# Phenotypically independent mental health symptom 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: 3464 words | Abstract: 350 words | Figures: 4 | References: 41

**Key points**

**Question:** How to statistically independent mental health profiles genetically correlate with each other, and with psychiatric disorders and cognitive traits?

**Findings:** Symptom profiles capturing different facets of mental health that were phenotypically uncorrelated were nonetheless genetically correlated. The symptom profiles also genetically correlated with psychiatric disorders and cognitive traits and although strongest correlations were typically observed between a given symptom profile and a disorder for which the symptom is common, specificity was overall limited.

**Meaning:** The genetic correlations of phenotypically independent symptom profiles may suggest that the known pleiotropy among common psychiatric disorders cannot be exclusively attributed to the overlapping symptomatology between the disorders.

**Abstract**

**Importance:** 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, and the predominant use of classical case-control designs have not allowed the study of variations in mental health independent of diagnosis.

**Objective:** To derive distinct profiles of mental symptoms in healthy individuals and to study how these genetically relate to each other and to common psychiatric disorders.

**Design:** This is a cross-sectional study using self-report mental health questionnaires and molecular genetic data.

**Setting:** We used population-based data from the UK Biobank.

**Participants:** Data from individuals with a diagnosed neurological or psychiatric disorder were excluded, allowing us to study variations in mental health in 139,006 healthy individuals, and genotypes in 117,088 healthy individuals with Caucasian ancestry.

**Main Outcomes and Measures:** We decomposed self-report mental health questionnaires into twelve distinct symptom profiles using independent component analysis, and performed a GWAS for each of them. Using GWAS summary statistics, we assessed genetic correlations between the symptom profiles, and between symptom profiles and common psychiatric disorders and cognitive traits.

**Results:** We found that symptom profiles were genetically correlated with a wide range of psychiatric disorders and cognitive traits (67 out of 108 correlations significant at $p < FDR$), with strongest effects typically observed between a given symptom profile and a disorder for which the symptom is common (e.g. depression symptoms and major depressive disorder, trauma experience and post-traumatic stress disorder). Strikingly, although the symptom profiles were phenotypically uncorrelated, many of them were genetically correlated with each other (31 out of 66 comparisons significant; $p < FDR$).

**Conclusions and Relevance:** This study provides evidence that statistically independent mental health profiles in healthy individuals partly share genetic underpinnings and show genetic overlaps with psychiatric disorders. These findings suggest that shared genetics across psychiatric disorders cannot be exclusively attributed to the overlapping symptomatology between and the heterogeneity within psychiatric disorders, and supports that moving from a classical case-control setting to a continuous mental health spectrum may complement the study of psychiatric genetics.

## 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 symptom profiles, and between symptom 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 symptom 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

*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

the National Health Service National Research Ethics Service (ref. 11/NW/0382). Participants with a diagnosed psychiatric or neurological disorder (F or G ICD10 diagnosis) were excluded from the analysis except for those with a nerve, nerve root and plexus disorders (categories G50 to G59). In addition, we excluded participants with more than 10% missing answers in the mental health questionnaires. This resulted in mental health data from 139,006 individuals, which was used in an independent component analysis. For the genetic analysis, we selected data from all Caucasian individuals with available genotypes, yielding a set of 117,088 participants aged 47-80 years (mean: 64, SD: 7.66, age at mental health assessment) and comprised 56.2% females.

*Processing of mental health data*

Fig. 1A depicts the analysis workflow. Of the 60 primary questions that were available as part of the mental health online assessment, we removed questions that asked specifically about symptoms occurring in the past two weeks to remove potential short-term temporal effects. We also removed questions on substance abuse due to the complex interactions between substance abuse and mental health. Furthermore, we excluded questions where more than 10% of the responses were missing (1 question excluded). In the resulting set of 37 questions (Suppl. Table 1), we imputed missing data using k-nearest neighbor imputation with k = 3 with the bnstruct package[30] in R[31] and z-standardized the data.

The resulting data covering 37 questions from 139,006 individuals was decomposed using independent component analysis (ICA). Using icasso[32] in MATLAB and by visually inspecting the loadings of the questions on the components, we estimated that a model order of 12 independent components yields the best clustering solution where the resulting components are stable and highly interpretable. The PCA identified 12 components with an eigenvalue larger than 1, and stability (Iq) was effectively 1. A model order lower than 12 would group together the psychosis questions with other symptom domains, which we preferred to keep separate. A model order larger than 12 was not reasonable as it would yield components that largely reflect single items. Next, we decomposed the items into 12 independent components using the fastICA algorithm as implemented in R[33]. Fig. 1B depicts how each of the 37 items loaded on the components, indicating independent components (ICs) that captured questions on sexual abuse (IC1), psychosis (IC2), anxiety, depression and mental distress (IC3), a diagnosis with a life-threatening illness (IC4), social instability (IC5), traumatic experiences (IC6), stress in the past month (IC7), experiences of feeling loved (IC8), thoughts around self-harm behavior (IC9), general happiness (IC10), addiction behavior and manic experiences (IC11), and experiences of emotional abuse (IC12). The individual level scores for each of the 12 components were subsequently residualized for age (both linear and quadratic term), sex, and the first 20 genetic principal components.

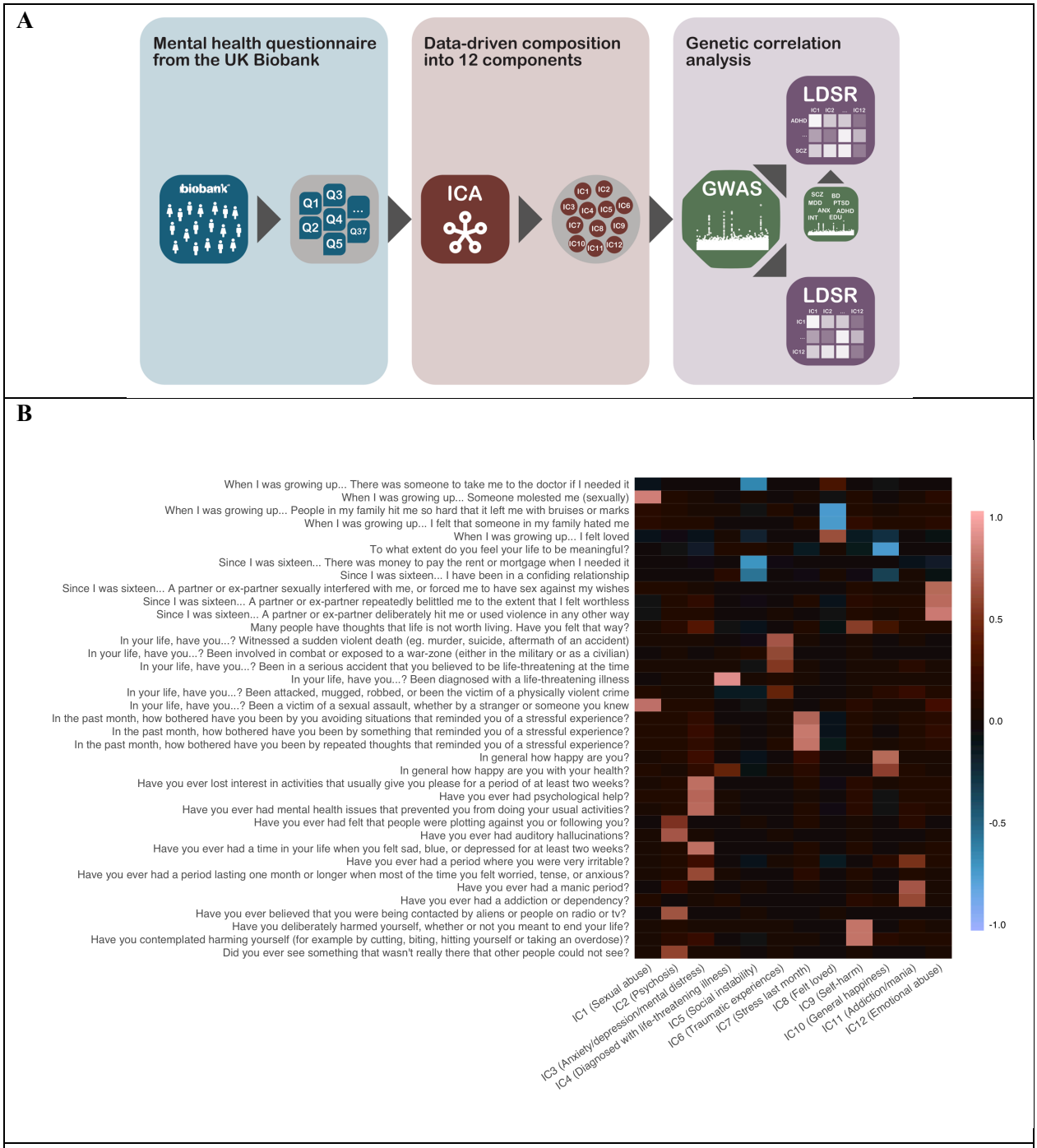



**Fig. 1. Workflow and variable weight matrix of the resulting decomposition A** Outline of the analysis workflow. **B** Weight matrix reflecting how each mental health question loaded on each IC. Brighter colors indicate higher loading, darker colors indicate lower loading. All 37 questions were captured by at least one of the 12 independent components. To facilitate interpretation, loadings of IC6, IC7, IC9, and IC11 were inverted so that all components showed the same direction of effect (higher component score indicating increased scoring on the items).

The distribution of IC2 indicated very few non-zero scores (Suppl. Fig. 2). This component loaded mostly on psychosis questions (Fig. 1B), indicating that only few of the included healthy individuals had symptoms in this domain. We therefore conducted an additional supplemental analysis in which we dichotomized IC2 such that loadings lower than 1 were labeled as "no/few symptoms", and loadings equal to or higher than 1 were labeled as "with symptoms".

*Processing of genetic data*

From the UK Biobank v3 imputed genetic data, we removed SNPs with an imputation quality score below 0.5, with a minor allele frequency below 0.001, missing in more than 5% of individuals, and that failed the Hardy-Weinberg equilibrium test at $p < 1e-9$. We removed also individuals with more than 10% missing data. We performed a genome-wide association analysis (GWAS) on each of the 12 independent components in PLINK 1.9[34,35]. Using a publicly available conversion toolbox for GWAS summary statistics (github.com/precimed/python_convert), we removed the MHC region and calculated a z-score for every SNP (8,165,726 SNPs after QC). We utilized linkage-disequilibrium score regression[10,36] to estimate genetic correlations between each of the independent components, and between the components and publicly available GWAS summary statistics for SCZ[1], BD[2], MD[37], ADHD[4], ASD[5], PTSD[6], ANX[7], as well as intelligence[38], and educational attainment[39] (Suppl. Table 2). For all aforementioned GWASs, we used those versions that did not have UK Biobank participants included. From the MD GWAS, we removed participants from the 23andMe dataset as well, leaving only cases with a diagnosed major depressive disorder (MDD). Prior to estimating genetic correlations, we set a threshold that only ICs with a heritability 1.96 times larger than its standard error should be included in the analysis and only those where visual quality control of corresponding Q-Q plots indicated genetic signal. These quality control steps were implemented to ensure that we did not make inferences on data that did not provide sufficient variance explained by genetics. Partitioned heritability[40] was estimated using the LDSC toolbox[36] and Q-Q plots were generated using custom scripts in R. Finally, we processed the GWAS summary statistics of each independent component through the Functional Mapping and Annotation toolbox (FUMA) to map lead SNPs onto genes[41]. FUMA parameters were kept as default, and we used the FUMA default European ancestry reference panel.

*Code and data availability*

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

## Results

Fig. 2 depicts SNP-based heritability (h2) for the 12 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. IC3, capturing questions on anxiety, depression and mental distress, had the highest heritability (h2 = 0.0738, SE = 0.0058), closely followed by IC10, capturing questions on general happiness (h2 = 0.0632, SE = 0.0052). The lowest heritability among the components was for IC2, reflecting psychosis questions (h2 = 0.0115, SE = 0.0047), likely owing to the low number of individuals with psychosis symptoms (Suppl. Fig. 2). We therefore performed a supplemental analysis to investigate if dichotomization of this IC would benefit the analysis (Suppl. Fig. 3). In brief, as dichotomization only marginally 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. 3. 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. 4) warranting inclusion of all components into subsequent genetic correlation analyses.

Except for IC4, all ICs showed genome wide significant SNPs at a threshold of 5e-8 (Suppl. Fig. 5). Using FUMA, we discovered 11 independent loci for IC2, 2 loci for IC3, IC5 and IC12, and 1 locus for IC6, IC7, IC8, IC9, and IC10, respectively. IC1 and IC4 had no significant genetic risk loci. Suppl. Table 4 provides a list of mapped genes for all ICs, illustrating that IC2 had the most mapped genes among all ICs (26 mapped genes).

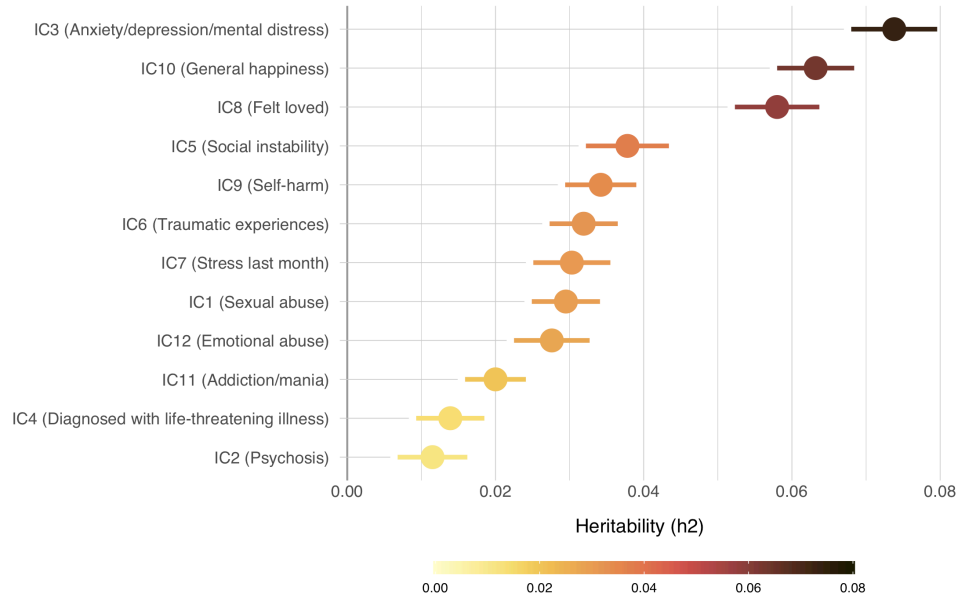



**Fig. 2. Heritability estimates of the independent components** SNP-based heritability for each IC sorted by decreasing heritability (h2). Heritability calculated using the LDSC toolbox[36]. Error bars reflect standard errors.

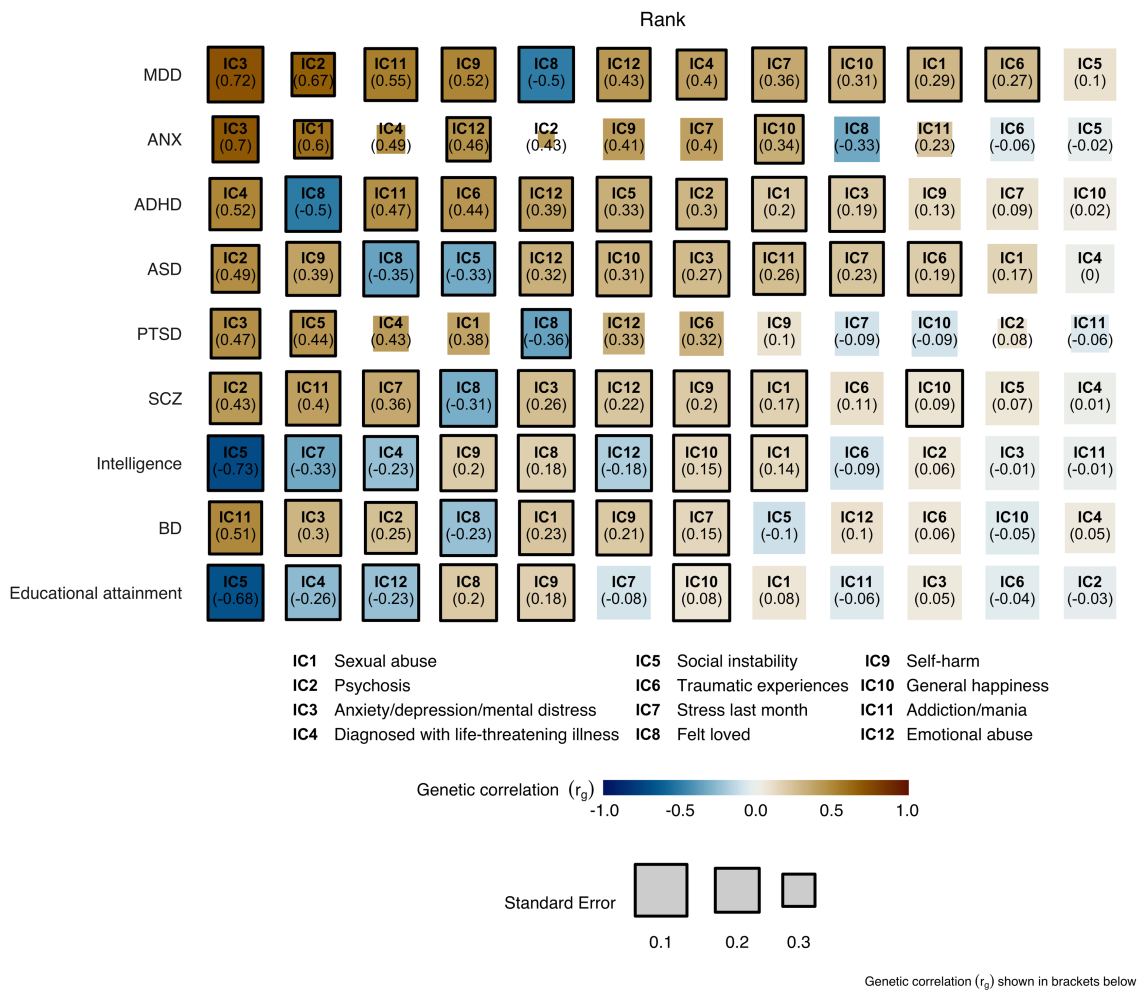



**Fig. 3. Genetic correlation between the independent components and disorders and cognitive traits** For each disorder, the associations with ICs are sorted by decreasing absolute genetic correlation such that the most leftward box reflects the strongest association between a given disorder and the 12 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 12 ICs and a set of psychiatric disorders as well as cognitive traits. Out of 108 comparisons, 67 were significant after FDR correction, which amounts to 62%. 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, both MDD and anxiety most strongly correlated with IC3, which reflects anxiety, depression, and mental distress (genetic correlation $r_g = 0.72$ and $= .70$

respectively, both $p_{FDR} < .00056$). SCZ was most highly correlated with IC2, which represents psychosis questions ($r_g = 0.43$, $p_{FDR} = .00055$). The highest genetic correlation of BD was with IC11, which represents addiction and mania ($r_g = 0.51$, $p_{FDR} = 2.7e\text{-}7$). For PTSD, the component reflecting traumatic experience (IC6) only ranked seventh among the sorted associations, yet the two ICs showing strongest association with PTSD reflected anxiety, depression, and mental distress (IC3; $r_g = 0.47$, $p_{FDR} = .0019$) and social instability (IC5; $r_g = 0.44$, $p_{FDR} = .037$), both of which are closely related to PTSD. ASD correlated strongest with IC2 (reflecting psychosis; $r_g = 0.49$, $p_{FDR} = .0066$) and ADHD correlated strongest with IC4 (diagnosed with life-threatening illness; $r_g = 0.52$, $p_{FDR} = 4.1e\text{-}6$). Educational attainment and intelligence were both strongest negatively correlated with the IC reflecting social instability (IC5, $r_g = -0.68$ and $r_g = -0.73$, respectively; both $p_{FDR} < 8.4e\text{-}40$) and showed generally negative genetic correlations with the other ICs. 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, almost half of the IC pairs were nonetheless significantly genetically correlated with each other (47%, $p < FDR$). IC3 (anxiety, depression, mental illness) and IC6 (traumatic experiences) were each genetically correlated with eight other ICs. IC1 (sexual abuse), IC8 (felt loved) and IC9 (self-harm) were each genetically correlated with seven other ICs. IC11 (addiction/mania), IC12 (emotional abuse) and IC2 (psychosis) were each genetically correlated with six other ICs. IC10 (general happiness) and IC4 (diagnosed with a life-threatening illness) were each correlated with 3 other ICs and IC7 (stress last month) was correlated with one other IC. IC5 (social instability) was the only IC without significant genetic correlation to other ICs. The analysis therefore revealed a large amount of genetic correlations despite statistical (phenotypic) independence of the symptom profiles.

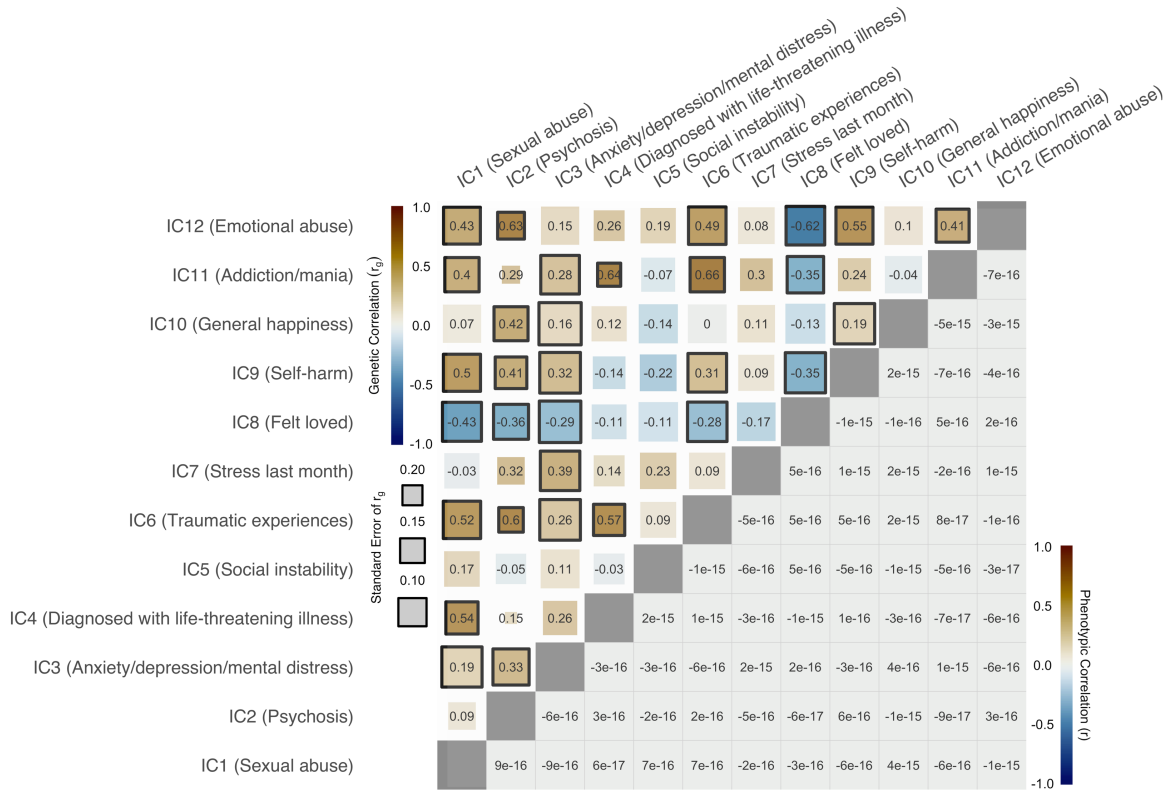



**Fig. 4. Phenotypic and genetic correlation between the ICs** The lower half of the IC by IC matrix depicts phenotypic correlations, reflecting the Pearson correlation of subject level component scores between independent components. As expected by ICA design, correlations were close to zero. The upper half of the matrix depicts the genetic correlations ($r_g$), indicating significant genetic correlations in 31 of 66 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 symptom data from more than 130,000 individuals into phenotypically distinct symptom 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 symptoms suggested limited specificity. Indeed, we found a large proportion of significant genetic correlations between the phenotypically uncorrelated symptom 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 symptom 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 associations with IC3, which reflects variance in anxiety, depression, and mental distress, was with MDD and 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 and SCZ showed significant correlations between all but one component. ADHD and ASD correlated with all but two 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. Most of the genetic correlations with intelligence and educational attainment pointed in the opposite direction, reflecting higher cognitive ability with fewer psychiatric symptoms. Importantly, when looking at the correlations between symptom 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 symptom 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.

295 *Limitations*

296 Notable strengths of the present study include the use of data-driven symptom decomposition in a large
297 sample of healthy individuals and its application to study pleiotropy in psychiatric genetics. Its main
298 limitations include the relatively low heritability of the resulting independent components, and the limited
299 number of individuals with psychosis symptoms yielding suboptimal distribution in IC2 (Suppl. Fig. 2).
300 First, it is important to note that all ICs passed quality control. Heritability of all ICs exceeded our pre-
301 defined heritability threshold of 1.96 times its standard error, and all Q-Q plots indicated genetic signal
302 (Suppl. Fig. 4). Furthermore, low heritability can still produce good genetic signal as a result from a low
303 number of genetic variants involved but where each has large effects[13]. For example, while IC2 had the
304 lowest heritability among the ICs, it showed the strongest genetic signal and yielded the largest number of
305 discoveries in FUMA. Second, although sample size and symptom distributions factored into the results,
306 these are mostly reflected in the standard error of genetic associations, not in a lack of effect. For example,
307 ANX[7] (n = 21,761) and PTSD[6] (n = 9,537) GWASs have relatively little power as reflected in the larger
308 standard errors in genetic correlations with these disorders, but nonetheless the strongest associations with
309 these disorders were with components that match symptoms of the disorders (IC3 with ANX, IC11 with
310 BD). Likewise, the suboptimal symptom distributions in IC2 and corresponding low heritability is reflected
311 in large standard errors of the resulting genetic correlations but nonetheless IC2, reflecting psychosis, was
312 most strongly associated with SCZ. Supplemental analysis with dichotomized IC2 also confirmed that the
313 distribution alone is unlikely to explain the observed associations (Suppl. Fig. 3).

314

315 *Conclusion*

316 In the present study, we revealed genetic overlap between statistically independent symptom profiles and
317 provide evidence that variations in mental health in healthy individuals relate genetically to psychiatric
318 disorders and cognitive traits. These findings support that pleiotropy between psychiatric disorders cannot
319 simply be explained by overlapping symptoms but may rather point to similar biological underpinnings of
320 distinct symptoms. Our results underscore the potential of data-driven approaches to the study of psychiatric
321 symptoms, and suggests that supplementing the classic case-control design with a dimensional approach
322 may improve our understanding of the genetic underpinnings of complex disorders of the mind.

323

324 **Acknowledgements**

325 The authors were funded by the Research Council of Norway (#276082 LifespanHealth, #223273
326 NORMENT, #283798 ERA-NET Neuron SYNSCHIZ, #249795), the South-East Norway Regional Health
327 Authority (2019101), and the European Research Council under the European Union's Horizon2020
328 Research and Innovation program (ERC Starting Grant #802998), as well as the Horizon2020 Research and
329 Innovation Action Grant CoMorMent (#847776). This research has been conducted using the UK Biobank

Resource (access code 27412, https://www.ukbiobank.ac.uk/). This work was performed on the TSD (Tjenester for Sensitive Data) facilities, owned by the University of Oslo, operated and developed by the TSD service group at the University of Oslo, IT-Department (USIT). Computations were also performed on resources provided by UNINETT Sigma2 - the National Infrastructure for High Performance Computing and Data Storage in Norway.

**Conflicts of interest**

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 to HealthLytix and received speakers honorarium from Lundbeck.

**Author contributions**

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

**References**

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

365 associated with schizophrenia and bipolar disorder using pleiotropy-informed conditional false
366 discovery rate. *PLoS Genet*. 2013;9(4):e1003455. doi:10.1371/journal.pgen.1003455
367 9. Anttila V, Bulik-Sullivan B, Finucane HK, et al. Analysis of shared heritability in common
368 disorders of the brain. *Science*. 2018;360(6395):eaap8757. doi:10.1126/science.aap8757
369 10. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human
370 diseases and traits. *Nat Genet*. 2015;47(11):1236-1241. doi:10.1038/ng.3406
371 11. Smeland OB, Bahrami S, Frei O, et al. Genome-wide analysis reveals extensive genetic overlap
372 between schizophrenia, bipolar disorder, and intelligence. *Mol Psychiatry*. January
373 2019:10.1038/s41380-018-0332-x. doi:10.1038/s41380-018-0332-x
374 12. Bansal V, Mitjans M, Burik CAP, et al. Genome-wide association study results for educational
375 attainment aid in identifying genetic heterogeneity of schizophrenia. *Nat Commun*. 2018;9(1):3078.
376 doi:10.1038/s41467-018-05510-z
377 13. Frei O, Holland D, Smeland OB, et al. Bivariate causal mixture model quantifies polygenic overlap
378 between complex traits beyond genetic correlation. *Nat Commun*. 2019;10(1):2417-2417.
379 doi:10.1038/s41467-019-10310-0
380 14. Bahrami S, Steen NE, Shadrin A, et al. Shared Genetic Loci Between Body Mass Index and Major
381 Psychiatric Disorders: A Genome-wide Association Study. *JAMA Psychiatry*. January
382 2020:10.1001/jamapsychiatry.2019.4188. doi:10.1001/jamapsychiatry.2019.4188
383 15. Amare AT, Schubert KO, Klingler-Hoffmann M, Cohen-Woods S, Baune BT. The genetic overlap
384 between mood disorders and cardiometabolic diseases: a systematic review of genome wide and
385 candidate gene studies. *Transl Psychiatry*. 2017;7(1):e1007-e1007. doi:10.1038/tp.2016.261
386 16. Andreasen NC. A unitary model of schizophrenia: Bleuler's "fragmented phrene" as
387 schizencephaly. *Arch Gen Psychiatry*. 1999;56(9):781-787. doi:10.1001/archpsyc.56.9.781
388 17. Craddock N, Owen MJ. The Kraepelinian dichotomy - going, going... but still not gone. *Br J*
389 *Psychiatry J Ment Sci*. 2010;196(2):92-95. doi:10.1192/bjp.bp.109.073429
390 18. Alnaes D, Kaufmann T, van der Meer D, et al. Brain Heterogeneity in Schizophrenia and Its
391 Association With Polygenic Risk. *JAMA Psychiatry*. 2019;76(7):739-748.
392 doi:10.1001/jamapsychiatry.2019.0257
393 19. Hibar DP, Westlye LT, Doan NT, et al. Cortical abnormalities in bipolar disorder: an MRI analysis
394 of 6503 individuals from the ENIGMA Bipolar Disorder Working Group. *Mol Psychiatry*.
395 2018;23(4):932-942. doi:10.1038/mp.2017.73
396 20. McGrath JJ, Saha S, Al-Hamzawi A, et al. Psychotic Experiences in the General Population: A
397 Cross-National Analysis Based on 31,261 Respondents From 18 Countries. *JAMA Psychiatry*.
398 2015;72(7):697-705. doi:10.1001/jamapsychiatry.2015.0575
399 21. Pirastu N, Cordioli M, Nandakumar P, et al. Genetic analyses identify widespread sex-differential

400 participation bias. *bioRxiv*. January 2020:2020.03.22.001453. doi:10.1101/2020.03.22.001453
401 22. Holland D, Frei O, Desikan R, et al. Beyond SNP Heritability: Polygenicity and Discoverability of
402 Phenotypes Estimated with a Univariate Gaussian Mixture Model. *bioRxiv*. January 2019:133132.
403 doi:10.1101/133132
404 23. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the
405 causes of a wide range of complex diseases of middle and old age. *PLoS Med*.
406 2015;12(3):e1001779. doi:10.1371/journal.pmed.1001779
407 24. Davis KAS, Coleman JRI, Adams M, et al. Mental health in UK Biobank – development,
408 implementation and results from an online questionnaire completed by 157 366 participants: a
409 reanalysis. *BJPsych Open*. 2020;6(2):e18. doi:10.1192/bjo.2019.100
410 25. Legge SE, Jones HJ, Kendall KM, et al. Association of Genetic Liability to Psychotic Experiences
411 With Neuropsychotic Disorders and Traits. *JAMA Psychiatry*. 2019;76(12):1256-1265.
412 doi:10.1001/jamapsychiatry.2019.2508
413 26. Caspi A, Houts RM, Belsky DW, et al. The p Factor: One General Psychopathology Factor in the
414 Structure of Psychiatric Disorders? *Clin Psychol Sci J Assoc Psychol Sci*. 2014;2(2):119-137.
415 doi:10.1177/2167702613497473
416 27. Jutten C, Hérault J. Détection de grandeurs primitives dans un message composite par une
417 architecture de calcul neuromimétique en apprentissage non supervisé. *10° Colloq Sur Trait Signal*
418 *Images FRA 1985*. 1985:1017-1022.
419 28. Comon P. Independent component analysis, A new concept? *High Order Stat*. 1994;36(3):287-314.
420 doi:10.1016/0165-1684(94)90029-9
421 29. Alnæs D, Kaufmann T, Doan NT, et al. Association of Heritable Cognitive Ability and
422 Psychopathology With White Matter Properties in Children and Adolescents. *JAMA Psychiatry*.
423 2018;75(3):287-295. doi:10.1001/jamapsychiatry.2017.4277
424 30. Alberto Franzin, Francesco Sambo, Barbara di Camillo. bnstruct: an R package for Bayesian
425 Network structure learning in the presence of missing data. *Bioinformatics*. 2017;33(8):1250-1252.
426 doi:10.1093/bioinformatics/btw807
427 31. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R
428 Foundation for Statistical Computing; 2017. www.R-project.org.
429 32. Himberg J, Hyvarinen A. Icasso: software for investigating the reliability of ICA estimates by
430 clustering and visualization. *2003 IEEE XIII Workshop Neural Netw Signal Process IEEE Cat*
431 *No03TH8718*. 2003:259-268. doi:10.1109/NNSP.2003.1318025
432 33. Marchini J, Heaton C, Ripley B. *FastICA: FastICA Algorithms to Perform ICA and Projection*
433 *Pursuit*.; 2019. https://CRAN.R-project.org/package=fastICA.
434 34. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK:

435 rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7. doi:10.1186/s13742-
436 015-0047-8
437 35. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and
438 population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-575. doi:10.1086/519795
439 36. Bulik-Sullivan BK, Loh PR, Finucane HK, et al. LD Score regression distinguishes confounding
440 from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47(3):291-295.
441 doi:10.1038/ng.3211
442 37. Wray NR, Ripke S, Mattheisen M, et al. Genome-wide association analyses identify 44 risk variants
443 and refine the genetic architecture of major depression. *Nat Genet*. 2018;50(5):668-681.
444 doi:10.1038/s41588-018-0090-3
445 38. Savage JE, Jansen PR, Stringer S, et al. Genome-wide association meta-analysis in 269,867
446 individuals identifies new genetic and functional links to intelligence. *Nat Genet*. 2018;50(7):912-
447 919. doi:10.1038/s41588-018-0152-6
448 39. Lee JJ, Wedow R, Okbay A, et al. Gene discovery and polygenic prediction from a genome-wide
449 association study of educational attainment in 1.1 million individuals. *Nat Genet*. 2018;50(8):1112-
450 1121. doi:10.1038/s41588-018-0147-3
451 40. Finucane HK, Bulik-Sullivan B, Gusev A, et al. Partitioning heritability by functional annotation
452 using genome-wide association summary statistics. *Nat Genet*. 2015;47(11):1228-1235.
453 doi:10.1038/ng.3404
454 41. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of
455 genetic associations with FUMA. *Nat Commun*. 2017;8(1):1826. doi:10.1038/s41467-017-01261-5
456
457 **Supplementary Tables**
458

| n° | 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 behaviors (such as gambling)? | 502 |
| 9 | 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 |
| 10 | 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 |
| 11 | Did you ever see something that wasn't really there that other people could not see? | 502 |
| 12 | 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 |
| 13 | When I was growing up... I felt that someone in my family hated me | 532 |
| 14 | When I was growing up... People in my family hit me so hard that it left me with bruises or marks | 532 |
| 15 | When I was growing up... I felt loved | 532 |
| 16 | When I was growing up... Someone molested me (sexually) | 532 |
| 17 | When I was growing up... There was someone to take me to the doctor if I needed it | 532 |
| 18 | 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 |
| 19 | 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 |
| 20 | 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 |
| 21 | Since I was sixteen... A partner or ex-partner repeatedly belittled me to the extent that I felt worthless | 532 |
| 22 | Since I was sixteen... I have been in a confiding relationship | 532 |

| | | |
|---|---|---|
| 23 | Since I was sixteen... A partner or ex-partner deliberately hit me or used violence in any other way | 532 |
| 24 | Since I was sixteen... A partner or ex-partner sexually interfered with me, or forced me to have sex against my wishes | 532 |
| 25 | Since I was sixteen... There was money to pay the rent or mortgage when I needed it | 532 |
| 26 | In your life, have you...? Been in a serious accident that you believed to be life-threatening at the time | 533 |
| 27 | In your life, have you...? Been involved in combat or exposed to a war-zone (either in the military or as a civilian) | 533 |
| 28 | In your life, have you...? Been diagnosed with a life-threatening illness | 533 |
| 29 | In your life, have you...? Been attacked, mugged, robbed, or been the victim of a physically violent crime | 533 |
| 30 | In your life, have you...? Witnessed a sudden violent death (e.g. murder, suicide, aftermath of an accident) | 533 |
| 31 | In your life, have you...? Been a victim of a sexual assault, whether by a stranger or someone you knew | 533 |
| 32 | Many people have thoughts that life is not worth living. Have you felt that way? | 535 |
| 33 | Have you deliberately harmed yourself, whether or not you meant to end your life? | 503 |
| 34 | Have you contemplated harming yourself (for example by cutting, biting, hitting yourself or taking an overdose)? | 535 |
| 35 | In general how happy are you? | 537 |
| 36 | In general how happy are you with your health? | 537 |
| 37 | To what extent do you feel your life to be meaningful? | 538 |

**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 139.006 individuals remained. There are in total 37 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

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 |

459

| 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.

460

| IC | Component represents | h2 (SE) | Lambda GC | Intercept (SE) |
|---|---|---|---|---|
| IC1 | Sexual abuse | 0.0295 (0.0046) | 1.0618 | 1.0007 (0.0068) |
| IC2 | Psychosis | 0.0115 (0.0047) | 1.0315 | 1.0005 (0.0069) |
| IC3 | Anxiety/depression/mental distress | 0.0738 (0.0058) | 1.1619 | 1.0057 (0.0083) |
| IC4 | Diagnosed with life-threatening illness | 0.0139 (0.0046) | 1.0466 | 1.0093 (0.0075) |
| IC5 | Social instability | 0.0378 (0.0056) | 1.0833 | 1.0098 (0.0081) |
| IC6 | Traumatic experiences | 0.0319 (0.0046) | 1.0649 | 1.0028 (0.0075) |
| IC7 | Stress last month | 0.0303 (0.0052) | 1.0557 | 0.9928 (0.0077) |
| IC8 | Felt loved | 0.058 (0.0057) | 1.1207 | 1.0035 (0.0078) |

| | | | | |
|---|---|---|---|---|
| IC9 | Self-harm | 0.0342 (0.0048) | 1.0802 | 1.0034 (0.007) |
| IC10 | General happiness | 0.0632 (0.0052) | 1.1587 | 1.0339 (0.0075) |
| IC11 | Addiction/mania | 0.02 (0.0041) | 1.0527 | 1.0077 (0.0068) |
| IC12 | Emotional abuse | 0.0276 (0.0051) | 1.0557 | 0.9932 (0.0075) |

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

461

| IC | Component represents | n° | Gene | Chr | $p_{min}$ | Individual Significant SNP |
|---|---|---|---|---|---|---|
| IC2 | Psychosis | 1 | PTH2R | 2 | 4.625e-08 | rs74977886 |
| IC2 | Psychosis | 2 | ETV1 | 7 | NA | rs571189241 |
| IC2 | Psychosis | 3 | DGKB | 7 | NA | rs571189241 |
| IC2 | Psychosis | 4 | C8orf47 | 8 | 5.636e-09 | rs2514338 |
| IC2 | Psychosis | 5 | ARHGAP42 | 11 | NA | rs182722613 |
| IC2 | Psychosis | 6 | TMEM133 | 11 | NA | rs182722613 |
| IC2 | Psychosis | 7 | PGR | 11 | NA | rs182722613 |
| IC2 | Psychosis | 8 | TRPC6 | 11 | 2.058e-10 | rs182722613 |
| IC2 | Psychosis | 9 | C11orf70 | 11 | NA | rs182722613 |
| IC2 | Psychosis | 10 | YAP1 | 11 | NA | rs182722613 |
| IC2 | Psychosis | 11 | UHRF1BP1L | 12 | NA | rs540594149 |
| IC2 | Psychosis | 12 | NR1H4 | 12 | 8.134e-05 | rs540594149 |
| IC2 | Psychosis | 13 | GAS2L3 | 12 | 2.356e-08 | rs540594149 |
| IC2 | Psychosis | 14 | ANO4 | 12 | 4.129e-07 | rs540594149 |
| IC2 | Psychosis | 15 | CUX2 | 12 | NA | rs139516917 |
| IC2 | Psychosis | 16 | FAM109A | 12 | 2.741e-07 | rs139516917 |
| IC2 | Psychosis | 17 | SH2B3 | 12 | 8.970e-08 | rs139516917 |
| IC2 | Psychosis | 18 | ATXN2 | 12 | 8.970e-08 | rs139516917 |
| IC2 | Psychosis | 19 | ACAD10 | 12 | 4.029e-07 | rs139516917 |
| IC2 | Psychosis | 20 | RP11-162P23.2 | 12 | 4.029e-07 | rs139516917 |
| IC2 | Psychosis | 21 | ALDH2 | 12 | 4.298e-07 | rs139516917 |
| IC2 | Psychosis | 22 | HECTD4 | 12 | 1.958e-04 | rs139516917 |

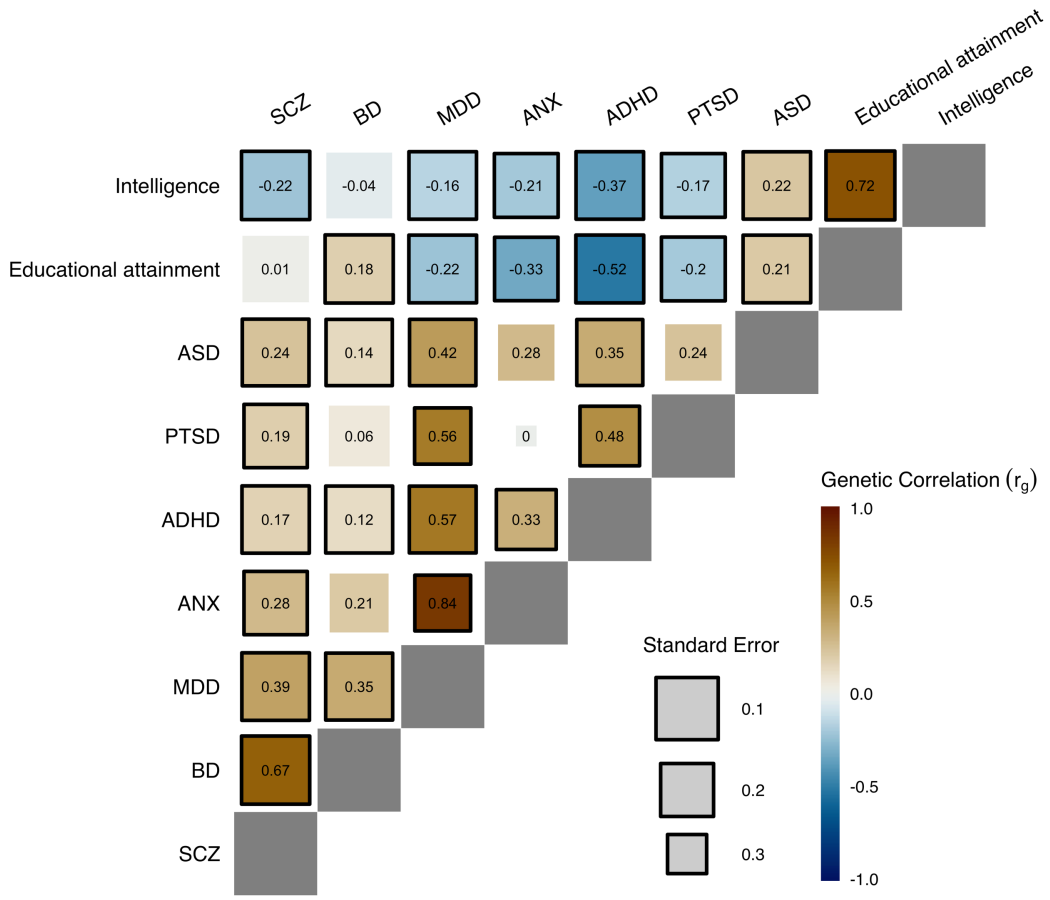
| | | | | | | |
|---|---|---|---|---|---|---|
| IC2 | Psychosis | 23 | RP11-65D24.2 | 13 | 2.944e-08 | rs143983012 |
| IC2 | Psychosis | 24 | POLRMT | 19 | 4.488e-08 | rs574550719 |
| IC2 | Psychosis | 25 | FGF22 | 19 | 2.701e-08 | rs574550719 |
| IC2 | Psychosis | 26 | RNF126 | 19 | 2.701e-08 | rs574550719 |
| IC3 | Anxiety/depression/mental distress | 1 | CABP1 | 12 | 2.226e-08 | rs73222787 |
| IC3 | Anxiety/depression/mental distress | 2 | MLEC | 12 | 6.657e-06 | rs73222787 |
| IC3 | Anxiety/depression/mental distress | 3 | UNC119B | 12 | 1.005e-05 | rs73222787 |
| IC3 | Anxiety/depression/mental distress | 4 | ACADS | 12 | 6.235e-06 | rs73222787 |
| IC3 | Anxiety/depression/mental distress | 5 | SPPL3 | 12 | 1.157e-05 | rs73222787 |
| IC3 | Anxiety/depression/mental distress | 6 | ZCCHC2 | 18 | 2.449e-08 | rs61614162 |
| IC5 | Social instability | 1 | AP1G1 | 16 | NA | rs8057124 |
| IC5 | Social instability | 2 | ATXN1L | 16 | 4.601e-08 | rs8057124 |
| IC5 | Social instability | 3 | IST1 | 16 | 3.624e-08 | rs8057124 |
| IC5 | Social instability | 4 | ZNF821 | 16 | 4.928e-08 | rs8057124 |
| IC6 | Traumatic experiences | 1 | SMYD2 | 1 | NA | rs183777150 |
| IC6 | Traumatic experiences | 2 | PTPN14 | 1 | NA | rs183777150 |
| IC6 | Traumatic experiences | 3 | KCNK2 | 1 | NA | rs183777150 |
| IC6 | Traumatic experiences | 4 | KCTD3 | 1 | NA | rs183777150 |
| IC8 | Felt loved | 1 | EYS | 6 | 3.493e-08 | rs183356400 |
| IC10 | General happiness | 1 | SND1 | 7 | 3.661e-08 | rs4731392 |
| IC10 | General happiness | 2 | LRRC4 | 7 | 1.791e-06 | rs4731392 |
| IC11 | Addiction/mania | 1 | KCNH8 | 3 | NA | rs561625344 |
| IC11 | Addiction/mania | 2 | EFHB | 3 | NA | rs561625344 |
| IC11 | Addiction/mania | 3 | KAT2B | 3 | 1.614e-08 | rs561625344 |
| IC12 | Emotional abuse | 1 | CDH19 | 18 | NA | rs181331621 |
| IC12 | Emotional abuse | 2 | PTPRT | 20 | 1.359e-08 | rs533568724 |

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

462

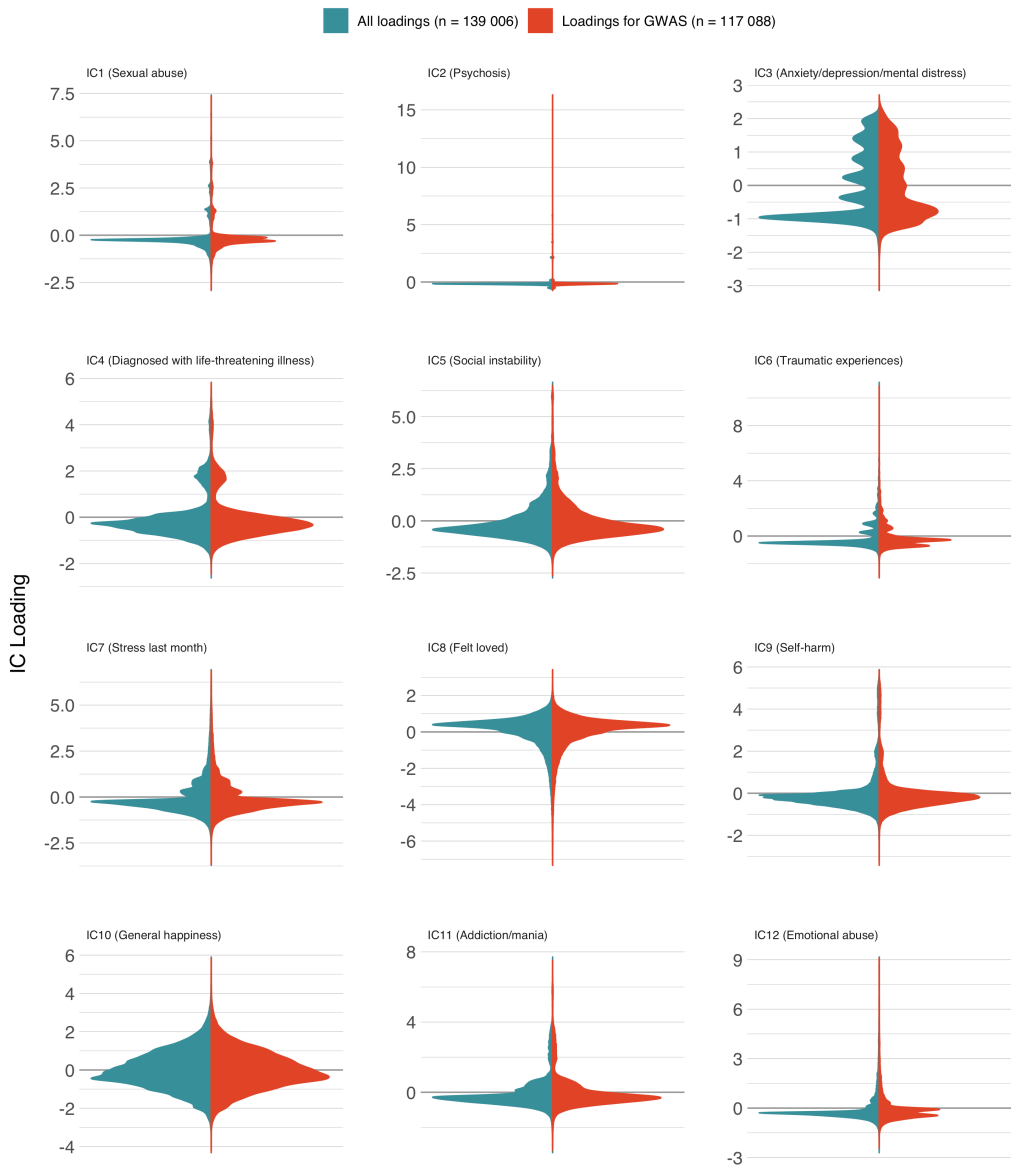
463 **Supplementary Figures**

464

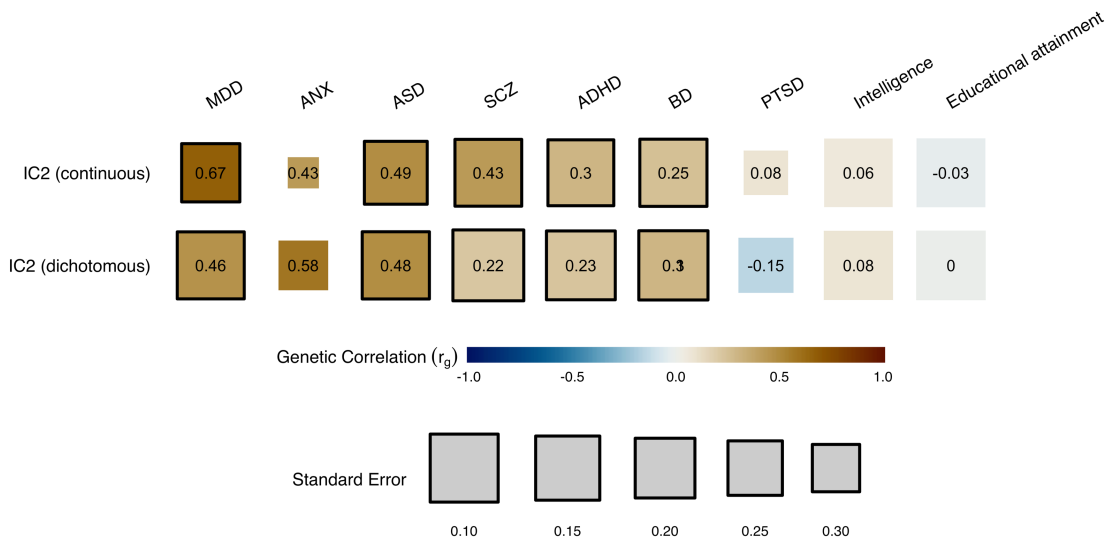



**Suppl. Fig. 1. Genetic correlation between the disorders and cognitive traits** Numbers inside the boxes denote correlation ($r_g$). Size of the boxes reflect standard error. Significant correlations ($p < FDR$) are indicated with a black border. In line with previous reports[9,25], the weakest correlation 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).

465

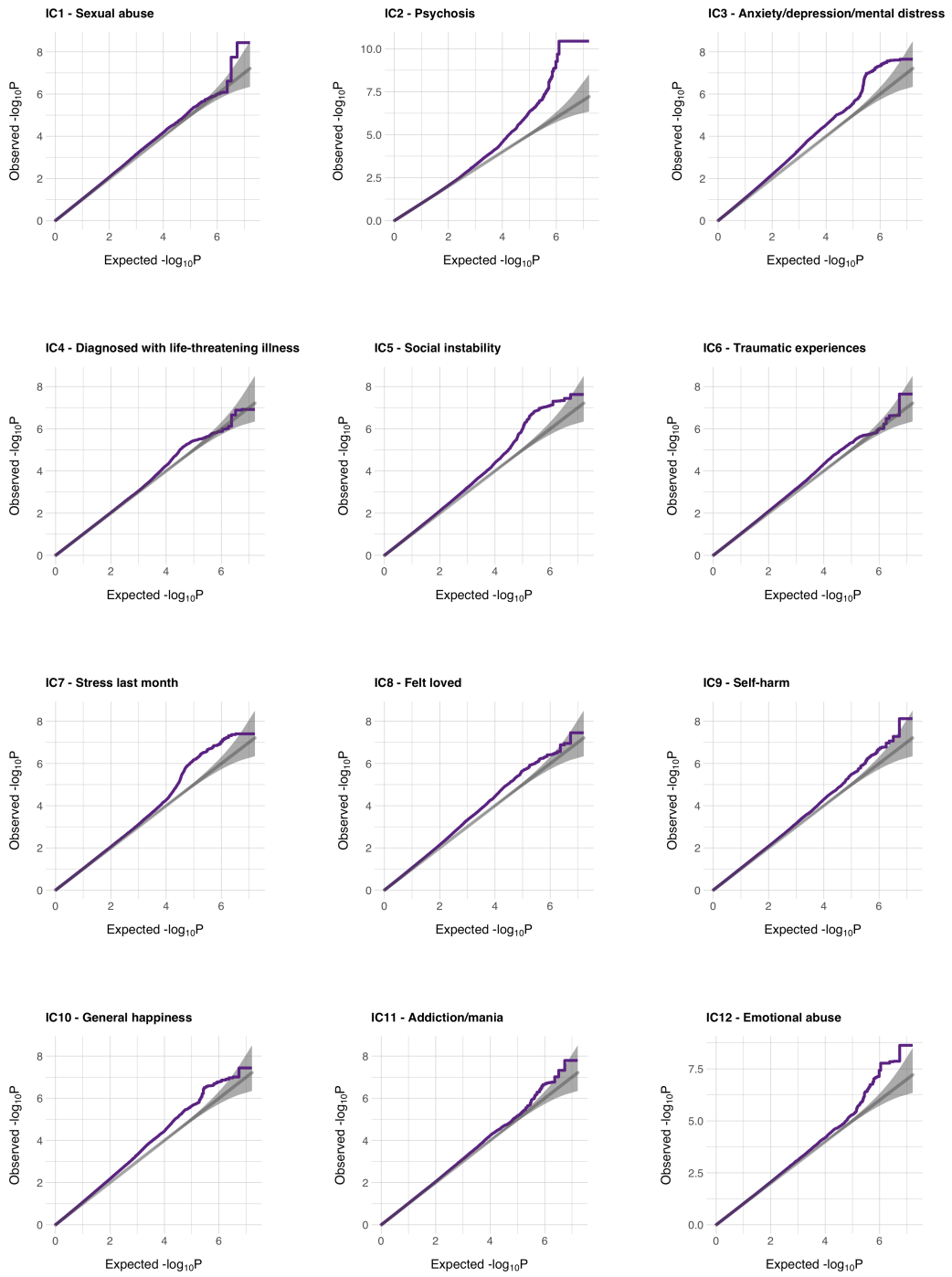



**Suppl. Fig. 2. ICA weight distributions** Density plot of the individual loadings for each independent component. Left half of the plot indicate the loadings for all individuals that were used for ICA. The right side reflects the sample from which genetic data was available and these are the loadings that were used for the GWASs. Distributions on the left show the loadings as provided by the ICA. Distributions on the right are regressed for age (linear and quadratic), sex, and the first 20 genetic principal components. Loadings for IC6, IC7, IC9, and IC11 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.

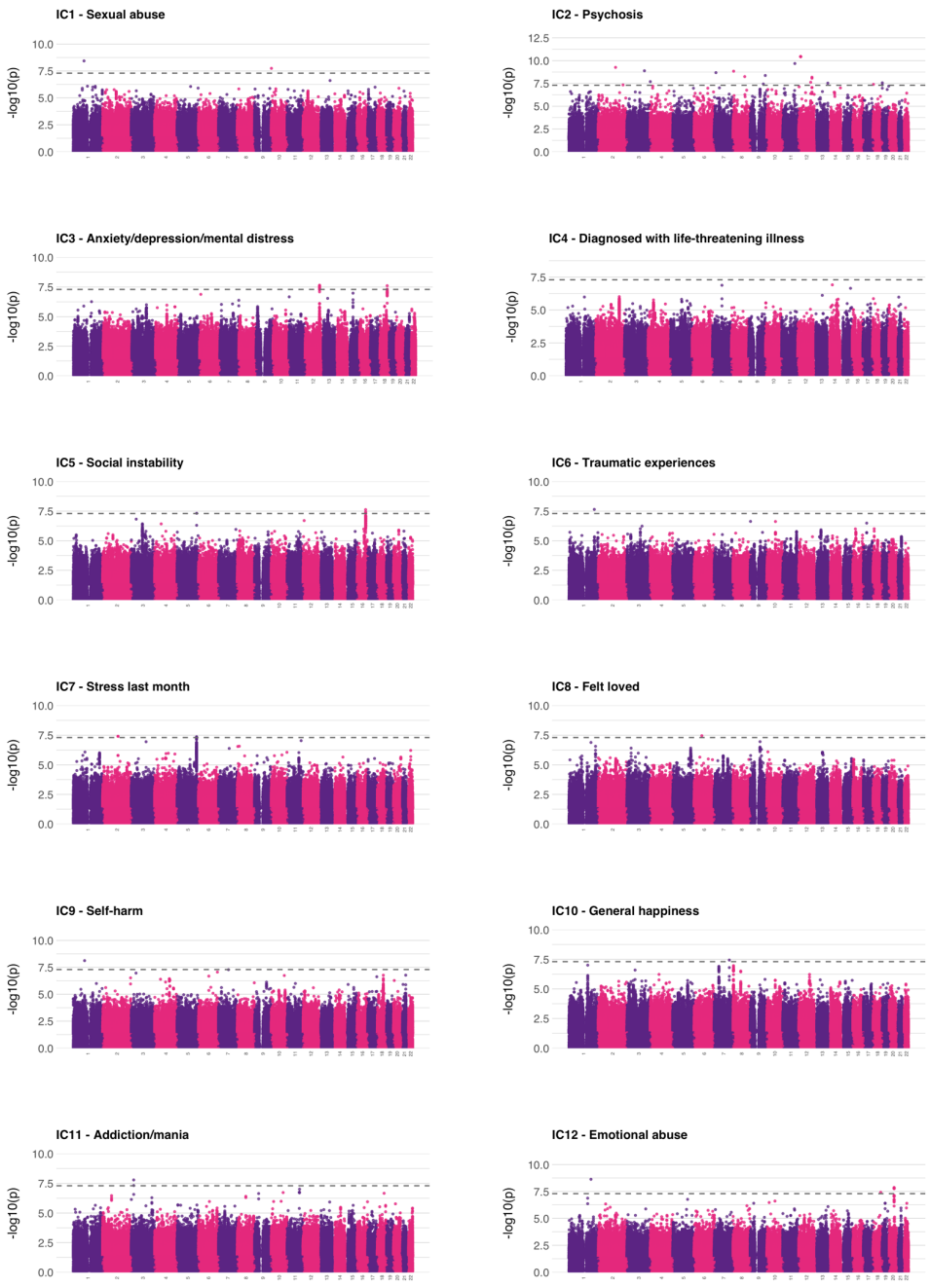



**Suppl. Fig. 3. Comparison between continuous and dichotomous IC2** Due to the low variance in IC2 particularly (Suppl. Fig. 2), we investigated if dichotomizing the loading on that component would improve heritability (i.e. loadings larger than 1 were used as cases, lower than 1 were used as controls 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 ($h2_{dichotomous}$ = 0.018 vs $h2_{continuous}$ = 0.0115) and the reported association with SCZ decreased from $r_{g\ continuous}$ = 0.4271 to $r_{g\ dichotomous}$ = 0.22. The standard error was almost halved after dichotomization ($SE_{continuous}$ = 0.116 vs $SE_{dichotomous}$ = 0.0585), 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 stronger with ANX. Errors were relatively similar.

467




**Suppl. Fig. 4. Q-Q plots** Q-Q plots showing the genetic signal from each of the ICs. IC2 showed the strongest signal despite having the lowest h2. None of the GWAS summary statistics showed any noticeable inflation.




**Suppl. Fig. 5. 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.

469