12	1.32971e-06	rs749014627
IC2	Psychosis	20	MAPKAPK5	12	5.77870e-07	rs749014627
IC2	Psychosis	21	TMEM116	12	1.19305e-05	rs749014627
IC2	Psychosis	22	HECTD4	12	4.04123e-06	rs749014627
IC3	Anxiety/depression/mental distress	1	CABP1	12	4.21106e-09	rs73222787
IC3	Anxiety/depression/mental distress	2	MLEC	12	1.62854e-06	rs73222787
IC3	Anxiety/depression/mental distress	3	UNC119B	12	2.60327e-06	rs73222787
IC3	Anxiety/depression/mental distress	4	ACADS	12	1.64855e-06	rs73222787
IC3	Anxiety/depression/mental distress	5	SPPL3	12	2.82900e-06	rs73222787
IC3	Anxiety/depression/mental distress	6	CABP1	12	4.21106e-09	rs73222787
IC3	Anxiety/depression/mental distress	7	MLEC	12	1.62854e-06	rs73222787
IC3	Anxiety/depression/mental distress	8	UNC119B	12	2.60327e-06	rs73222787
IC3	Anxiety/depression/mental distress	9	ACADS	12	1.64855e-06	rs73222787
IC3	Anxiety/depression/mental distress	10	SPPL3	12	2.82900e-06	rs73222787
IC5	Social instability	1	AP1G1	16	3.00910e-06	rs8057124
IC5	Social instability	2	ATXN1L	16	1.83621e-08	rs8057124
IC5	Social instability	3	IST1	16	1.04493e-08	rs8057124
IC5	Social instability	4	ZNF821	16	2.05589e-08	rs8057124
IC5	Social instability	5	AP1G1	16	3.00910e-06	rs8057124
IC5	Social instability	6	ATXN1L	16	1.83621e-08	rs8057124
IC5	Social instability	7	IST1	16	1.04493e-08	rs8057124
IC5	Social instability	8	ZNF821	16	2.05589e-08	rs8057124
IC7	Stress last month	1	ACTN1	14	1.10472e-05	rs761897
IC7	Stress last month	2	ACTN1	14	1.10472e-05	rs761897

IC	Component represents	no	Gene	Chr	pmin	Individual Significant SNP
IC8	Felt loved	1	EYS	6	4.70962e-08	rs183356400
IC8	Felt loved	2	RBFOX1	16	4.39126e-08	rs13332228
IC8	Felt loved	3	EYS	6	4.70962e-08	rs183356400
IC8	Felt loved	4	RBFOX1	16	4.39126e-08	rs13332228
IC10	General happiness	1	CSMD1	8	4.77632e-08	rs2554644
IC10	General happiness	2	CSMD1	8	4.77632e-08	rs2554644
IC11	Addiction/mania	1	CADM2	3	3.68665e-08	rs9866089
IC11	Addiction/mania	2	CADM2	3	3.68665e-08	rs9866089
IC12	Emotional abuse	1	PTPRT	20	1.63897e-08	rs533568724
IC12	Emotional abuse	2	PTPRT	20	1.63897e-08	rs533568724
IC13	Alcohol abuse	1	EIF2B4	2	2.34369e-08	rs1260326
IC13	Alcohol abuse	2	SNX17	2	2.34369e-08	rs1260326
IC13	Alcohol abuse	3	ZNF513	2	2.34369e-08	rs1260326
IC13	Alcohol abuse	4	PPM1G	2	2.34369e-08	rs1260326
IC13	Alcohol abuse	5	GCKR	2	1.70337e-10	rs1260326
IC13	Alcohol abuse	6	AC109829.1	2	1.16193e-05	rs1260326
IC13	Alcohol abuse	7	SIX3	2	2.32692e-07	rs528301 rs504675
IC13	Alcohol abuse	8	RFC1	4	4.44299e-08	rs12643682
IC13	Alcohol abuse	9	KLB	4	8.52705e-17	rs12643682 rs58015370 rs13125440 rs6836420
IC13	Alcohol abuse	10	TSPAN5	4	4.55114e-10	rs4699663
IC13	Alcohol abuse	11	METAP1	4	1.26232e-13	rs146788033
IC13	Alcohol abuse	12	ADH5	4	3.35425e-25	rs145452708
IC13	Alcohol abuse	13	ADH6	4	1.73597e-16	rs11733695
IC13	Alcohol abuse	14	ADH1B	4	5.89840e-52	rs1229984 rs145452708
IC13	Alcohol abuse	15	C4orf17	4	5.76240e-15	rs543669349
IC13	Alcohol abuse	16	TRMT10A	4	5.76240e-15	rs543669349

IC	Component represents	no	Gene	Chr	pmin	Individual Significant SNP
IC13	Alcohol abuse	17	DAPP1	4	NA	rs543669349 rs188514326
IC13	Alcohol abuse	18	LAMTOR3	4	NA	rs543669349 rs188514326
IC13	Alcohol abuse	19	DNAJB14	4	8.67314e-10	rs543669349 rs188514326
IC13	Alcohol abuse	20	BANK1	4	6.49446e-08	rs13107325
IC13	Alcohol abuse	21	SLC39A8	4	1.26159e-13	rs13107325 rs34333163
IC13	Alcohol abuse	22	ARHGAP27	17	4.06505e-09	rs62062288
IC13	Alcohol abuse	23	PLEKHM1	17	1.02001e-13	rs62062288 rs62053943
IC13	Alcohol abuse	24	CRHR1	17	1.00879e-13	rs62062288 rs62053943 rs9303521 rs12944712
IC13	Alcohol abuse	25	SPPL2C	17	1.27409e-13	rs62062288 rs62053943 rs12944712
IC13	Alcohol abuse	26	MAPT	17	2.31437e-14	rs62062288 rs62053943
IC13	Alcohol abuse	27	STH	17	1.94674e-13	rs62062288 rs62053943
IC13	Alcohol abuse	28	KANSL1	17	5.27802e-14	rs62062288 rs62053943
IC13	Alcohol abuse	29	ARL17B	17	1.06779e-13	rs62062288 rs62053943
IC13	Alcohol abuse	30	LRRC37A	17	1.06779e-13	rs62062288 rs62053943
IC13	Alcohol abuse	31	NSF	17	9.06747e-13	rs62062288 rs62053943

IC	Component represents	no	Gene	Chr	pmin	Individual Significant SNP
IC13	Alcohol abuse	32	WNT3	17	7.33108e-12	rs62062288
IC13	Alcohol abuse	33	FUT2	19	2.24286e-08	rs28894750
IC13	Alcohol abuse	34	MAMSTR	19	2.24286e-08	rs28894750
IC13	Alcohol abuse	35	RASIP1	19	4.60338e-08	rs28894750
IC13	Alcohol abuse	36	IZUMO1	19	2.76726e-06	rs28894750
IC13	Alcohol abuse	37	FUT1	19	2.76726e-06	rs28894750
IC13	Alcohol abuse	38	EIF2B4	2	2.34369e-08	rs1260326
IC13	Alcohol abuse	39	SNX17	2	2.34369e-08	rs1260326
IC13	Alcohol abuse	40	ZNF513	2	2.34369e-08	rs1260326
IC13	Alcohol abuse	41	PPM1G	2	2.34369e-08	rs1260326
IC13	Alcohol abuse	42	GCKR	2	1.70337e-10	rs1260326
IC13	Alcohol abuse	43	AC109829.1	2	1.16193e-05	rs1260326
IC13	Alcohol abuse	44	SIX3	2	2.32692e-07	rs528301 rs504675
IC13	Alcohol abuse	45	RFC1	4	4.44299e-08	rs12643682
IC13	Alcohol abuse	46	KLB	4	8.52705e-17	rs12643682 rs58015370 rs13125440 rs6836420
IC13	Alcohol abuse	47	TSPAN5	4	4.55114e-10	rs4699663
IC13	Alcohol abuse	48	METAP1	4	1.26232e-13	rs146788033
IC13	Alcohol abuse	49	ADH5	4	3.35425e-25	rs145452708
IC13	Alcohol abuse	50	ADH6	4	1.73597e-16	rs11733695
IC13	Alcohol abuse	51	ADH1B	4	5.89840e-52	rs1229984 rs145452708
IC13	Alcohol abuse	52	C4orf17	4	5.76240e-15	rs543669349
IC13	Alcohol abuse	53	TRMT10A	4	5.76240e-15	rs543669349
IC13	Alcohol abuse	54	DAPP1	4	NA	rs543669349 rs188514326
IC13	Alcohol abuse	55	LAMTOR3	4	NA	rs543669349 rs188514326

IC	Component represents	no	Gene	Chr	pmin	Individual Significant SNP
IC13	Alcohol abuse	56	DNAJB14	4	8.67314e-10	rs543669349 rs188514326
IC13	Alcohol abuse	57	BANK1	4	6.49446e-08	rs13107325
IC13	Alcohol abuse	58	SLC39A8	4	1.26159e-13	rs13107325 rs34333163
IC13	Alcohol abuse	59	ARHGAP27	17	4.06505e-09	rs62062288
IC13	Alcohol abuse	60	PLEKHM1	17	1.02001e-13	rs62062288 rs62053943
IC13	Alcohol abuse	61	CRHR1	17	1.00879e-13	rs62062288 rs62053943 rs9303521 rs12944712
IC13	Alcohol abuse	62	SPPL2C	17	1.27409e-13	rs62062288 rs62053943 rs12944712
IC13	Alcohol abuse	63	MAPT	17	2.31437e-14	rs62062288 rs62053943
IC13	Alcohol abuse	64	STH	17	1.94674e-13	rs62062288 rs62053943
IC13	Alcohol abuse	65	KANSL1	17	5.27802e-14	rs62062288 rs62053943
IC13	Alcohol abuse	66	ARL17B	17	1.06779e-13	rs62062288 rs62053943
IC13	Alcohol abuse	67	LRRC37A	17	1.06779e-13	rs62062288 rs62053943
IC13	Alcohol abuse	68	NSF	17	9.06747e-13	rs62062288 rs62053943
IC13	Alcohol abuse	69	WNT3	17	7.33108e-12	rs62062288
IC13	Alcohol abuse	70	FUT2	19	2.24286e-08	rs28894750
IC13	Alcohol abuse	71	MAMSTR	19	2.24286e-08	rs28894750
IC13	Alcohol abuse	72	RASIP1	19	4.60338e-08	rs28894750

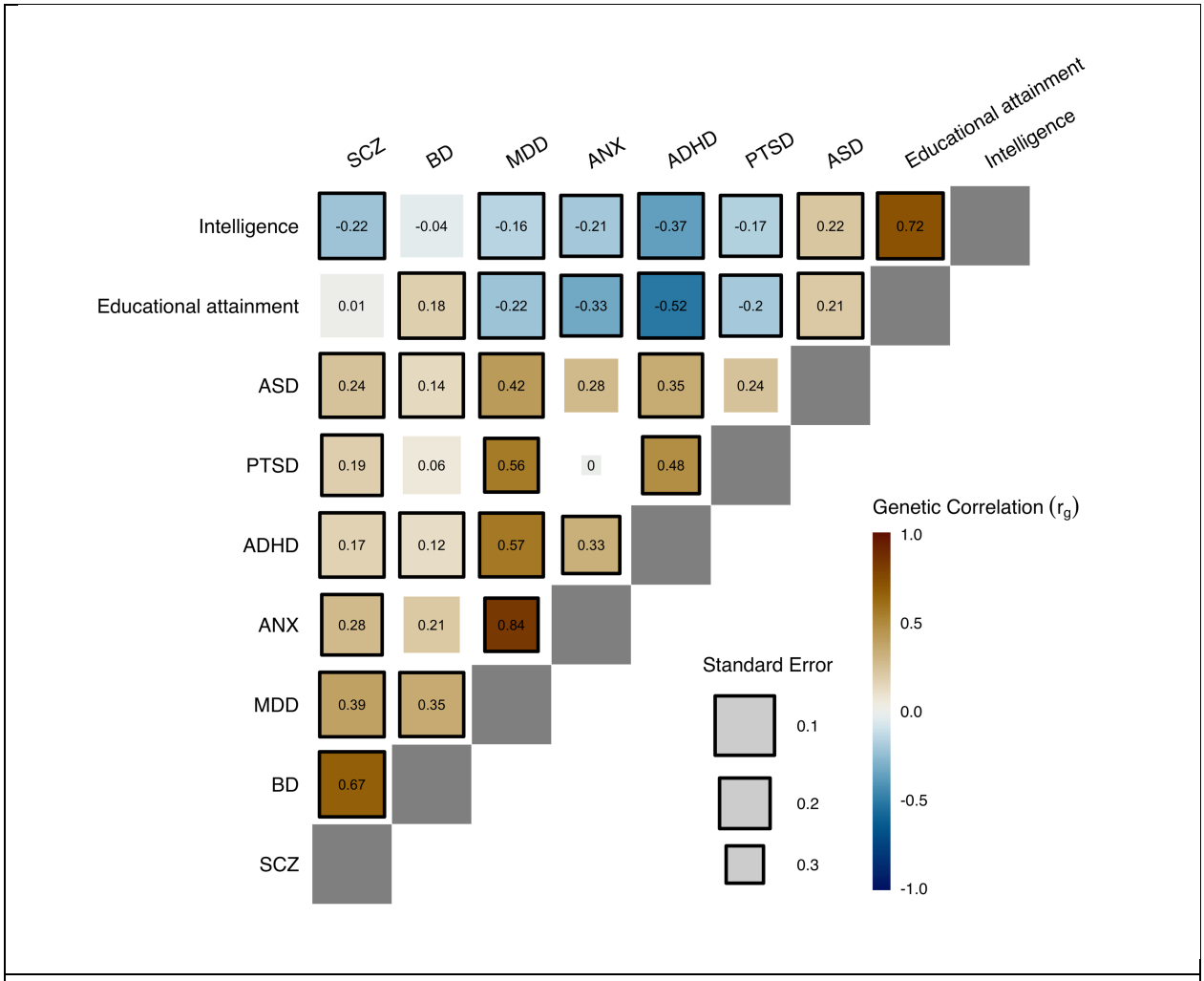
IC	Component represents	no	Gene	Chr	pmin	Individual Significant SNP
IC13	Alcohol abuse	73	IZUMO1	19	2.76726e-06	rs28894750
IC13	Alcohol abuse	74	FUT1	19	2.76726e-06	rs28894750

Suppl. Table 4. Mapped genes from FUMA This table shows the mapped genes and individual lead SNPs as provided by FUMA.

452

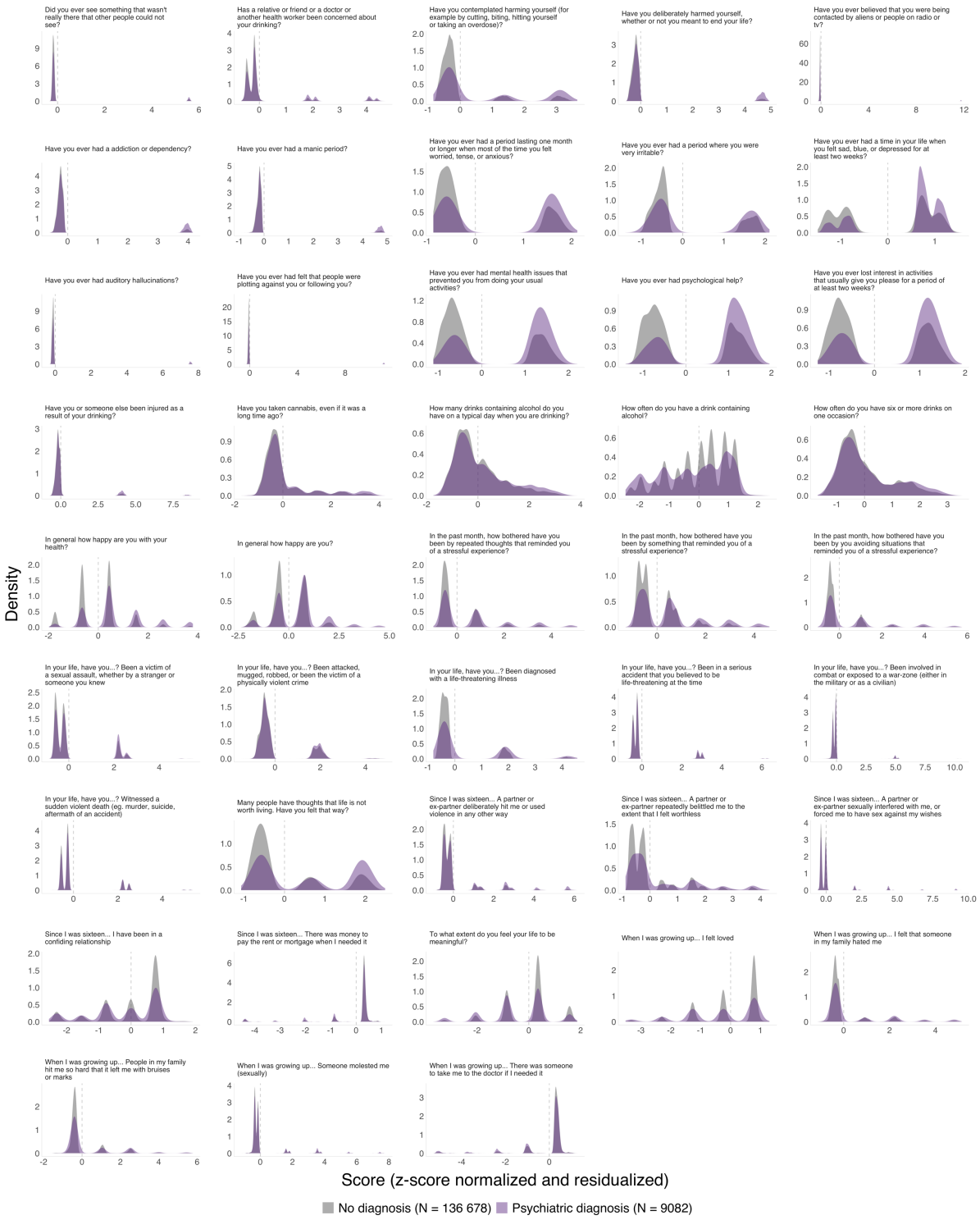
453 **Supplementary Figures**

454



Suppl. Fig. 1. Genetic correlation between the disorders and cognitive traits Numbers inside the boxes denote correlation (r_g). Size of the boxes reflect standard error. Significant correlations ($p < FDR$) are indicated with a black border. In line with previous reports[9,25], the weakest correlation was between PTSD and ANX ($r_g = -0.004$, SE = 0.3408) and the strongest between ANX and MDD ($r_g = 0.8441$, SE = 0.1724).

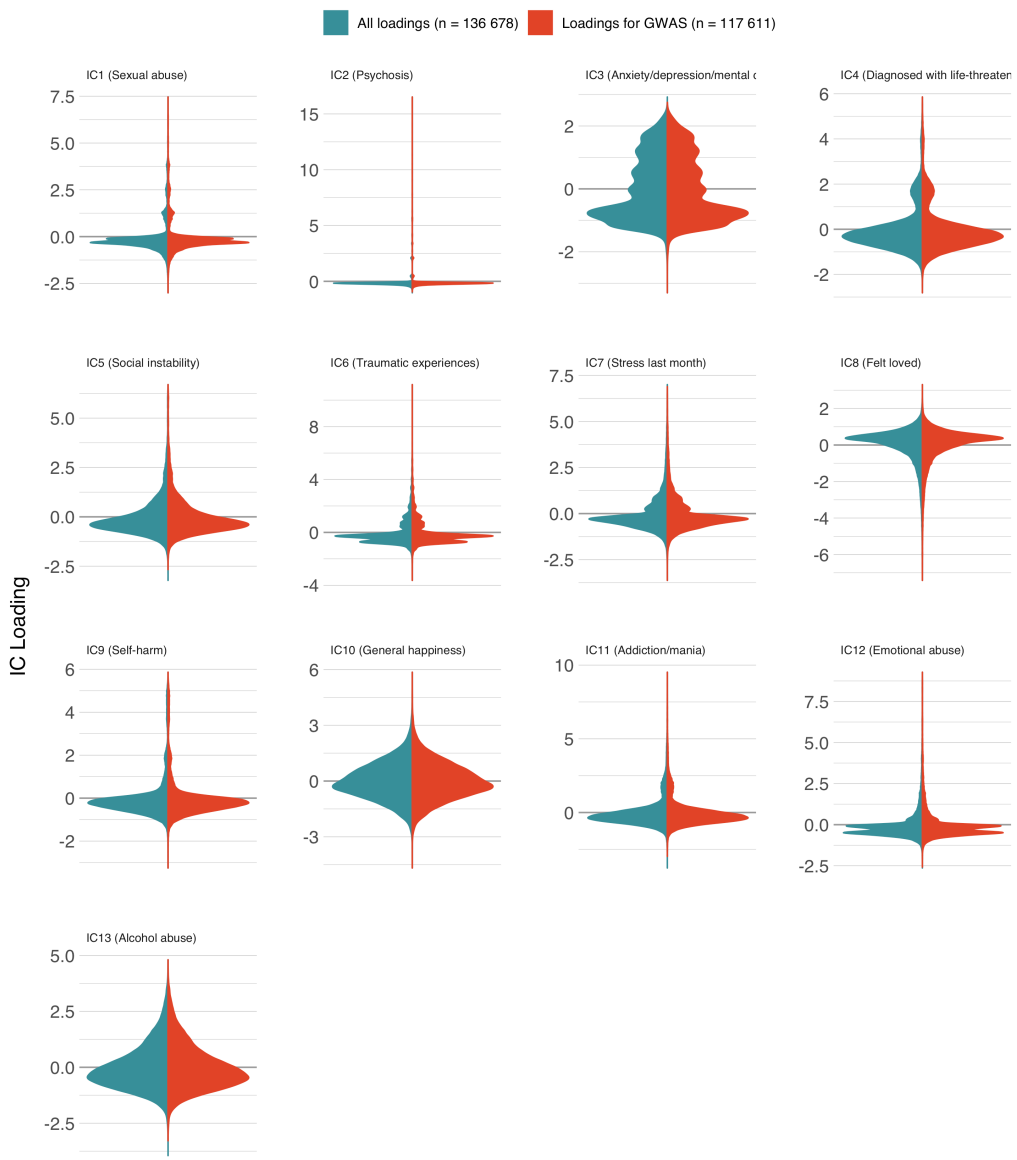
455



Suppl. Fig. 2. Z-normalized question scores Density plot of the scores comparing the sample of individuals without psychiatric or neurological diagnosis to the individuals with a diagnosed psychiatric disorder. With a few notable exceptions, distributions are fairly similar, supporting that mental health data from healthy individuals can be used to study psychiatric disorders. As expected, in the exceptions

where distributions were different, individuals with a psychiatric diagnosis had higher scores overall compared to healthy individuals.

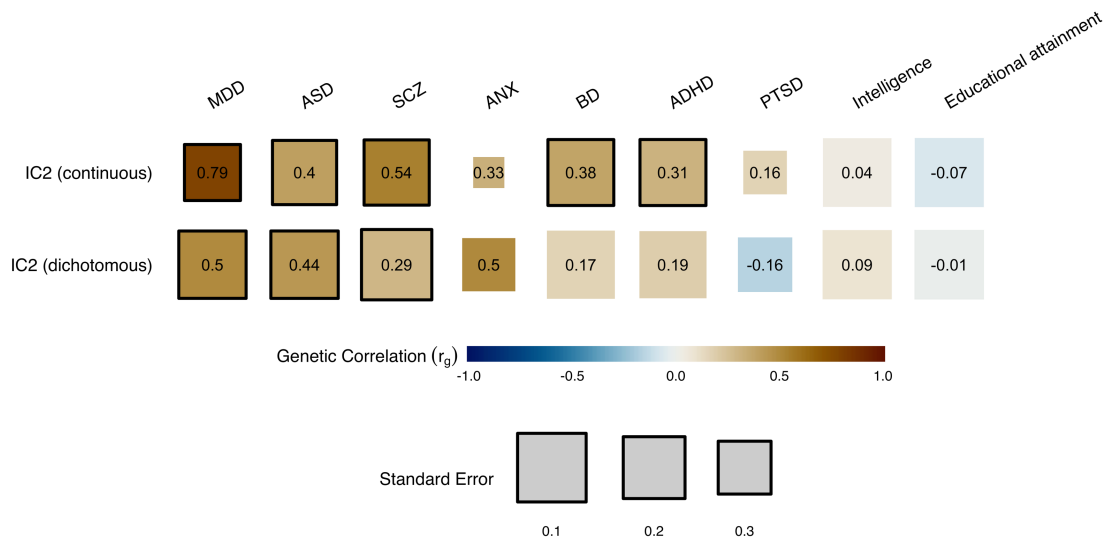
456



Suppl. Fig. 3. ICA weight distributions Density plot of the individual loadings for each independent component. Left half of the plot indicate the loadings for all individuals that were used for ICA. The right side reflects the sample from which genetic data was available and these are the loadings that were used for the GWASs. Distributions on the left show the loadings as provided by the ICA. Distributions were regressed for age (linear and quadratic), sex, and the first 20 genetic principal components before ICA decomposition. Loadings of IC1, IC2, IC5, IC9, IC10, IC11, and IC12 were inverted so that higher

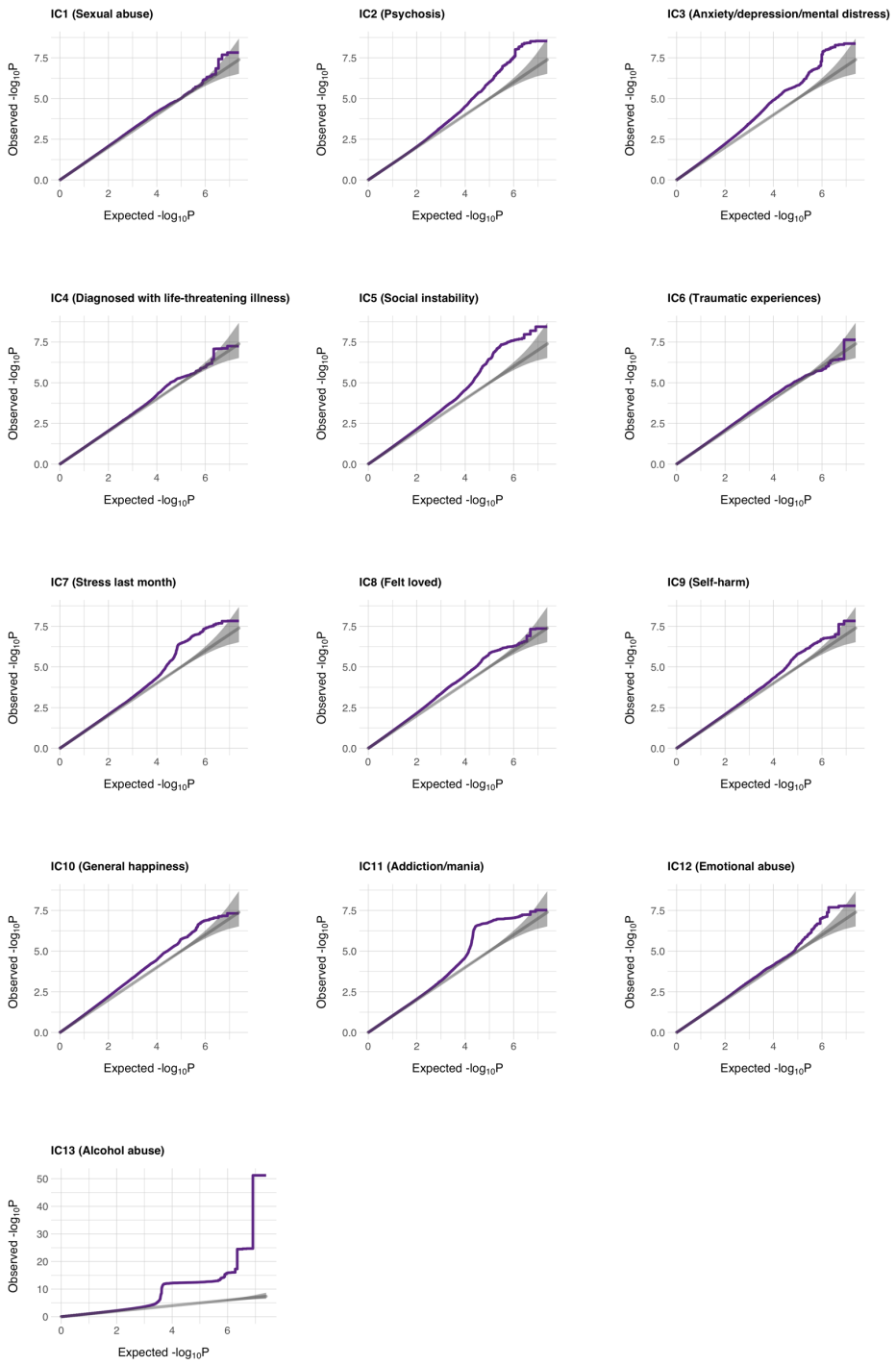
loadings reflect higher symptoms. IC2 has a very narrow distribution, with a few high loadings. IC10 showed a wide distribution due to the universal nature of the questions that comprise this component.

457

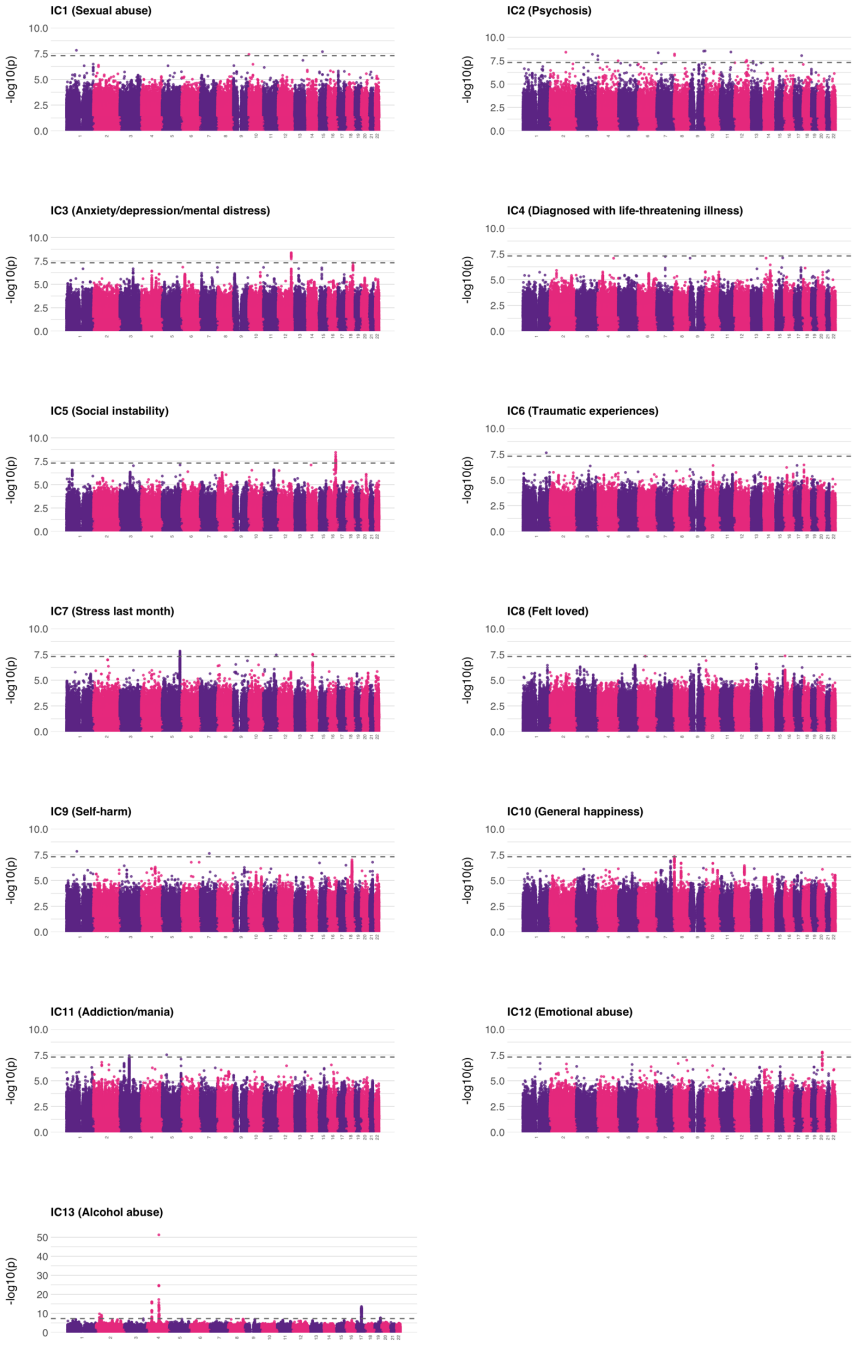


Suppl. Fig. 4. Comparison between continuous and dichotomous IC2 Due to the low variance in IC2 particularly (Suppl. Fig. 3), we investigated if dichotomizing the loading on that component would improve heritability (i.e. loadings larger than 1 were used as cases, lower than 1 were used as controls in a case-control GWAS). This dichotomous approach is conceptually similar to the approach taken by Legge and colleagues[25] who recently reported a genetic correlation between SCZ and dichotomous psychosis symptoms (yes/no). However, dichotomizing did not improve the estimates strongly. Heritability was only slightly higher ($h2_{dichotomous}$ = 0.0133 vs $h2_{continuous}$ = 0.0089) and the reported association with SCZ decreased from $r_{g\ continuous}$ = 0.5427 to $r_{g\ dichotomous}$ = 0.2945. The standard error was more than halved after dichotomization ($SE_{continuous}$ = 0.1527 vs $SE_{dichotomous}$ = 0.0711), suggesting a slight benefit of the dichotomization in scenarios where data distributions limit the continuous approach. In the main analyses, we decided to stay consistent with the other components and keep the phenotype as a continuous measure for IC2, yet we provide results with the dichotomized component for comparison in this figure. Correlations are mostly the same for the continuous and the dichotomous component. The continuous IC2 correlated stronger with MDD compared to the dichotomous IC2, which correlated less strongly with MDD, but stronger with ASD and ANX compared to the continuous IC2. Errors were relatively similar.

458



Suppl. Fig. 5. Q-Q plots Q-Q plots showing the genetic signal from each of the ICs. IC13 showed the strongest signal. IC2 showed a strong signal despite having the lowest h2. None of the GWAS summary statistics showed any noticeable inflation.



Suppl. Fig. 6. Manhattan plots for the ICs Significance threshold set at 5e-8. Only IC4 had no genome-wide significant hits, a few more had no lead SNPs, according to FUMA.

460