

Symptom-level genetic modelling identifies novel risk loci and unravels the shared genetic architecture of anxiety and depression

Jackson G. Thorp^{a,b}, Adrian I. Campos^{b,c}, Andrew D. Grotzinger^d, Zachary Gerring^a, Jiyuan An^e, Jue-Sheng Ong^e, Wei Wang^f, 23andMe Research Team^f, Suyash Shringarpure^f, Enda M. Byrne^g, Stuart MacGregor^e, Nicholas G. Martin^c, Sarah E. Medland^h, Christel M. Middeldorp^{i,j,k}, Eske M. Derks^a

^a Translational Neurogenomics, QIMR Berghofer Medical Research Institute, Brisbane, Australia

^b Faculty of Medicine, University of Queensland, Brisbane, Australia

^c Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, Australia

^d Department of Psychology, University of Texas at Austin, Austin, United States

^e Statistical Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Australia

^f 23andMe, Inc., Sunnyvale, United States

^g Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia

^h Psychiatric Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Australia

ⁱ Child Health Research Centre, University of Queensland, Brisbane, Australia

^j Child and Youth Mental Health Service, Children's Health Queensland Hospital and Health Service, Brisbane, Australia

^k Department of Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands

Correspondence:

Jackson Thorp
Translational Neurogenomics
300 Herston Road
Brisbane, QLD 4006
Jackson.Thorp@qimrberghofer.edu.au

Eske Derks
Translational Neurogenomics
300 Herston Road
Brisbane, QLD 4006
Eske.Derks@qimrberghofer.edu.au

1

Abstract

2 Depression and anxiety are highly prevalent and comorbid psychiatric traits that cause
3 considerable burden worldwide. Previous studies have revealed substantial genetic overlap
4 between depression, anxiety, and a closely related personality trait – neuroticism. Here, we
5 use factor analysis and genomic structural equation modelling (Genomic SEM) to investigate
6 the genetic factor structure underlying 28 items assessing depression, anxiety and
7 neuroticism. Symptoms of depression and anxiety loaded on two distinct, although
8 genetically correlated factors, while neuroticism items were partitioned between them. We
9 leveraged this factor structure to conduct multivariate genome-wide association analyses on
10 latent factors of anxiety symptoms and depressive symptoms, using data from over 400,000
11 individuals in the UK Biobank. We identified 89 independent variants for the depressive factor
12 (61 genomic loci, 29 novel) and 102 independent variants for the anxiety factor (73 loci, 71
13 novel). Of these variants, 72% and 78%, respectively, replicated in an independent 23andMe
14 cohort of ~1.9 million individuals with self-reported diagnosis of depression (634,037 cases)
15 and anxiety (624,615 cases). A pairwise GWAS analysis revealed substantial genetic overlap
16 between anxiety and depression but also showed trait-specific genetic influences; e.g.
17 genomic regions specific to depressive symptoms were associated with hypertriglyceridemia,
18 while regions specific to anxiety symptoms were linked to blood pressure phenotypes. The
19 substantial genetic overlap between the two traits was further evidenced by a lack of trait-
20 specificity in polygenic prediction of depressive and anxiety symptoms. Our results provide
21 novel insight into the genetic architecture of depression and anxiety and comorbidity
22 between them.

23

24 **Symptom-level genetic modelling identifies novel risk loci and unravels the**
25 **shared genetic architecture of anxiety and depression**

26 Depression and anxiety are the two most prevalent psychiatric disorders and cause
27 substantial disease burden, accounting for over 10% of years lived with disability worldwide
28 (Vigo, Thornicroft, and Atun 2016; World Health Organization 2017). They are highly
29 comorbid; around three quarters of people with an anxiety disorder also meet diagnostic
30 criteria for major depressive disorder (Lamers et al. 2011). Genetic factors play a substantial
31 role in liability to these disorders, with heritability estimates between 30 to 40 percent for
32 both depression and anxiety (Hettema, Neale, and Kendler 2001; Sullivan, Neale, and Kendler
33 2000). Twin and family studies suggest that their comorbidity is largely explained by shared
34 genetic risk factors (Middeldorp et al. 2005).

35 Neuroticism, characterised as the tendency to experience emotional negativity, such as mood
36 swings, sadness, and worry (McCrae and Costa 1985; Eysenck and Eysenck 1985), is a shared
37 risk factor for depression and anxiety (Kotov et al. 2010; Gray and McNaughton 2000; Ormel
38 et al. 2013; Zinbarg et al. 2016). Genetic factors explain around ~40% of variation in
39 neuroticism (Vukasović and Bratko 2015), and these factors largely overlap with those that
40 affect depression and anxiety (Hettema, Prescott, and Kendler 2004; Jardine, Martin, and
41 Henderson 1984; Fanous et al. 2002; Hettema et al. 2006; Middeldorp et al. 2005). Recent
42 molecular genetic studies have uncovered extensive pleiotropy between these three traits
43 (Purves et al. 2019; Meier et al. 2019; Wray et al. 2018), but little is known about their genetic
44 overlap at a symptom-based level. Here, we investigate the genetic relationship between
45 individual symptoms of anxiety, depression and neuroticism, in order to elucidate their
46 genetic overlap and gain insight into the biological mechanisms underlying comorbidity
47 between anxiety and depression.

48 Genome-wide association studies (GWAS) have accelerated our progress in unravelling the
49 genetic architecture of these psychiatric traits. The general observation is that complex traits
50 are influenced by large numbers of genetic variants with small individual effect sizes (i.e. high
51 polygenicity) and consequently require very large sample sizes in order to detect them.
52 Recent GWAS have identified over 100 independent and robustly associated single nucleotide
53 polymorphisms (SNPs) for depression and neuroticism (Howard et al. 2019; Wray et al. 2018;
54 Nagel, Jansen, et al. 2018; Luciano et al. 2018). In comparison, genetic studies of anxiety are
55 still underpowered, with the largest studies to date having identified 5 genetic risk loci for
56 lifetime anxiety disorder (Purves et al. 2019) and 6 loci for anxiety symptoms (Levey et al.
57 2020). While the full-scope of risk conferring genetic loci remains to be discovered for these
58 traits, bivariate genomic methods (e.g., LD Score Regression; Bulik-Sullivan et al. 2015) have
59 been used to obtain estimates of overall levels of genetic overlap. Consistent with family-
60 based results, large SNP-based genetic correlations have been reported across depression,
61 anxiety, and neuroticism ($r_g > 0.70$) (Wray et al. 2018; Nagel, Jansen, et al. 2018). Moreover,
62 pairwise comparison of genomic loci implicated in neuroticism and major depression found a
63 substantial portion (~70%) of regions associated with major depression are also associated
64 with neuroticism (Adams et al. 2019).

65 The extensive genetic overlap between depression, anxiety, and neuroticism may partly
66 reflect overlap in item content and diagnostic criteria used to measure these traits (Ormel,
67 Riese, and Rosmalen 2012). Numerous scales of neuroticism include “sub-scales” or “facets”
68 of both depression and anxiety (e.g. NEO Personality Inventory-Revised (NEO-PI-R); California
69 Psychological Inventory (CPI) Big Five), and many items within these scales closely resemble
70 symptom measures of both depression and anxiety. For example, NEO-PI-R neuroticism items
71 *“Sometimes I feel completely worthless”* and *“I have sometimes experienced a deep sense of*

72 *guilt or sinfulness*” are very similar to the DSM-5 major depression symptom “*Feelings of*
73 *worthlessness or excessive or inappropriate guilt*”. Neuroticism is therefore not operationally
74 distinct, implying that by studying its underlying components one could gain valuable insight
75 into symptoms of anxiety and depression and their genetic influences. Indeed, hierarchical
76 clustering of individual neuroticism items in the Short-form Eysenck Personality
77 Questionnaire-Revised (EPQR-S) (Eysenck, Eysenck, and Barrett 1985) revealed two genetic
78 item clusters, termed ‘depressed affect’ and ‘worry’, displaying stronger genetic overlap with
79 depression and anxiety, respectively (Nagel, Watanabe, et al. 2018).

80 Item or symptom-level genetic analyses allow investigation of the underlying genetic
81 structure of a trait and have proven useful in disentangling genetic and phenotypic
82 heterogeneity of neuroticism and depression (Nagel, Watanabe, et al. 2018) (Thorp et al.
83 2019). In the present study, we extend this approach across multiple traits and investigate
84 the genetic factor structure underlying 28 symptoms of depression, anxiety, and neuroticism.
85 We apply genomic structural equation modelling (Genomic SEM) (Grotzinger et al. 2019), a
86 recently developed multivariate method which enables estimation of the joint genetic
87 architecture of multiple complex traits based on summary statistics from GWAS. This allows
88 genetic subtypes or combinations of genetically similar symptoms to be identified, leading to
89 increased statistical power for the discovery of genetic loci and improved understanding of
90 the comorbidity and genetic overlap across traits. We sought to answer three questions: (1)
91 how do items used to measure neuroticism genetically relate to symptoms of depression and
92 anxiety; (2) can we leverage genetic overlap with neuroticism to boost power for the
93 discovery of genetic risk loci for anxiety and depressive symptoms; and (3) can we identify
94 genetic loci that differentiate anxiety and depressive symptoms? First, we model the genetic
95 factor structure across the three traits using item-level questionnaire data from the UK

6

96 Biobank (anxiety and depressive symptoms, $N = \sim 135,000$; neuroticism, $N = \sim 400,000$). Then,
97 we leverage this factor structure to identify genetic loci for latent factors of depressive
98 symptoms and anxiety symptoms using Genomic SEM. Finally, we identify genomic regions
99 that are unique to or shared by depressive and anxiety symptoms to gain insight into the
100 genetic architecture of these traits and the comorbidity between them.

101

102 **Results**

103 **Factor analysis of symptoms of depression, anxiety, and neuroticism**

104 We explored genetic overlap between anxiety symptoms, depressive symptoms, and
105 neuroticism by modelling the genetic factor structure of items used to measure these traits.
106 Item-level genome-wide association analyses were conducted individually on each of 28 items
107 of neuroticism (12 items; EPQR-S), anxiety (7 items; GAD-7), and depression (9 items; PHQ-9),
108 in ~135,000 UKB participants (see Supplementary table 1 for item-specific sample sizes). LD
109 Score regression was used to calculate genetic correlations between all item pairs and an
110 exploratory factor analysis (EFA) was then conducted on this genetic correlation matrix. A
111 minimum average partial (MAP) test suggested the optimal number of factors to extract is
112 three (consistent with the eigenvalue-greater-than-one rule; see Supplementary Table 2).
113 Factor loadings of the 3-factor model are presented in Figure 1 (and Supplementary Table 3).
114 Depression items had high loadings (> 0.4) on genetic factor 1, anxiety items had high loadings
115 on genetic factor 2 (except for the item 'irritability', which loaded onto factor 1), and
116 neuroticism items loaded either highly on factor 1 (5 items) or factor 2 (7 items). Genetic
117 factor 3 is characterised by relatively low loadings, which are positive for items of depression
118 and anxiety, and negative for items of neuroticism. The items with the highest loadings on
119 factor 3 are mostly somatic symptoms; therefore this factor may contain variance that
120 separates a psychosomatic facet of depression and anxiety from neuroticism. Factor 3 is
121 largely underpowered for further genetic analysis; we therefore restrict subsequent analyses
122 to factor 1 and factor 2. In this paper, we refer to these two genetic latent factors as DEP
123 (depressive symptoms) and ANX (anxiety symptoms).

124 We then submitted this factor structure model to Genomic SEM (retaining standardised
125 loadings > 0.4) to assess its fit to the data (taking into account uncertainty in covariance
126 estimates). The model provided adequate fit to the data (CFI = 0.890; SRMR = 0.087; factor
127 loadings in Supplementary Table 4). As a first validation of the DEP and ANX latent factors that
128 included the neuroticism items, we calculated genetic correlations between the latent factors
129 and sum-scores of depressive symptoms (PHQ-9; N = 135,149) and anxiety symptoms (GAD-
130 7; N = 135,747). The genetic correlations between the PHQ-9 sum-score and the DEP latent
131 factor ($r_g = 0.94$, 95% CI [0.87,1.01]), and the GAD-7 sum-score and ANX latent factor ($r_g = 0.93$,
132 95% CI [0.86,1.00]) were not significantly different from one, suggesting the DEP and ANX
133 latent factors are good proxies for anxiety and depressive symptoms. The genetic correlation
134 between the DEP and ANX factors was moderately high ($r_g = 0.80$, 95% CI [0.77,0.83]) and was
135 similar to the correlation between the PHQ-9 and GAD-7 sum-scores ($r_g = 0.83$, 95% CI
136 [0.72,0.95]).

137 **Multivariate GWAS of anxiety and depressive symptoms**

138 Having identified the genetic latent factor structure within the UKB sample, our next step was
139 to leverage this structure to identify genomic risk loci for the DEP and ANX latent factors. To
140 maximise power in the multivariate GWAS, we expanded the neuroticism items to the full
141 UKB set (i.e. an additional ~270k individuals who completed the neuroticism questionnaire
142 but not the depressive or anxiety symptoms questionnaires were included; each neuroticism
143 item N = ~400k). We examined whether this changed the factor structure by fitting the EFA-
144 derived model to the genetic covariance matrix of the full UKB set. The EFA-derived factor
145 structure retained adequate fit to the data (CFI = 0.893; SRMR = 0.088), while standardised
146 factor loadings were concordant with loadings before expanding the sample size of

147 neuroticism ($r = 0.95$; factor loadings in Supplementary Table 5). The genetic correlation
148 between the DEP and ANX factors remained the same ($r_g = 0.79$, 95% CI [0.77,0.81]).

149 Multivariate GWAS were conducted by estimating the effects of individual SNPs on the DEP
150 and ANX latent factors using Genomic SEM. The GWAS of the DEP factor identified 7,677
151 genome-wide significant SNPs ($P < 5 \times 10^{-8}$), tagging 89 independent SNPs in 62 genomic risk
152 loci (see Figure 2a and Supplementary Table 6). Of these loci, 33 overlap with a previous
153 GWAS of depressive symptoms or major depression (Wray et al. 2018; Nagel, Jansen, et al.
154 2018; Howard et al. 2019; Okbay et al. 2016; Howard et al. 2018; Hyde et al. 2016; Turley et
155 al. 2018; Baselmans et al. 2019), while 29 are novel depression loci. For the ANX factor, 11,163
156 SNPs reached genome-wide significance, tagging 102 independent SNPs in 73 loci (see Figure
157 2a and Supplementary Table 7). Two loci overlapped with a previous GWAS of anxiety
158 disorders (Purves et al. 2019), and 71 loci are novel for anxiety. LD Score Regression analyses
159 indicated that our GWAS results were not due to uncontrolled inflation for either DEP
160 (intercept = 0.991, s.e. = 0.010) or ANX (intercept = 1.003, s.e. = 0.011). Effect sizes of
161 independent significant SNPs showed large concordance between DEP and ANX. Three
162 variants had significantly ($P < 2.76 \times 10^{-4}$) different effect sizes: rs613872 was associated only
163 with DEP, and two SNPs (rs62250713 and rs391957) with ANX only (see Figure 2b).

164 We conducted a replication of the significant independent SNPs in a cohort of research
165 participants from 23andMe with information on self-reported diagnosis of depression
166 (634,037 cases; 1,308,690 controls) and anxiety (624,615 cases; 1,310,854 controls). For the
167 DEP replication, 81 variants were tested (8 SNPs were unavailable or of insufficient quality in
168 the 23andMe cohort). Of these variants, 58 were significant after Bonferroni correction ($\alpha =$
169 $0.05/81$; $P < 6.17 \times 10^{-4}$) and had the same direction of effect, and 40 reached genome-wide

170 significance ($P < 5 \times 10^{-8}$). 93 ANX variants were tested for replication (9 SNPs were of
171 insufficient quality in the 23andMe cohort), of which 73 were significant after Bonferroni
172 correction ($\alpha = 0.05/93$; $P < 5.38 \times 10^{-4}$) and had the same direction of effect, and 39 reached
173 genome-wide significance ($P < 5 \times 10^{-8}$).

174 MAGMA was used to conduct gene-based association tests and gene-set enrichment
175 analyses. We identified 255 genes and 9 gene sets associated with DEP, and 325 genes and
176 21 gene sets associated with ANX (significant after Bonferroni correction; Supplementary
177 Tables 8-11). There was substantial overlap with respect to enriched functional categories
178 between the two traits (110 genes and 5 gene sets).

179 **Polygenic risk prediction**

180 To further validate the latent factors, we used polygenic risk scores (PRS) derived from the
181 ANX and DEP summary statistics to predict both depressive and anxiety symptoms in an
182 independent sample ($N = 4,434$). The PRS includes all SNPs, using Bayesian multiple regression
183 to account for linkage disequilibrium (see Methods). PRS for DEP significantly predicted
184 depressive symptoms ($P = 2.69 \times 10^{-10}$), explaining 1.05% of variance (see Figure 3). Similarly,
185 PRS for ANX significantly predicted anxiety symptoms ($P = 4.80 \times 10^{-14}$), explaining 1.53% of
186 variance. For comparison, we also calculated PRS from PHQ-9 sum-score, GAD-7 sum-score,
187 neuroticism, major depression (Wray et al. 2018), depression (Howard et al. 2019) and anxiety
188 disorders (Purves et al. 2019). The DEP and ANX latent factors explained a greater amount of
189 variance than the sum-scores (proportional increase in explained variance was 185% and
190 237%, respectively), indicating that polygenic prediction was improved by the combination of
191 leveraging information from the neuroticism items and taking a more psychometrically
192 informed approach (i.e., factor analysis) to phenotype construction. Overall, specificity in

193 polygenic prediction was low; PRS for depression phenotypes explained an equal or greater
194 amount of variance in anxiety symptoms than depressive symptoms (see Figure 3 and
195 Supplementary Table 12).

196 **Genetic correlations with other complex traits**

197 We estimated genetic correlations between the ANX and DEP latent factors and a range of
198 human complex traits (see Figure 4 and Supplementary Table 13). While patterns of
199 correlations were similar in magnitude and direction across most of the traits, some traits
200 showed differential genetic overlap with DEP and ANX. Smoking related phenotypes
201 (initiation, age of initiation, cigarettes per day, and cessation) genetically correlated with DEP
202 ($|r_g| > 0.27$), but not with ANX. Genetic overlap with socioeconomic traits (Townsend
203 deprivation index, household income, and educational attainment) was consistently larger for
204 DEP than ANX. Conversely, ANX showed stronger overlap with obsessive compulsive disorder,
205 anorexia nervosa, and schizophrenia.

206 **Shared and trait-specific genetic influences**

207 We sought to identify trait-specific genomic regions by conducting a pairwise analysis of the
208 DEP and ANX GWAS summary statistics in order to characterise regions as pleiotropic, or
209 uniquely associated with either DEP or ANX. We used gwas-pw (Pickrell et al. 2016) to
210 estimate the posterior probability that a given genomic region is associated with (1) DEP only,
211 (2) ANX only, (3) both DEP and ANX, and (4) both DEP and ANX but via separate causal
212 variants. Out of the 1703 tested regions, 123 (7%) had a posterior probability greater than 0.5
213 of containing a causal variant for at least one of the two traits. Of these regions, 10 were
214 uniquely associated with DEP, 20 were uniquely associated with ANX, 71 were associated with

215 both DEP and ANX, and 22 were associated with both traits but via separate variants (see
216 Figure 5a and Supplementary Table 14).

217 Next, we conducted gene-based association tests separately for regions that were specific to
218 DEP, specific to ANX, or shared. We identified 26 genes significantly associated with DEP-
219 specific regions, 47 genes with ANX-specific regions, and 144 with shared regions (see
220 Supplementary Tables 15-17). To further identify genes for trait-specific and shared regions
221 we also mapped SNPs (that reached genome-wide significance in the GWAS) to genes based
222 on proximity, eQTL, and chromatin interactions. These three strategies mapped 49 genes to
223 DEP-specific regions, 74 genes to ANX-specific regions and 470 genes to shared regions (see
224 Supplementary Tables 18-20). The total number of genes identified (across all methods) was
225 63, 102 and 509 genes for DEP-specific, ANX-specific and shared regions, respectively (see
226 Figure 5b).

227 Using all genes linked to trait-specific and shared regions, we conducted gene-set enrichment
228 analysis against gene-sets defined by traits in the NHGRI-EBI GWAS Catalog (Buniello et al.
229 2019). Genes prioritised for DEP-specific regions were significantly enriched in a gene-set for
230 hypertriglyceridemia ($P = 1.92 \times 10^{-7}$). Genes mapped to ANX-specific regions showed
231 enrichment in multiple gene-sets: schizophrenia ($P = 1.99 \times 10^{-11}$), autism spectrum disorder
232 or schizophrenia ($P = 2.03 \times 10^{-22}$), response to cognitive-behavioural therapy in major
233 depressive disorder ($P = 1.30 \times 10^{-13}$), and multiple gene sets related to blood pressure: mean
234 arterial pressure ($P = 4.37 \times 10^{-7}$), systolic blood pressure ($P = 3.24 \times 10^{-7}$), pulse pressure
235 ($P = 2.65 \times 10^{-7}$), and hypertension ($P = 1.27 \times 10^{-5}$). Genes prioritised for shared regions were
236 significantly enriched in 109 gene-sets (see Supplementary Table 21), including autism
237 spectrum disorder or schizophrenia ($P = 1.71 \times 10^{-137}$), blood protein levels ($P = 1.04 \times 10^{-40}$),

13

238 sarcoidosis ($P = 7.89 \times 10^{-33}$), systolic blood pressure ($P = 3.24 \times 10^{-4}$), and triglyceride levels

239 ($P = 1.81 \times 10^{-3}$).

240

241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263

Discussion

Recent studies have revealed substantial genetic correlations between depression, anxiety, and neuroticism (Wray et al. 2018; Levey et al. 2020; Howard et al. 2019; Nagel, Jansen, et al. 2018). The extensive genetic overlap partly reflects the overlap in items used to measure these traits, which motivated us to explore the genetic factor structure underlying symptoms of depression, anxiety, and neuroticism. Leveraging the underlying factor structure, we conduct GWAS on latent factors of anxiety and depressive symptoms using data from over 400,000 individuals in the UK Biobank. We identify novel and robust genetic associations and present findings for the largest GWAS of anxiety to date (102 GWS independent variants, ~78% replicate in an independent cohort of self-reported diagnosis of anxiety). We also characterise shared and trait-specific genetic influences and report on gene-set analyses targeted at understanding shared and trait-specific aetiology.

Modelling the genetic factor structure of symptoms of depression, anxiety, and neuroticism revealed two key findings. First, symptoms of anxiety and depression loaded onto different factors, although the genetic correlation between the factors was high. This implies that while symptoms of depression and anxiety are closely related, symptoms are genetically more similar to symptoms within the same disorder than symptoms between disorders. This observation is consistent with a phenotypic-level network analysis of the PHQ-9 and GAD-7 items within a psychiatric sample (Beard et al. 2016), which found that symptom connections were higher within each disorder than between disorders. The large degree of genetic overlap between symptoms of anxiety and depression is in agreement with twin-based symptom-level analyses (Kendler et al. 1987), although our results suggest some specificity of common genetic risk factors. Second, neuroticism items loaded highly onto either the DEP or ANX

264 factors rather than forming their own factor, suggesting that at a genetic level neuroticism is
265 not itself a distinct construct, and likely encapsulates (sub-clinical) symptoms of both
266 depression and anxiety. The partitioning of neuroticism items among two distinct factors is in
267 line with the results of a hierarchical clustering analysis of the genetic correlation matrix
268 derived from these neuroticism items when analysed in isolation (Nagel, Watanabe, et al.
269 2018).

270 By leveraging item-level genetic overlap with neuroticism, we substantially increased
271 statistical power to identify genomic risk loci for depressive and anxiety symptoms. Our GWAS
272 of depressive symptoms identified 62 genomic risk loci, of which 29 are novel. For anxiety
273 symptoms, we identified 73 genomic loci (71 novel), a substantial increase from previous
274 studies which have found 6 loci for anxiety disorders (Purves et al. 2019; Meier et al. 2019)
275 and 6 loci for anxiety symptoms (Levey et al. 2020). The identification of a large number of
276 replicable genomic loci for anxiety unlocks the possibility of leveraging statistical genetic
277 approaches (e.g. Mendelian randomisation or drug repositioning) which were not possible
278 with previous anxiety GWAS. We implemented one such approach, a regional pairwise
279 analysis, in order to disentangle the shared genetic architecture of anxiety and depression
280 and identify trait-specific genetic loci.

281 Our results suggest that depression-specific genomic regions are linked to
282 hypertriglyceridemia. This is consistent with several studies that have found an association
283 between depression and triglyceride levels (Igna, Julkunen, and Vanhanen 2011; Richter,
284 Juckel, and Assion 2010; Akbaraly et al. 2009; Glueck et al. 1993), and previous GWAS of
285 depression have reported significant genetic correlations with triglycerides ($r_g = 0.14$; Wray
286 et al. 2018; Howard et al. 2019). Shared regions were also enriched in a gene-set for

287 triglyceride levels, suggesting that while both depression and anxiety contain a metabolic
288 component, this component may be larger in depression than anxiety. Conversely, genes
289 mapped to anxiety-specific regions were enriched in gene-sets related to multiple blood
290 pressure phenotypes. While perhaps unsurprising given increased blood pressure is a direct
291 physiological effect of the stress response, anxiety has also been linked to increased risk of
292 hypertension (Pan et al. 2015). Interestingly, genes unique to ANX were also enriched in a
293 gene-set linked to response to cognitive-behavioural therapy in major depression (Rayner et
294 al. 2019), suggesting that the presence of comorbid anxiety symptoms may influence
295 treatment response for depression. Indeed, comorbid anxiety and major depressive disorders
296 are associated with higher symptom severity and impairment, disorder persistence, and
297 reduced response rates (Young, Mufson, and Davies 2006; Kessler et al. 2007; Emmanuel,
298 Simmonds, and Tyrer 1998; Walker et al. 2000; Altamura et al. 2004). There was also
299 significant enrichment of ANX-specific regions in schizophrenia gene-sets, consistent with a
300 larger genetic correlation of schizophrenia with the ANX factor than with DEP factor. Anxiety
301 symptoms are highly prevalent in patients with schizophrenia (Achim et al. 2009) and are
302 associated with the positive symptom domain of schizophrenia (Emsley et al. 1999).

303 Genetic correlations with other complex traits were largely concordant in direction and
304 magnitude across DEP and ANX. Smoking related phenotypes however, were genetically
305 correlated with depression but not with anxiety. Observational studies on the association of
306 smoking with anxiety and depression are largely mixed with regards to the direction of effect
307 (Fluharty et al. 2017). We find moderate genetic correlations between DEP and smoking
308 initiation, cigarettes per day, cessation (positive), and age of initiation (negative). Genetic
309 correlations were not significant between any smoking phenotype and ANX, suggesting a
310 unique relationship between smoking behaviour and depression.

311 It is well established that depression and anxiety share a substantial amount of genetic
312 liability. Our results provide additional evidence for this notion, with a high genetic correlation
313 between the DEP and ANX factors ($r_g = 0.80$), consistent with the genetic overlap between
314 major depression and anxiety disorders in previous studies ($r_g = 0.75-0.80$). Further, the
315 amount of polygenic overlap (i.e., the fraction of genetic variants causally influencing both
316 traits) was considerable, with 71 out of 123 genomic regions containing a causal variant that
317 influences both traits. We note that this is likely an underestimate of the proportion of shared
318 genetic effects, as the correction for sample overlap in pairwise GWAS is conservative and
319 some truly shared genetic effects may be corrected out (Pickrell et al. 2016). Given the high
320 genetic overlap and substantial comorbidity between depression and anxiety (Kessler et al.
321 2007; Lamers et al. 2011), it is unsurprising that there was very little differentiation in
322 polygenic prediction of depressive and anxiety symptoms with PRS derived from latent
323 factors. This was not specific to the latent factors; poor specificity was also seen with PRS
324 derived from clinical depression and anxiety phenotypes (Wray et al. 2018; Purves et al.
325 2019).

326 We performed multiple analyses to validate the latent factors representing depression and
327 anxiety. First, genetic correlations between the latent factors and the PHQ-9 and GAD-7 sum-
328 scores (i.e., between the DEP factor and the sum-score of the 9 depressive symptoms, and
329 between the ANX factor and the sum-score of the 7 anxiety symptoms) were very high and
330 not significantly different from one. These correlations persisted when incorporating an
331 additional ~270,000 individuals with neuroticism information only into the latent factors.
332 Second, a substantial proportion (~72% for DEP and ~78% for ANX) of genome-wide
333 significant variants replicated in a large, independent cohort of self-reported diagnosis of

334 anxiety and depression. Third, polygenic risk score analyses showed that the latent factors
335 significantly predicted depressive and anxiety symptoms in a second independent sample.

336 The findings of the present study should be interpreted in light of some key limitations. The
337 DEP and ANX latent factors represent depressive and anxiety symptoms within a population-
338 based cohort. While these factors had relatively high genetic correlations with clinical
339 phenotypes, it is likely that they do not capture the entire spectrum of genetic influences on
340 major depression and anxiety disorders (Schwabe et al. 2019; Kendler et al. 2019; Cai et al.
341 2020). For example, the episodic nature of major depression or the persistence of generalised
342 anxiety are not well captured by symptom questionnaires. Second, by leveraging neuroticism
343 items in the extended UKB sample, we increased statistical power to identify variants for
344 symptoms which overlap with depression and anxiety (predominately 'psychological'
345 symptoms). However, the somatic, motor, and neurovegetative symptoms characteristic of
346 major depression and anxiety disorders are not well represented in neuroticism, and
347 consequently not within our phenotypes.

348 Over a decade of molecular genetic studies have confirmed the presence of widespread
349 pleiotropy across psychiatric disorders (Smoller et al. 2019; Lee et al. 2019; Watanabe et al.
350 2019). The substantial sharing of genetic risk factors challenges the utility of analysing discrete
351 diagnostic categories of psychopathology defined by current classification systems (such as
352 the Diagnostic and Statistical Manual of Mental Disorders 5th edition; DSM-5) in the discovery
353 of genetic risk loci and prediction of disease. There has been recent interest in analysing
354 psychiatric endophenotypes as an alternative (Sanchez-Roige and Palmer 2020), with
355 initiatives such as the Hierarchical Taxonomy of Psychopathology (HiTOP) (Waszczuk et al.
356 2020) and Research Domain Criteria (RDoC) (Insel et al. 2010) facilitating such efforts.

357 Symptom-level analyses are one such endophenotypic approach which may prove useful in
358 in advancing our understanding of the genetic aetiology of psychopathology, by allowing the
359 discovery of symptom-specific genetic influences and the identification of genetic subtypes
360 and trans-diagnostic factors of genetic liability. We show that analysing genetically
361 homogenous combinations of symptoms across traits can increase statistical power to
362 identify loci and elucidate comorbidity and genetic overlap between psychiatric phenotypes.

363

364

Methods

365 UK Biobank

366 Data for the main analyses came from the UK Biobank (UKB), a major health data resource
367 containing phenotypic information on a wide range of health-related measures and
368 characteristics in over 500,000 participants from the United Kingdom general population
369 (Bycroft et al. 2018). Participants were excluded from the present study based on ancestry,
370 relatedness and withdrawn consent. Participants were included if they were of white British
371 ancestry, identified through self-reported ethnicity and ancestral principal components.
372 Participants who self-reported as not white British, but for whom the first two genetic
373 principal components indicated them to be genetically similar to those of white British
374 ancestry were also included in order to maximise sample size (MacGregor et al. 2018).

375 *Depressive and anxiety symptoms*

376 Depressive symptoms were assessed with the 9-item Patient Health Questionnaire (PHQ-9)
377 (Kroenke, Spitzer, and Williams 2001), and anxiety symptoms with the 7-item Generalised
378 Anxiety Disorder scale (GAD-7) (Spitzer et al. 2006). Over 150,000 participants completed the
379 PHQ-9 and GAD-7 as part of a UKB mental health follow-up questionnaire administered online
380 in 2016 (Davis et al. 2020). Each item assesses the frequency of a particular symptom over
381 the past two weeks, rated on a four-point ordinal scale: (0) not at all, (1) several days, (2)
382 more than half the days, (3) nearly every day.

383 The ordinal scale of measurement of these items complicates interpretation of SNP-based
384 heritability (h^2_{SNP}) estimates. As h^2_{SNP} estimates are utilised in Genomic SEM, we transformed
385 each item into a binary phenotype in order for interpretable h^2_{SNP} estimates to be produced.
386 Items were dichotomised such that an item was considered to be endorsed if the item score

387 was one or greater (several days, more than half the days, or nearly every day), and not
388 endorsed if the score was zero (not at all). A cut-off score of one was used in order to
389 maximise the number of participants who endorsed an item and hence statistical power
390 (genetic correlations between the ordinal-scale items and binary items were all > 0.95 ;
391 median $r_g = 0.98$).

392 ***Neuroticism***

393 Neuroticism was measured using the 12-item Eysenck Personality Questionnaire-Revised
394 Short Form (EPQR-S) (Eysenck, Eysenck, and Barrett 1985), with each item assessed on a
395 dichotomous scale ('yes' or 'no'). The questionnaire was administered to the entire UKB
396 cohort (~500,000 participants).

397 **Genome-wide association analyses**

398 GWAS analyses of the 28 individual items (9 depression items, 7 anxiety items, 12 neuroticism
399 items) were conducted via logistic regression in PLINK v2.00a (Chang et al. 2015). If two
400 individuals in the sample were related ($\pi\text{-hat} > 0.2$) one individual was removed
401 (preferentially from the control set if the related individuals were in both case and control
402 sets). GWAS analyses of the PHQ-9, GAD-7, and EPQR-S sum-scores were conducted via linear
403 regression. Analyses were limited to autosomal SNPs with high imputation quality score (INFO
404 score ≥ 0.80) and a minor allele frequency of 1% or higher, resulting in 9,417,325 SNPs being
405 tested for association. Age, sex, genotyping array, and 20 principal components were
406 included as covariates.

407 **Factor analysis**

408 We first explored the genetic factor structure by conducting an exploratory factor analysis
409 (EFA) based on the genetic correlation matrix of the 28 items. The EFA was conducted using

410 only participants who completed the UKB mental health questionnaire (N range = 132,602 –
411 137,461; see Supplementary Table 1 for item-specific sample sizes). As participants who
412 completed the UKB mental health questionnaire differ significantly from the entire UKB
413 cohort (e.g. higher educational attainment, higher socio-economic status, lower rates of
414 smoking; Davis et al. 2020), we restricted the EFA to a subset for neuroticism to ensure these
415 systematic differences did not bias the EFA. Cross-trait LD Score Regression was used to
416 estimate genetic correlations between each of the 28 items. These estimates are not biased
417 by sample overlap (Bulik-Sullivan et al. 2015). The ‘psych’ R package was used to conduct the
418 EFA, with an ordinary least squares extraction method and oblimin rotation method. Two
419 procedures were used to decide on the optimal number of factors to extract: (1) a minimum
420 average partial (MAP) test (Velicer 1976) (the lowest average squared partial correlation
421 indicates the number of factors to extract), and the eigenvalue-greater-than-one rule (Kaiser
422 1960) (factors with an eigenvalue above 1 are extracted).

423 The factor model identified using EFA (retaining factor loadings > 0.4) were subsequently
424 carried forward in a confirmatory factor analysis (CFA) in Genomic SEM (Grotzinger et al.
425 2019). This was done to assess the fit of the factor model to the data while taking into account
426 uncertainty in covariance estimates, and to allow the estimation of genetic correlations
427 between latent factors and external traits (i.e. PHQ-9 and GAD-7 sum-scores). The default
428 Diagonally Weighted Least Squares (DWLS) estimator was used. SNP-based heritability
429 estimates (diagonal of the genetic covariance matrix) were converted to the liability scale,
430 where the population prevalence of the items was estimated from the UKB sample
431 (population prevalence = sample prevalence; see Supplementary Table 1).

432

433 **Multivariate genome-wide association analyses**

434 GWAS of the latent factors of anxiety and depressive symptoms were conducted in Genomic
435 SEM. All summary statistics were standardized with respect to the variance in the phenotype
436 (i.e., STDY) using the *sumstats* function in Genomic SEM. The factor structure identified in the
437 EFA was specified as the model. SNPs tested for association in the univariate item-level GWAS
438 and also contained in the 1000 genomes phase 3 reference sample (with MAF > 0.01) were
439 included, resulting in the analysis of 7,746,079 SNPs. We applied the conventional genome-
440 wide significance threshold of $P < 5 \times 10^{-8}$.

441 The results were annotated using FUMA (Watanabe et al. 2017). Significant SNPs were
442 clumped into blocks high in linkage disequilibrium (the non-random association of alleles at a
443 specific locus; LD) using a threshold of $r^2 < 0.10$ (correlation between allele frequencies of two
444 SNPs). Genomic risk loci were identified by merging independent SNPs if $r^2 \geq 0.10$ and their
445 LD blocks were physically close to each other at a distance of 1000 kb.

446 **23andMe replication cohort**

447 In the 23andMe replication analysis, case-control status was determined by self-reported
448 depression (634,037 cases; 1,308,690 controls) or self-reported anxiety (624,615 cases;
449 1,310,854 controls) from samples of European ancestry (close relatives removed) in the
450 23andMe research cohort. The self-reported phenotype of depression was defined as cases
451 if samples have ever been diagnosed with depression, or controls if samples have never been
452 diagnosed with depression; the self-reported phenotype of anxiety was defined as cases if
453 samples have ever been diagnosed with anxiety, or controls if samples have never been
454 diagnosed with anxiety. All individuals included in the analyses provided informed consent

455 and participated in the research online, under a protocol approved by the external AAHRPP-
456 accredited IRB, Ethical & Independent Review Services (<http://www.eandireview.com>).

457 Association analyses were conducted by 23andMe; a logistic regression assuming an additive
458 model for allelic effects was used with adjustment for age, sex, indicator variables to
459 represent the genotyping platforms and the first five genotype principal components. The
460 summary statistics were provided for independent genome-wide significant SNPs in the
461 depression and anxiety latent factors GWAS. In the replication analysis of self-reported
462 depression, 8 SNPs were unavailable or of insufficient quality in the replication sample,
463 therefore 81 variants were tested for replication. In the replication analysis of self-reported
464 anxiety, 9 SNPs were of insufficient quality or unavailable in the replication samples, therefore
465 93 variants were tested for replication.

466 **Polygenic risk prediction**

467 The target sample consisted of an adult cohort (N = 4,434) from the Australian Twin Registry.
468 Depressive and anxiety symptoms were assessed by the Delusions-Symptoms-States
469 Inventory: Anxiety and Depression Scales (DSSI/sAD) (Bedford, Foulds, and Sheffield 1976),
470 which consists of seven anxiety and seven depression items. Each item assesses the degree
471 of distress due to a particular symptom, rated on a four-point ordinal scale: (0) none, (1) a
472 little, (2) a lot, (3) unbearably. Additional details of the cohort and assessment procedures are
473 reported elsewhere (Jardine, Martin, and Henderson 1984).

474 In total, nine polygenic risk scores (PRS) were created, using SNP weights from: DEP latent
475 factor (UKB), ANX latent factor (UKB), PHQ-9 sum-score (UKB; N = 135,149), GAD-7 sum-score
476 (UKB; N = 135,747), neuroticism (UKB - MHQ subset; N = 136,212), neuroticism (UKB; N =
477 338,812), major depression (Wray et al. 2018 - QIMR and 23andMe cohorts excluded; N =

478 159,598), depression (Howard et al. 2019 - QIMR and 23andMe cohorts excluded; N =
479 494,258), and anxiety disorders (Purves et al. 2019; N = 114,019). We used SBayesR (Lloyd-
480 Jones et al. 2019) to account for the correlation in effect sizes arising from linkage
481 disequilibrium (LD). Briefly, SBayesR implements Bayesian multiple regression to jointly
482 analyse all SNPs and account for LD between SNPs. As recommended, a shrunk matrix derived
483 from ~3 million SNPs (MAF > 0.01) on 50k participants from the UKB was used as the LD
484 correlation matrix. Polygenic risk scores were calculated in PLINK v1.90 (Purcell et al. 2007).
485 For each set of PRS we estimated the proportion of variance explained by the PRS in sum-
486 scores of both anxiety symptoms and depressive symptoms in the target sample, using
487 genomic restricted maximum likelihood (GCTA-GREML) in order to account for relatedness
488 within the target sample. We also estimated the amount of variance explained by PRS in log-
489 transformed and inverse normal-transformed sum-scores of depressive and anxiety
490 symptoms (presented in Supplementary Table 12).

491 **Gene-based tests and gene-set analysis**

492 MAGMA v1.07 (de Leeuw et al. 2015) was used to conduct gene-based and gene-set analyses
493 on the summary statistics of the DEP and ANX latent factors. The gene-based analysis tested
494 18,756 protein coding genes for association. A Bonferroni corrected significance threshold
495 was applied ($P < 2.67 \times 10^{-6}$). The gene-set analysis tested 7,250 gene sets for association with
496 DEP and ANX factors. A Bonferroni corrected significance threshold was applied
497 ($P < 6.90 \times 10^{-6}$).

498 **Pairwise analysis of GWAS summary statistics**

499 The pairwise GWAS analysis was implemented using gwas-pw (Pickrell et al. 2016). First, the
500 genome is split into 1703 approximately independent regions (Berisa and Pickrell 2016). Then

501 the posterior probability of each of the following models is calculated: (1) the region is
502 associated with DEP only, (2) the region is associated with ANX only, (3) the region is
503 associated with both DEP and ANX, and (4) there are separate associations for ANX and DEP
504 within that region. To account for sample overlap across the two traits, gwas-pw requires the
505 correlation between effect sizes in the two traits in non-associated regions. We used fgwas
506 (Pickrell et al. 2014) to calculate the posterior probability of association (PPA) for each region
507 with both traits. We then calculated the correlation in effect sizes for SNPs in regions with a
508 PPA < 0.2 for both ANX and DEP. Given one of the models has a posterior probability > 0.5,
509 we report the model with the highest posterior probability. Results are presented in the form
510 of an ideogram, created in the Complex-Traits Genetics Virtual Lab (Cuéllar-Partida et al.
511 2019).

512 **Gene-mapping of trait-specific or shared regions**

513 First, MAGMA was used to conduct gene-based tests separately on genomic regions reported
514 to be associated with only DEP (120 protein-coding genes tested; Bonferroni corrected
515 significance threshold, $P < 4.17 \times 10^{-4}$), regions associated with only ANX (390 protein-coding
516 genes tested; Bonferroni corrected significance threshold, $P < 1.57 \times 10^{-4}$), and regions
517 associated with both ANX and DEP (1038 protein-coding genes tested; Bonferroni corrected
518 significance threshold, $P < 4.82 \times 10^{-5}$).

519 Three additional methods implemented in FUMA were used to map SNPs in trait-specific or
520 shared regions (GWAS $P < 5 \times 10^{-8}$) to genes. (1) Positional mapping: SNPs are mapped to
521 genes based on proximity (within a 10kb window). (2) eQTL mapping: SNPs are mapped to a
522 gene if they have a significant (FDR < 0.05) association with the expression level of that gene.
523 We used eQTL information from GTEx v8 (Lonsdale et al. 2013), the CommonMind

524 Consortium (Fromer et al. 2016), and BRAINEAC (Ramasamy et al. 2014). (3) Chromatin
525 interaction mapping: genes are mapped if there is a significant ($FDR < 1 \times 10^{-6}$) chromatin
526 interaction between a genomic region (within a genomic risk locus) and promotor regions of
527 genes 250 bp upstream and 500 bp downstream of the transcription start site. Hi-C sequence
528 data was used to identify chromatin interactions from 23 human tissue and cell types (Schmitt
529 et al. 2016).

530 All prioritised genes for trait-specific and shared regions were used to conduct gene-set
531 enrichment analyses (hypergeometric test performed in FUMA) against gene-sets defined by
532 traits in the NHGRI-EBI GWAS Catalog (Buniello et al. 2019). Multiple testing was corrected
533 for with a Benjamini-Hochberg false discovery rate (FDR) of 0.05.

534

535 **Acknowledgements**

536 We thank the research participants of all cohorts for making this study possible. This work
537 was conducted using the UK Biobank Resource (application number 25331). NGM received
538 funding from the Australian National Health & Medical Research Council (NHMRC) to conduct
539 surveys in the Australian Twin Registry. SM is supported by an NHMRC Fellowship.

540 Members of the 23andMe Research Team include: Michelle Agee, Stella Aslibekyan, Adam
541 Auton, Robert K. Bell, Katarzyna Bryc, Sarah K. Clark, Sayantan Das, Sarah L. Elson, Kipper
542 Fletez-Brant, Pierre Fontanillas, Nicholas A. Furlotte, Pooja M. Gandhi, Karl Heilbron, Barry
543 Hicks, David A. Hinds, Karen E. Huber, Ethan M. Jewett, Yunxuan Jiang, Kira Kalkus, Aaron
544 Kleinman, Keng-Han Lin, Nadia K. Litterman, Marie K. Luff, Matthew H. McIntyre, Kimberly F.
545 McManus, Steven J. Micheletti, Sahar V. Mozaffari, Joanna L. Mountain, Priyanka
546 Nandakumar, Elizabeth S. Noblin, Carrie A.M. Northover, Jared O'Connell, Aaron A.
547 Petrakovitz, Steven J. Pitts, G. David Poznik, J. Fah Sathirapongsasuti, Janie F. Shelton, Chao
548 Tian, Joyce Y. Tung, Robert J. Tunney, Vladimir Vacic, and Xin Wang.

549 **Conflict of interest**

550 The authors have no conflicts of interest to declare.

References

- Achim, Amélie M., Michel Maziade, Éric Raymond, David Olivier, Chantal Mérette, and Marc-André Roy. 2009. 'How Prevalent Are Anxiety Disorders in Schizophrenia? A Meta-Analysis and Critical Review on a Significant Association', *Schizophrenia Bulletin*, 37: 811-21.
- Adams, Mark J., David M. Howard, Michelle Luciano, Toni-Kim Clarke, Gail Davies, W. David Hill, . . . Andrew M. McIntosh. 2019. 'Genetic stratification of depression by neuroticism: revisiting a diagnostic tradition', *Psychological Medicine*: 1-10.
- Akbaraly, Tasnime N., Mika Kivimäki, Eric J. Brunner, Tarani Chandola, Michael G. Marmot, Archana Singh-Manoux, and Jane E. Ferrie. 2009. 'Association Between Metabolic Syndrome and Depressive Symptoms in Middle-Aged Adults', *Diabetes Care*, 32: 499.
- Altamura, A. C., C. Montresor, D. Salvadori, and E. Mundo. 2004. 'Does comorbid subthreshold anxiety affect clinical presentation and treatment response in depression? A preliminary 12-month naturalistic study', *Int J Neuropsychopharmacol*, 7: 481-7.
- Baselmans, B. M. L., R. Jansen, H. F. Ip, J. van Dongen, A. Abdellaoui, M. P. van de Weijer, . . . M. Bartels. 2019. 'Multivariate genome-wide analyses of the well-being spectrum', *Nat Genet*, 51: 445-51.
- Beard, C., A. J. Millner, M. J. Forgeard, E. I. Fried, K. J. Hsu, M. T. Treadway, . . . T. Bjorgvinsson. 2016. 'Network analysis of depression and anxiety symptom relationships in a psychiatric sample', *Psychol Med*, 46: 3359-69.
- Bedford, A., G. A. Foulds, and B. F. Sheffield. 1976. 'A new personal disturbance scale (DSSI/sAD)', *Br J Soc Clin Psychol*, 15: 387-94.
- Berisa, T., and J. K. Pickrell. 2016. 'Approximately independent linkage disequilibrium blocks in human populations', *Bioinformatics*, 32: 283-5.
- Bulik-Sullivan, Brendan K, Hilary K. Finucane, Verner Anttila, Alexander Gusev, Felix R. Day, Po-Ru Loh, . . . Benjamin M. Neale. 2015. 'An atlas of genetic correlations across human diseases and traits', *Nature Genetics*, 47: 1236.
- Buniello, A., J. A. L. MacArthur, M. Cerezo, L. W. Harris, J. Hayhurst, C. Malangone, . . . H. Parkinson. 2019. 'The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019', *Nucleic Acids Res*, 47: D1005-d12.
- Bycroft, Clare, Colin Freeman, Desislava Petkova, Gavin Band, Lloyd T. Elliott, Kevin Sharp, . . . Jonathan Marchini. 2018. 'The UK Biobank resource with deep phenotyping and genomic data', *Nature*, 562: 203-09.
- Cai, Na, Joana A. Revez, Mark J. Adams, Till F. M. Andlauer, Gerome Breen, Enda M. Byrne, . . . M. D. D. Working Group of the Psychiatric Genomics Consortium. 2020. 'Minimal phenotyping yields genome-wide association signals of low specificity for major depression', *Nature Genetics*.
- Chang, Christopher C., Carson C. Chow, Laurent Cam Teller, Shashaank Vattikuti, Shaun M. Purcell, and James J. Lee. 2015. 'Second-generation PLINK: rising to the challenge of larger and richer datasets', *GigaScience*, 4: 7-7.
- Cuéllar-Partida, Gabriel, Mischa Lundberg, Pik Fang Kho, Shannon D'Urso, Luis F. Gutiérrez-Mondragón, Trung Thanh Ngo, and Liang-Dar Hwang. 2019. 'Complex-Traits Genetics Virtual Lab: A community-driven web platform for post-GWAS analyses', *bioRxiv*: 518027.
- Davis, Katrina A. S., Jonathan R. I. Coleman, Mark Adams, Naomi Allen, Gerome Breen, Breda Cullen, . . . Matthew Hotopf. 2020. 'Mental health in UK Biobank – development, implementation and results from an online questionnaire completed by 157 366 participants: a reanalysis', *BJPsych Open*, 6: e18.
- de Leeuw, Christiaan A., Joris M. Mooij, Tom Heskes, and Danielle Posthuma. 2015. 'MAGMA: Generalized Gene-Set Analysis of GWAS Data', *PLOS Computational Biology*, 11: e1004219.
- Emmanuel, J., S. Simmonds, and P. Tyrer. 1998. 'Systematic review of the outcome of anxiety and depressive disorders', *British Journal of Psychiatry*, 173: 35-41.

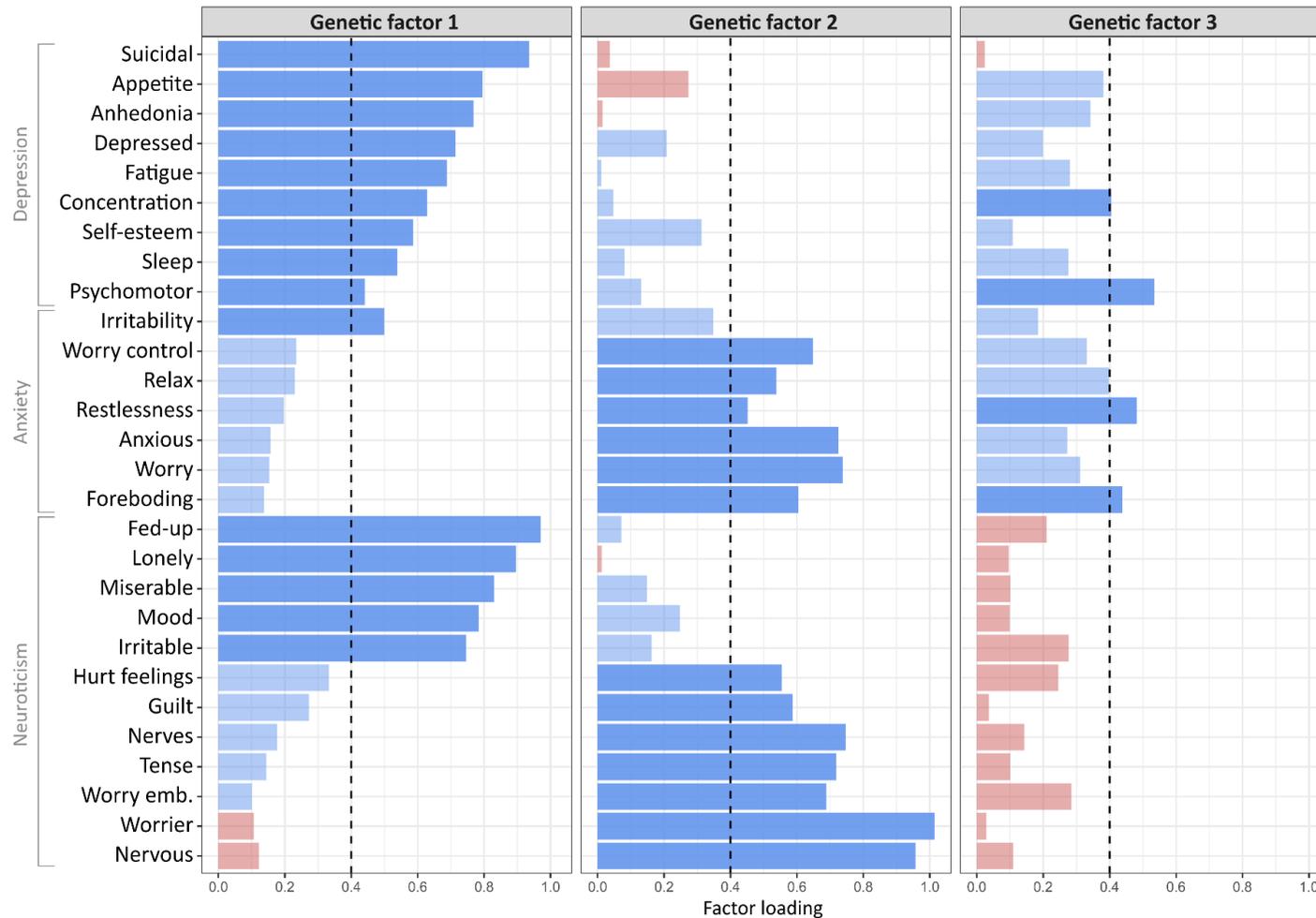
- Emsley, R. A., P. P. Oosthuizen, A. F. Joubert, M. C. Roberts, and D. J. Stein. 1999. 'Depressive and anxiety symptoms in patients with schizophrenia and schizophreniform disorder', *J Clin Psychiatry*, 60: 747-51.
- Eysenck, Hans J, and Michael W Eysenck. 1985. *Personality and individual differences: A natural science approach* (Plenum: New York, NY).
- Eysenck, S. B. G., H. J. Eysenck, and Paul Barrett. 1985. 'A revised version of the psychoticism scale', *Personality and Individual Differences*, 6: 21-29.
- Fanous, A., C. O Gardner, C. A Prescott, R. Cancro, and K. S Kendler. 2002. 'Neuroticism, major depression and gender: a population-based twin study', *Psychological Medicine*, 32: 719-28.
- Fromer, Menachem, Panos Roussos, Solveig K. Sieberts, Jessica S. Johnson, David H. Kavanagh, Thanneer M. Perumal, . . . Pamela Sklar. 2016. 'Gene expression elucidates functional impact of polygenic risk for schizophrenia', *Nature Neuroscience*, 19: 1442.
- Glueck, Charles J., Murray Tieger, Robert Kunkel, Trent Tracy, James Speirs, Patricia Streicher, and Ellen Illig. 1993. 'Improvement in symptoms of depression and in an index of life stressors accompany treatment of severe hypertriglyceridemia', *Biological Psychiatry*, 34: 240-52.
- Gray, Jeffrey A., and Neil McNaughton. 2000. *The Neuropsychology of Anxiety. An Enquiry into the Functions of the Septo-Hippocampal System* (Oxford University Press: Oxford).
- Grotzinger, Andrew D., Mijke Rhemtulla, Ronald de Vlaming, Stuart J. Ritchie, Travis T. Mallard, W. David Hill, . . . Elliot M. Tucker-Drob. 2019. 'Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits', *Nature Human Behaviour*, 3: 513-25.
- Hettema, J. M., M. C. Neale, and K. S. Kendler. 2001. 'A review and meta-analysis of the genetic epidemiology of anxiety disorders', *Am J Psychiatry*, 158: 1568-78.
- Hettema, J. M., M. C. Neale, J. M. Myers, C. A. Prescott, and K. S. Kendler. 2006. 'A population-based twin study of the relationship between neuroticism and internalizing disorders', *Am J Psychiatry*, 163: 857-64.
- Hettema, J. M., C. A. Prescott, and K. S. Kendler. 2004. 'Genetic and environmental sources of covariation between generalized anxiety disorder and neuroticism', *Am J Psychiatry*, 161: 1581-7.
- Howard, David M., Mark J. Adams, Toni-Kim Clarke, Jonathan D. Hafferty, Jude Gibson, Masoud Shirali, . . . Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. 2019. 'Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions', *Nature Neuroscience*.
- Howard, David M., Mark J. Adams, Masoud Shirali, Toni-Kim Clarke, Riccardo E. Marioni, Gail Davies, . . . Andrew M. McIntosh. 2018. 'Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways', *Nature Communications*, 9: 1470.
- Hyde, Craig L., Michael W. Nagle, Chao Tian, Xing Chen, Sara A. Paciga, Jens R. Wendland, . . . Ashley R. Winslow. 2016. 'Identification of 15 genetic loci associated with risk of major depression in individuals of European descent', *Nature Genetics*, 48: 1031.
- Igna, C. V., J. Julkunen, and H. Vanhanen. 2011. 'Vital exhaustion, depressive symptoms and serum triglyceride levels in high-risk middle-aged men', *Psychiatry Res*, 187: 363-9.
- Insel, Thomas, Bruce Cuthbert, Marjorie Garvey, Robert Heinssen, Daniel S Pine, Kevin Quinn, . . . Philip Wang. 2010. "Research domain criteria (RDoC): toward a new classification framework for research on mental disorders." In.: Am Psychiatric Assoc.
- Jardine, R., N. G. Martin, and A. S. Henderson. 1984. 'Genetic covariation between neuroticism and the symptoms of anxiety and depression', *Genetic Epidemiology*, 1: 89-107.
- Kaiser, Henry F. 1960. 'The application of electronic computers to factor analysis', *Educational and psychological measurement*, 20: 141-51.
- Kendler, K. S., A. C. Heath, N. G. Martin, and L. J. Eaves. 1987. 'Symptoms of anxiety and symptoms of depression. Same genes, different environments?', *Arch Gen Psychiatry*, 44: 451-7.

- Kendler, Kenneth S., Charles O. Gardner, Michael C. Neale, Steve Aggen, Andrew Heath, Lucía Colodro-Conde, . . . Nathan A. Gillespie. 2019. 'Shared and specific genetic risk factors for lifetime major depression, depressive symptoms and neuroticism in three population-based twin samples', *Psychological Medicine*, 49: 2745-53.
- Kessler, R. C, M. Gruber, J. M Hettema, I. Hwang, N. Sampson, and K. A Yonkers. 2007. 'Co-morbid major depression and generalized anxiety disorders in the National Comorbidity Survey follow-up', *Psychological Medicine*, 38: 365-74.
- Kotov, Roman, Wakiza Gamez, Frank Schmidt, and David Watson. 2010. 'Linking "big" personality traits to anxiety, depressive, and substance use disorders: a meta-analysis', *Psychological bulletin*, 136: 768.
- Kroenke, Kurt, Robert L. Spitzer, and Janet B. W. Williams. 2001. 'The PHQ-9', *Journal of General Internal Medicine*, 16: 606-13.
- Lamers, Femke, Patricia Oppen, Hannie Comijs, Johannes Smit, Philip Spinhoven, Anton Balkom, . . . B. W. Penninx. 2011. 'Comorbidity Patterns of Anxiety and Depressive Disorders in a Large Cohort Study: the Netherlands Study of Depression and Anxiety (NESDA)', *J Clin Psychiatry*, 72: 341-8.
- Lee, Phil H., Verner Anttila, Hyejung Won, Yen-Chen A. Feng, Jacob Rosenthal, Zhaozhong Zhu, . . . Jordan W. Smoller. 2019. 'Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders', *Cell*, 179: 1469-82.e11.
- Levey, Daniel F., Joel Gelernter, Renato Polimanti, Hang Zhou, Zhongshan Cheng, Mihaela Aslan, . . . Murray B. Stein. 2020. 'Reproducible Genetic Risk Loci for Anxiety: Results From ~200,000 Participants in the Million Veteran Program', *American Journal of Psychiatry*: appi.ajp.2019.19030256.
- Lloyd-Jones, Luke R., Jian Zeng, Julia Sidorenko, Loïc Yengo, Gerhard Moser, Kathryn E. Kemper, . . . Peter M. Visscher. 2019. 'Improved polygenic prediction by Bayesian multiple regression on summary statistics', *Nature Communications*, 10: 5086.
- Lonsdale, John, Jeffrey Thomas, Mike Salvatore, Rebecca Phillips, Edmund Lo, Saboor Shad, . . . Helen F. Moore. 2013. 'The Genotype-Tissue Expression (GTEx) project', *Nature Genetics*, 45: 580.
- Luciano, Michelle, Saskia P. Hagenaars, Gail Davies, W. David Hill, Toni-Kim Clarke, Masoud Shirali, . . . Ian J. Deary. 2018. 'Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism', *Nature Genetics*, 50: 6-11.
- MacGregor, S., J. S. Ong, J. An, X. Han, T. Zhou, O. M. Siggs, . . . A. W. Hewitt. 2018. 'Genome-wide association study of intraocular pressure uncovers new pathways to glaucoma', *Nat Genet*, 50: 1067-71.
- McCrae, Robert R., and Paul T. Costa. 1985. 'Updating Norman's "adequacy taxonomy": Intelligence and personality dimensions in natural language and in questionnaires', *Journal of Personality and Social Psychology*, 49: 710-21.
- Meier, Sandra M., Kalevi Tronteri, Kirstin L. Purves, Thomas Damm Als, Jakob Grove, Mikaela Laine, . . . Ole Mors. 2019. 'Genetic Variants Associated With Anxiety and Stress-Related Disorders: A Genome-Wide Association Study and Mouse-Model Study', *JAMA Psychiatry*, 76: 924-32.
- Middeldorp, C. M., D. C. Cath, R. Van Dyck, and D. I. Boomsma. 2005. 'The co-morbidity of anxiety and depression in the perspective of genetic epidemiology. A review of twin and family studies', *Psychol Med*, 35: 611-24.
- Nagel, Mats, Philip R. Jansen, Sven Stringer, Kyoko Watanabe, Christiaan A. de Leeuw, Julien Bryois, . . . Team andMe Research. 2018. 'Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways', *Nature Genetics*, 50: 920-27.
- Nagel, Mats, Kyoko Watanabe, Sven Stringer, Danielle Posthuma, and Sophie van der Sluis. 2018. 'Item-level analyses reveal genetic heterogeneity in neuroticism', *Nature Communications*, 9: 905.

- Okbay, Aysu, Bart M. L. Baselmans, Jan-Emmanuel De Neve, Patrick Turley, Michel G. Nivard, Mark Alan Fontana, . . . David Cesarini. 2016. 'Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses', *Nature Genetics*, 48: 624.
- Ormel, Johan, Bertus F. Jeronimus, Roman Kotov, Harriette Riese, Elisabeth H. Bos, Benjamin Hankin, . . . Albertine J. Oldehinkel. 2013. 'Neuroticism and common mental disorders: meaning and utility of a complex relationship', *Clinical psychology review*, 33: 686-97.
- Ormel, Johan, Harriette Riese, and Judith G. M. Rosmalen. 2012. 'Interpreting neuroticism scores across the adult life course: immutable or experience-dependent set points of negative affect?', *Clinical psychology review*, 32: 71-79.
- Pan, Yu, Wenpeng Cai, Qi Cheng, Wei Dong, Ting An, and Jin Yan. 2015. 'Association between anxiety and hypertension: a systematic review and meta-analysis of epidemiological studies', *Neuropsychiatric disease and treatment*, 11: 1121-30.
- Pickrell, Joseph K., Tomaz Berisa, Jimmy Z. Liu, Laure Séguirel, Joyce Y. Tung, and David A. Hinds. 2016. 'Detection and interpretation of shared genetic influences on 42 human traits', *Nature Genetics*, 48: 709-17.
- Purcell, Shaun, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A. R. Ferreira, David Bender, . . . Pak C. Sham. 2007. 'PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses', *The American Journal of Human Genetics*, 81: 559-75.
- Purves, Kirstin L., Jonathan R. I. Coleman, Sandra M. Meier, Christopher Rayner, Katrina A. S. Davis, Rosa Cheesman, . . . Thalia C. Eley. 2019. 'A major role for common genetic variation in anxiety disorders', *Molecular psychiatry*.
- Ramasamy, Adaikalavan, Daniah Trabzuni, Sebastian Guelfi, Vibin Varghese, Colin Smith, Robert Walker, . . . Michael E. Weale. 2014. 'Genetic variability in the regulation of gene expression in ten regions of the human brain', *Nature Neuroscience*, 17: 1418-28.
- Rayner, Christopher, Jonathan R. I. Coleman, Kirstin L. Purves, John Hodson, Kimberley Goldsmith, Georg W. Alpers, . . . Thalia C. Eley. 2019. 'A genome-wide association meta-analysis of prognostic outcomes following cognitive behavioural therapy in individuals with anxiety and depressive disorders', *Translational Psychiatry*, 9: 150.
- Richter, N., G. Juckel, and H. J. Assion. 2010. 'Metabolic syndrome: a follow-up study of acute depressive inpatients', *European Archives of Psychiatry and Clinical Neuroscience*, 260: 41-49.
- Sanchez-Roige, Sandra, and Abraham A. Palmer. 2020. 'Emerging phenotyping strategies will advance our understanding of psychiatric genetics', *Nature Neuroscience*.
- Schmitt, Anthony D, Ming Hu, Inkyung Jung, Zheng Xu, Yunjiang Qiu, Catherine L Tan, . . . Bing Ren. 2016. 'A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome', *Cell Reports*, 17: 2042-59.
- Schwabe, I., Y. Milaneschi, Z. Gerring, P. F. Sullivan, E. Schulte, N. P. Suppli, . . . C. M. Middeldorp. 2019. 'Unraveling the genetic architecture of major depressive disorder: merits and pitfalls of the approaches used in genome-wide association studies', *Psychological Medicine*, 49: 2646-56.
- Smoller, J. W., O. A. Andreassen, H. J. Edenberg, S. V. Faraone, S. J. Glatt, and K. S. Kendler. 2019. 'Psychiatric genetics and the structure of psychopathology', *Mol Psychiatry*, 24: 409-20.
- Spitzer, Robert L., Kurt Kroenke, Janet B. W. Williams, and Bernd Löwe. 2006. 'A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7', *Archives of Internal Medicine*, 166: 1092-97.
- Sullivan, P. F., M. C. Neale, and K. S. Kendler. 2000. 'Genetic epidemiology of major depression: review and meta-analysis', *Am J Psychiatry*, 157: 1552-62.
- Thorp, Jackson G., Andries T. Marees, Jue-Sheng Ong, Jiyuan An, Stuart MacGregor, and Eske M. Derks. 2019. 'Genetic heterogeneity in self-reported depressive symptoms identified through genetic analyses of the PHQ-9', *Psychological Medicine*: 1-12.

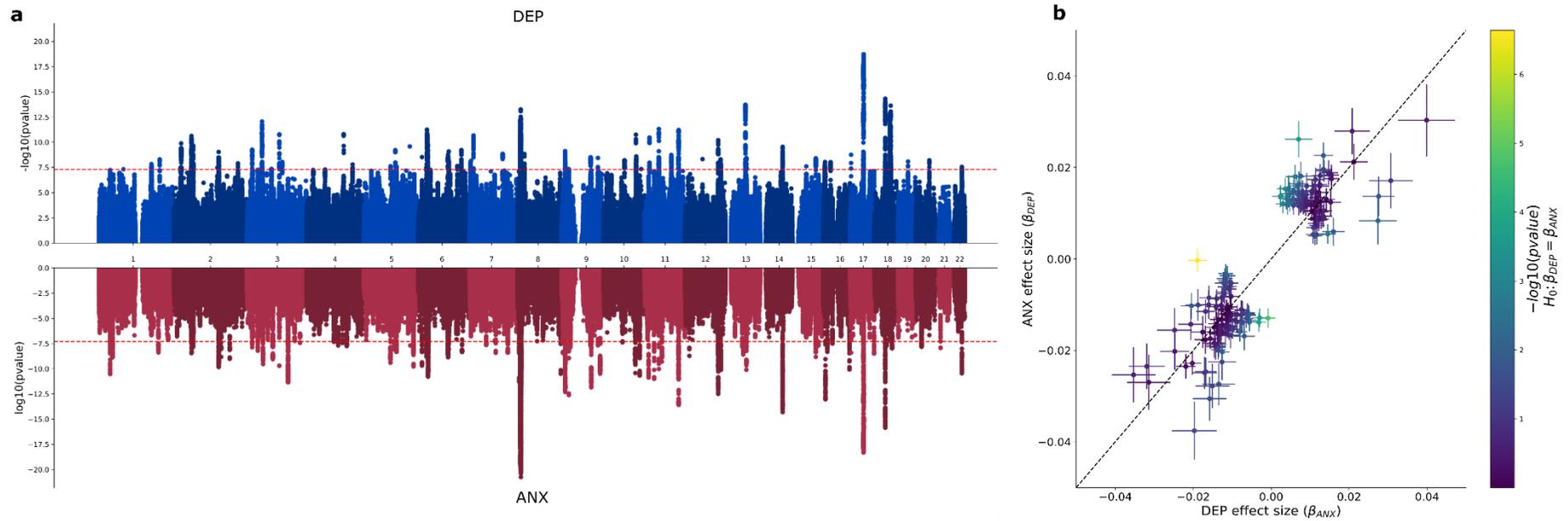
- Turley, Patrick, Raymond K. Walters, Omeed Maghizian, Aysu Okbay, James J. Lee, Mark Alan Fontana, . . . Consortium Social Science Genetic Association. 2018. 'Multi-trait analysis of genome-wide association summary statistics using MTAG', *Nature Genetics*, 50: 229-37.
- Velicer, Wayne F. 1976. 'Determining the number of components from the matrix of partial correlations', *Psychometrika*, 41: 321-27.
- Vigo, Daniel, Graham Thornicroft, and Rifat Atun. 2016. 'Estimating the true global burden of mental illness', *The Lancet Psychiatry*, 3: 171-78.
- Vukasović, Tena, and Denis Bratko. 2015. 'Heritability of personality: A meta-analysis of behavior genetic studies', *Psychological bulletin*, 141: 769-85.
- Walker, E. A., W. J. Katon, J. Russo, M. Von Korff, E. Lin, G. Simon, . . . J. Unützer. 2000. 'Predictors of outcome in a primary care depression trial', *Journal of General Internal Medicine*, 15: 859-67.
- Waszczuk, Monika A., Nicholas R. Eaton, Robert F. Krueger, Alexander J. Shackman, Irwin D. Waldman, David H. Zald, . . . Roman Kotov. 2020. 'Redefining phenotypes to advance psychiatric genetics: Implications from hierarchical taxonomy of psychopathology', *J Abnorm Psychol*, 129: 143-61.
- Watanabe, K., S. Stringer, O. Frei, M. Umicovic Mirkov, C. de Leeuw, T. J. C. Polderman, . . . D. Posthuma. 2019. 'A global overview of pleiotropy and genetic architecture in complex traits', *Nat Genet*, 51: 1339-48.
- Watanabe, Kyoko, Erdogan Taskesen, Arjen Bochoven, and Danielle Posthuma. 2017. 'Functional mapping and annotation of genetic associations with FUMA', *Nature Communications*, 8: 1826.
- World Health Organization. 2017. 'Depression and other common mental disorders: global health estimates'.
- Wray, Naomi R., Stephan Ripke, Manuel Mattheisen, Maciej Trzaskowski, Enda M. Byrne, Abdel Abdellaoui, . . . the Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. 2018. 'Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression', *Nature Genetics*, 50: 668-81.
- Young, J. F., L. Mufson, and M. Davies. 2006. 'Impact of comorbid anxiety in an effectiveness study of interpersonal psychotherapy for depressed adolescents', *J Am Acad Child Adolesc Psychiatry*, 45: 904-12.
- Zinbarg, Richard E., Susan Mineka, Lyuba Bobova, Michelle G. Craske, Suzanne Vrshek-Schallhorn, James W. Griffith, . . . Deepika Anand. 2016. 'Testing a Hierarchical Model of Neuroticism and Its Cognitive Facets: Latent Structure and Prospective Prediction of First Onsets of Anxiety and Unipolar Mood Disorders During 3 Years in Late Adolescence', *Clinical Psychological Science*, 4: 805-24.

Figure 1: Standardised factor loadings from a genetic exploratory factor analysis of 28 items of depression, anxiety and neuroticism.



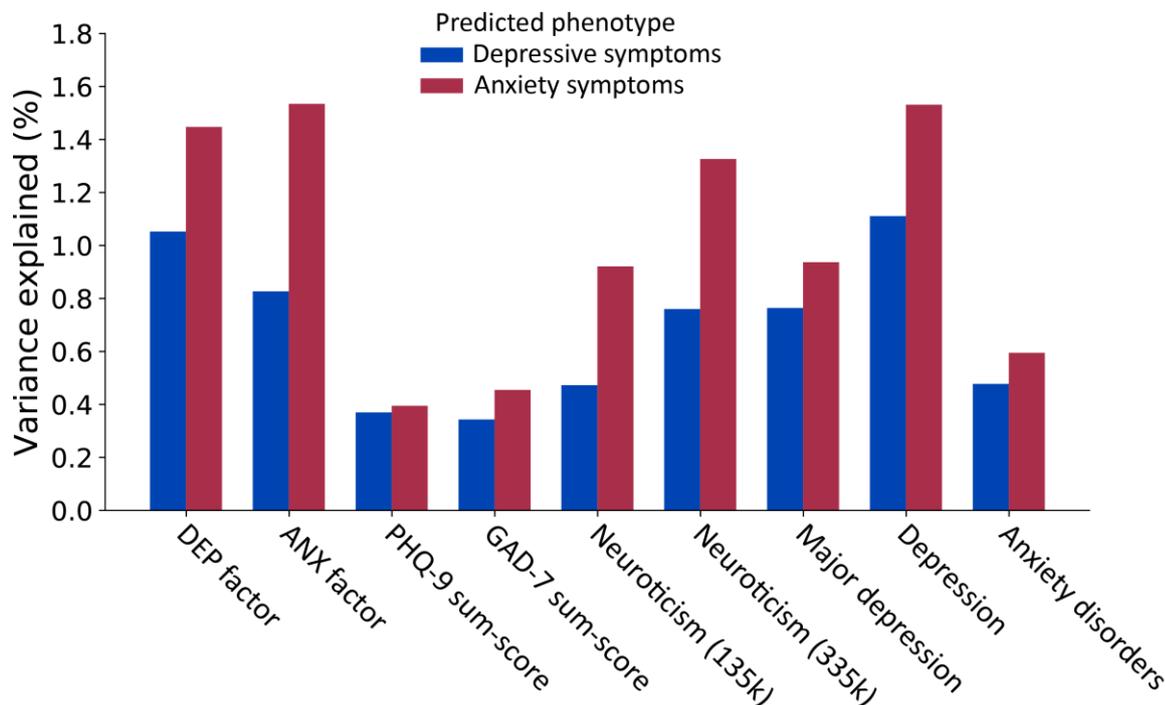
Note: Positive loadings are indicated in blue and negative loadings in red. Items with a standardised loading less than 0.4 are shown as transparent.

Figure 2: SNP-based associations of the DEP and ANX latent factors and comparison of effect sizes



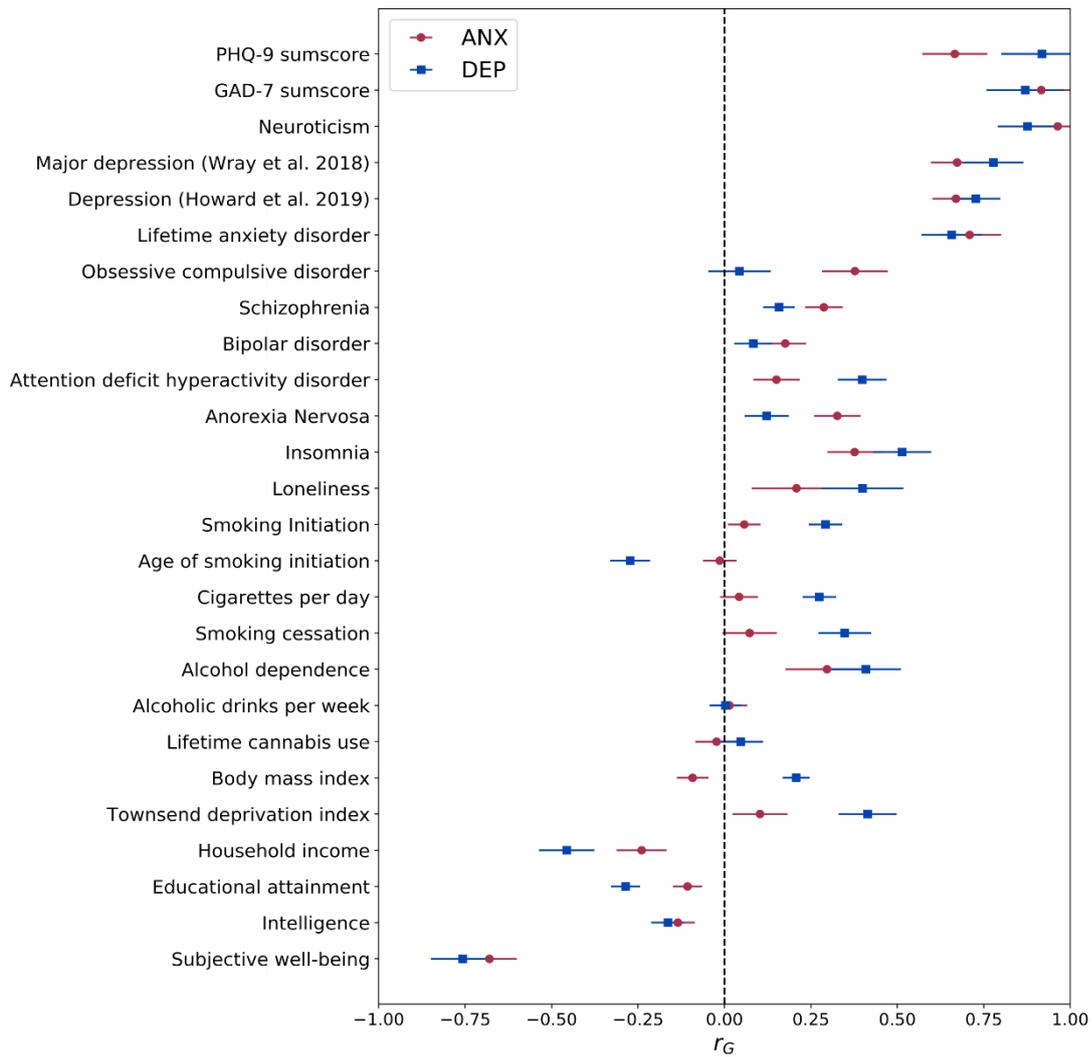
Note: (a) Manhattan plots for GWAS of the DEP (top; blue) and ANX (bottom; red) latent factors. The red line denotes the p-value threshold for genome-wide significance ($P < 5 \times 10^{-8}$). (b) Comparison of effect sizes for independent SNPs that reached genome-wide significance in the GWAS of either factor.

Figure 3: Polygenic risk prediction of depressive and anxiety symptoms



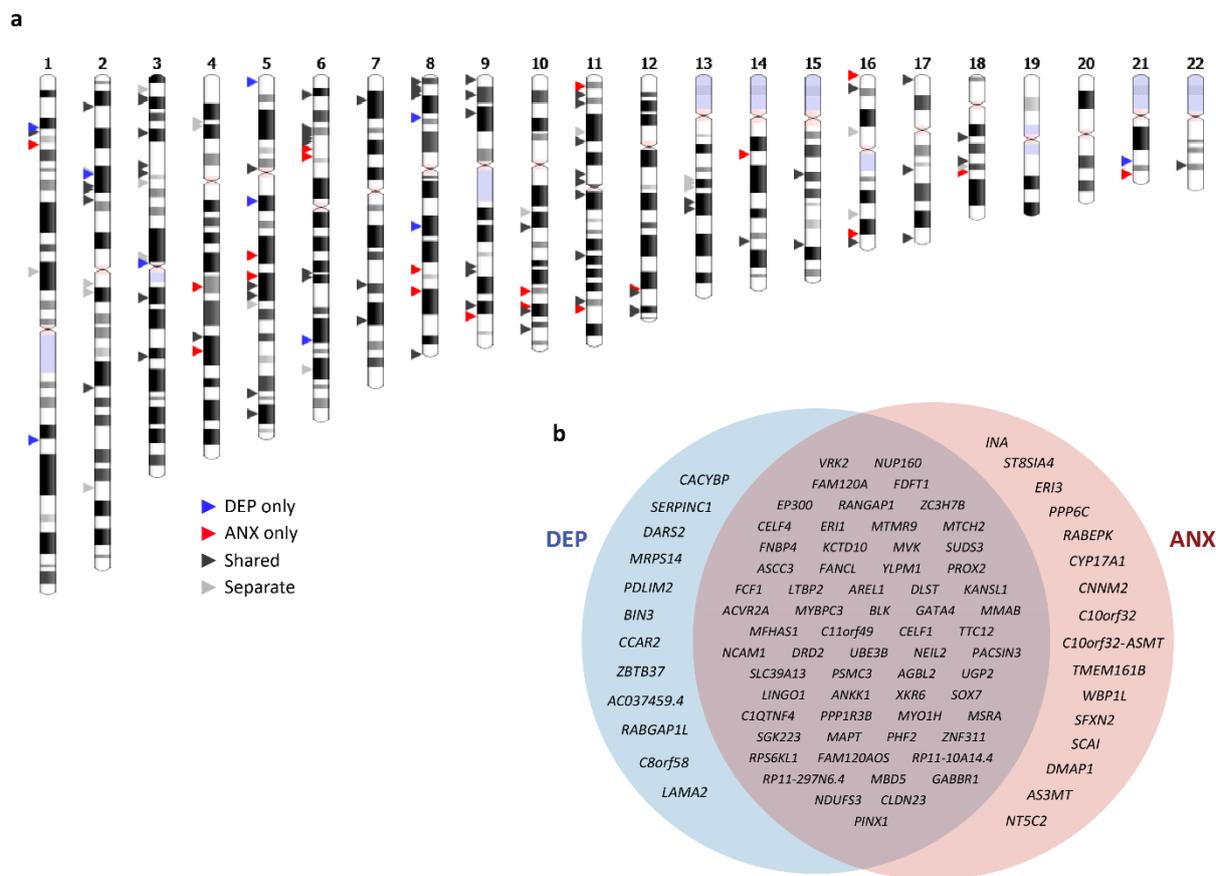
Note: amount of variance in depressive and anxiety symptoms explained by PRS for the DEP and ANX latent factors, and a range of related phenotypes: PHQ-9 sum-score (N = 135,149), GAD-7 sum-score (N = 135,747), neuroticism (N = 136,212; N = 338,812), major depression (Wray et al. 2018; QIMR and 23andMe cohorts excluded; N = 159,598), depression (Howard et al. 2019; QIMR and 23andMe cohorts excluded; N = 494,258), and anxiety disorders (N = 114,019). All PRS predictions were significant after Bonferroni correction ($\alpha = 0.05/18$; $P < 2.78 \times 10^{-3}$).

Figure 4: Genetic correlations with other complex traits



Note: Genetic correlations between the latent factors and 26 psychiatric, substance use and socioeconomic complex traits. Error bars represent 95% confidence intervals

Figure 5: Shared and trait-specific genetic influences on depressive and anxiety symptoms.



Note: (a) Genomic regions that are specific to DEP (blue arrows), specific to ANX (red arrows), shared (dark grey arrows), or influence both traits via separate variants (light grey arrows). (b) Prioritized genes for shared and trait-specific regions. Genes that reached significance (after Bonferroni correction) in the gene-based test and were identified by at least two additional methods (position, eQTL, or chromatin interaction) are presented (see Supplementary Tables 15-20 for all mapped genes).