

1 The stratification of major depressive disorder into genetic subgroups

2

3 David M. Howard*¹, Toni-Kim Clarke¹, Mark J. Adams¹, Jonathan D. Hafferty¹, Eleanor M.
4 Wigmore¹, Yann Zeng^{1,2}, Lynsey S. Hall¹, Jude Gibson¹, Thibaud S. Boutin², Caroline Hayward^{2,8},
5 Pippa A. Thomson^{3,4}, David J. Porteous³, Blair H. Smith^{5,8}, Alison D. Murray^{6,8}, Major Depressive
6 Disorder Working Group of the Psychiatric GWAS Consortium[†], Chris S. Haley², Ian J. Deary^{4,7,8},
7 Heather C. Whalley¹ and Andrew M. McIntosh^{1,7,8}

8

9 Affiliations:

10 ¹Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, UK

11 ²Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine,
12 University of Edinburgh, Edinburgh, UK

13 ³Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine,
14 University of Edinburgh, Edinburgh, UK

15 ⁴Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK

16 ⁵Division of Population Health Sciences, University of Dundee, Dundee, UK

17 ⁶Aberdeen Biomedical Imaging Centre, University of Aberdeen, Aberdeen, UK

18 ⁷Department of Psychology, The University of Edinburgh, Edinburgh, UK

19 ⁸Generation Scotland, Institute of Genetics and Molecular Medicine, University of Edinburgh,
20 Edinburgh, UK

21 [†]Consortium members and their affiliations are listed in supplementary information

22 *Corresponding author: David M. Howard

23 Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, UK

24 +44 131 537 6268

25 (e-mail: D.Howard@ed.ac.uk)

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40 **Major depressive disorder (MDD) is a heritable condition ($h^2 = 37\%$)¹ and a leading cause of**
41 **disability worldwide². MDD is clinically heterogeneous and comorbid with a variety of**
42 **conditions and it has been hypothesised that this causal heterogeneity may have confounded**
43 **previous attempts to elucidate its genetic architecture³⁻⁵. We applied a relatively new technique,**
44 **Buhmbox⁶, to identify the presence of heterogeneous sub-groups within MDD using summary**
45 **data from genome-wide association studies. We analysed two independent cohorts ($n_{\text{total}} =$**
46 **31,981) and identified significant evidence ($P_{\text{corrected}} < 0.05$) for 10 sub-groups across both**
47 **cohorts, including subgroups with a liability for migraine, alcohol consumption and eczema.**
48 **The most notable subgroups ($P_{\text{corrected}} \leq 2.57 \times 10^{-8}$ in both cohorts) were for blood levels of**
49 **cholesterol and triglycerides, and blood pressure, indicating subgroups within MDD cases of**
50 **individuals with a genetic predisposition for anomalous levels of these metabolic traits. Our**
51 **findings provide strong evidence for novel causal heterogeneity of MDD and identify avenues**
52 **for both stratification and treatment.**

53 MDD is a complex and clinically heterogeneous condition that is characterised by symptoms
54 including low mood and/or anhedonia persisting for at least two weeks. Many unique combinations of
55 symptoms may lead to the same diagnosis and it has been suggested that this symptomatic
56 heterogeneity may be due to, as yet unproven, causal heterogeneity⁷. In support of the causal
57 heterogeneity hypothesis, MDD is frequently observed to be comorbid with many diseases including
58 cancer⁸, cardiovascular disease,⁹ and other psychiatric illnesses^{10,11}.

59 We sought to test the presence of causal heterogeneity in MDD according to a number of disease and
60 quantitative traits using a newly available tool, Buhmbox⁶. Buhmbox examines the weighted pairwise
61 correlations of the risk allele dosages for these diseases and traits within MDD cases and controls,
62 based on effect size and frequency, and assigns a *P*-value based on the likelihood of the observed
63 correlations between the cases and controls. We used two cohort studies, Generation Scotland:
64 Scottish Family Health Study (GS:SFHS)¹² and UK Biobank¹³, both of which have whole-genome
65 genotyping data and information relating to MDD status. Study demographics for each cohort are
66 provided in **Table 1**. Within each cohort, we examined 34 traits with a reported comorbidity with

67 MDD and tested whether evidence of subgroups for these traits could be detected within our MDD
68 cases. Further information regarding the 34 traits and their sources is provided in **Supplementary**
69 **Table 1**. For the traits anorexia nervosa, neuroticism and MDD, summary statistics from different
70 publications were assessed and are numbered accordingly, i.e. MDD 1, MDD 2 and MDD 3. In the
71 case of MDD, this allowed us to examine whether different sets of associated loci, drawn from
72 different populations and diagnostic criteria for MDD, would form a heterogeneous subgroup within
73 our GS:SFHS and/or UK Biobank MDD cases.

74 **Table 1.** Study demographics of Generation Scotland: Scottish Family Health Study (GS:SFHS) and
75 UK Biobank

	GS:SFHS	UK Biobank
N	6,946	25,035
Age range	19 - 93	40 - 79
Mean age in years (S.D)	51.5 (13.2)	57.8 (8.0)
Males / Females	3,013 / 3,933	12,528 / 12,507
MDD Cases / Controls	975 / 5,971	8,508 / 16,527

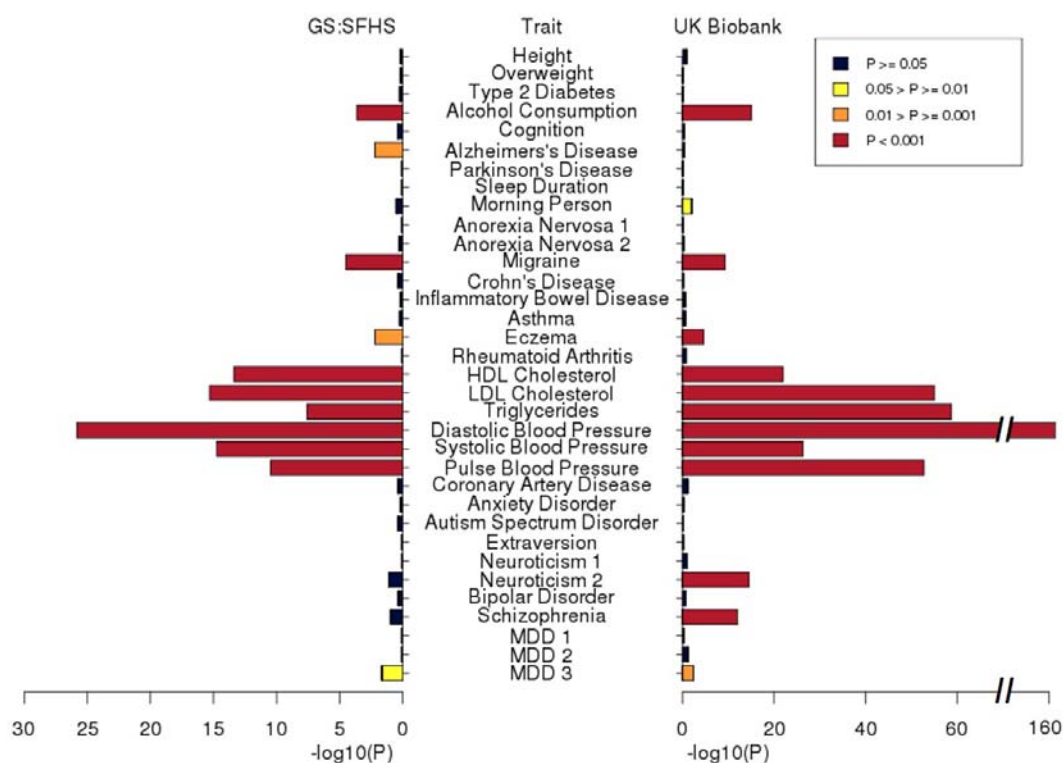
76
77 To account for multiple testing, *P*-values were adjusted using a false discovery rate adjustment and all
78 reported values have been adjusted¹⁴. Ten traits showed significant evidence ($P < 0.05$) of MDD
79 subgroup heterogeneity across both cohorts: total HDL cholesterol levels, total LDL cholesterol
80 levels, serum triglycerides levels, diastolic blood pressure, systolic blood pressure, pulse pressure,
81 alcohol consumption, migraine, eczema and MDD 3. Four traits: Alzheimer's disease, neuroticism,
82 schizophrenia and being a 'morning person' were identified as stratifying traits within one, but not
83 both, cohorts. The subgroup heterogeneity *P*-values obtained within each cohort and for each trait are
84 shown in **Table 2** and **Figure 1**.

85 **Table 2.** The false discovery rate adjusted *P*-values for subgroup heterogeneity of the shown disease
86 or quantitative trait within MDD for Generation Scotland: Scottish Family Health Study (GS:SFHS)
87 and UK Biobank. Bold values indicate statistical significance ($P < 0.05$).

	GS:SFHS	UK Biobank
Height	0.657	0.141
Overweight	0.693	0.876
Type 2 Diabetes	0.587	0.854
Alcohol Consumption	2.29 x 10⁻⁴	9.37 x 10⁻¹⁶
Cognition	0.387	0.361
Alzheimer's Disease	0.006	0.361

Parkinson's Disease	0.851	0.854
Sleep Duration	0.806	0.876
Morning Person	0.300	0.010
Anorexia Nervosa 1	0.851	0.781
Anorexia Nervosa 2	0.545	0.588
Migraine	3.08 x 10⁻⁵	5.30 x 10⁻¹⁰
Crohn's Disease	0.443	0.606
Inflammatory Bowel Disease	0.651	0.280
Asthma	0.587	0.264
Eczema	6.56 x 10⁻³	2.49 x 10⁻⁵
Rheumatoid Arthritis	0.918	0.174
HDL Cholesterol	4.09 x 10⁻¹⁴	1.14 x 10⁻²²
LDL Cholesterol	4.91 x 10⁻¹⁶	1.04 x 10⁻⁵⁵
Triglycerides	2.57 x 10⁻⁸	1.94 x 10⁻⁵⁹
Diastolic Blood Pressure	1.48 x 10⁻²⁶	3.84 x 10⁻¹⁶²
Systolic Blood Pressure	1.84 x 10⁻¹⁵	5.07 x 10⁻²⁷
Pulse Blood Pressure	3.36 x 10⁻¹¹	1.90 x 10⁻⁵³
Coronary Artery Disease	0.443	0.086
Anxiety Disorder	0.693	0.602
Autism Spectrum Disorder	0.443	0.648
Extraversion	0.784	0.606
Neuroticism 1	0.918	0.102
Neuroticism 2	0.081	3.20 x 10⁻¹⁵
Bipolar Disorder	0.453	0.189
Schizophrenia	0.117	1.13 x 10⁻¹²
MDD 1	0.851	0.588
MDD 2	0.851	0.073
MDD 3	0.023	0.004

88



89

90 **Figure 1.** The false discovery rate adjusted P -values for evidence of subgroup heterogeneity of the
 91 shown disease or quantitative trait within MDD for both Generation Scotland: Scottish Family Health
 92 Study (GS:SFHS; $n = 6,946$) and UK Biobank ($n = 25,035$).

93

94 The most striking results were those relating to metabolic traits ($P \leq 2.57 \times 10^{-8}$ across both cohorts).

95 Although these traits are unlikely to be independent of one another, the statistical significance

96 obtained across both studies suggest robust subgroup heterogeneity. Buhmbox does not identify the

97 individuals within the subgroup and future work to address this would aid in determining the degree

98 of overlap between the observed P -values. Milaneschi, et al.¹⁵ also reported evidence of a genetic

99 correlation using profile risk scores between triglycerides and a severe atypical MDD subgroup and it

100 would be beneficial to examine the blood pressure and cholesterol traits in other populations, social,

101 economic and health care settings. Although elevated blood pressure and cholesterol levels are

102 positively correlated with coronary artery disease (CAD), no evidence ($P \geq 0.05$) for a subgroup for

103 CAD was found. This apparent anomaly merits further study, but may reflect stratification

104 independent of CAD.

1105 There is substantial comorbidity between alcohol abuse and MDD¹⁶ with studies demonstrating a
1106 bidirectional relationship between alcohol and depression^{17,18}. Both cohorts used in our study provided
1107 evidence ($P < 0.05$) of a novel alcohol consumption subgroup within MDD. Ellingson, et al.¹⁹ have
1108 suggested that negative emotionality and behavioral control may mediate the genetic overlap between
1109 alcohol consumption and MDD. Further work could examine whether the subgroups within our MDD
1110 cases could also possess loci that influence those mediatory factors.

1111 Similarly to alcohol consumption, migraine has also been shown to have a bidirectional relationship
1112 with depression²⁰. We found evidence ($P < 0.05$) of subgroup heterogeneity within MDD that had a
1113 genetic predisposition for migraine in both of the cohorts that we studied. Ours is the first study to
1114 report the existence of a heterogeneous subgroup within MDD cases of individuals with a migraine-
1115 like genetic profile.

1116 Subgroup heterogeneity for MDD was observed ($P < 0.05$) in both GS:SFHS and UK Biobank for the
1117 MDD 3 trait obtained from Hyde, et al.²¹. MDD 3 was based on a self-reported diagnosis of MDD
1118 within a large and predominately European population. The diagnosis of MDD within UK Biobank
1119 was also self-reported and the existence of an MDD 3 subgroup suggests that there is a common
1120 shared genetic basis to this phenotype, but that it was not shared across all cases. MDD 1 was
1121 extracted from a study examining recurrent depression in Han Chinese women within a hospital
1122 setting and although we didn't find evidence of a subgroup ($P \geq 0.05$), this was not completely
1123 unexpected within our population and UK-based cohorts. No evidence ($P \geq 0.05$) was found for a
1124 MDD 2 subgroup, however a polygenic risk score approach has provided evidence of pleiotropy
1125 between MDD 2 cases and GS:SFHS cases ($P < 1.37 \times 10^{-10}$) and MDD 2 cases and UK Biobank
1126 cases ($P < 1.92 \times 10^{-8}$), Hall, et al., (manuscript in preparation).

1127 The body's inflammatory response has been highlighted as a potential contributor to depression²².
1128 Crohn's disease, inflammatory bowel disease and asthma were examined, but neither cohort provided
1129 evidence ($P \geq 0.05$) for subgroup heterogeneity. Asthma is frequently comorbid with eczema and both
1130 cohorts examined in our study provided evidence ($P < 0.05$) for a subgroup of individuals with a
1131 genetic predisposition for eczema within our MDD cases. Associations between eczema and

132 depression are well reported in the literature and potentially mediated by health anxiety²³. Chronic
133 inflammation seen in some cases of eczema, and a growing appreciation for the impotence of
134 inflammation in depression, provides another possible explanation for subgroup heterogeneity.

135 A degree of cognitive impairment has been reported in individuals that are currently experiencing a
136 depressive episode^{24,25}. However, we found no evidence ($P \geq 0.05$) for MDD subgroup heterogeneity
137 for general fluid cognitive ability. Alzheimer's disease is also associated with a decline in cognitive
138 ability and previous studies have demonstrated depression to be a risk factor for the disease^{26,27}.
139 Within GS:SFHS, there was evidence ($P = 0.006$) for a subgroup of MDD cases which harboured the
140 loci associated with Alzheimer's disease, however this was not replicated in UK Biobank ($P \geq 0.05$).
141 Parkinson's disease is another condition that is associated with neuropathology that is more likely to
142 occur later in life. We found no evidence ($P \geq 0.05$) in either cohort for a subgroup of Parkinson's
143 disease within MDD cases.

144 Subgroup heterogeneity was observed ($P = 1.13 \times 10^{-12}$) in UK Biobank for schizophrenia which
145 substantiates the work of Milaneschi, et al. ⁷ who demonstrated correlated genomic profile risk scores
146 between schizophrenia and a severe typical MDD subtype. However, Han, et al. ⁶ and our GS:SFHS
147 cohort provided no evidence ($P > 0.05$) of a schizophrenia subgroup within MDD cases, which
148 suggests that evidence of subgroup heterogeneity for schizophrenia is population and/or diagnosis
149 dependent.

150 We also examined a number of developmental and personality traits due to the impact that depression
151 can have on social interaction and feelings of self-worth. Evidence of subgroup heterogeneity ($P =$
152 3.20×10^{-15}) was only found within UK Biobank for neuroticism 2 drawn from the Smith, et al. ²⁸
153 study. The neuroticism 2 trait had a much greater number of associated loci compared to neuroticism
154 1 and therefore neuroticism 1 may have been underpowered to detect an effect, but this is also
155 dependent on the effect sizes of the associated loci, the number of MDD cases and the size of any
156 subgroup. There are similarities in the way that individuals respond to stressful events between
157 neuroticism and MDD and it may be the heritable component that underpins this response that is
158 driving the observed subgrouping within UK Biobank.

159 The diagnosis of MDD within GS:SFHS was based on the DSM-IV criteria²⁹, which includes
160 questions related to sleep and eating patterns. Therefore, sleep duration, being a ‘morning person’ and
161 anorexia nervosa were included in our study. However, it was only within UK Biobank that a
162 significant *P*-value for being a ‘morning person’ ($P = 0.010$) was observed. Anorexia nervosa 1 and 2
163 and sleep duration had low numbers of associated loci available (**Supplementary Table 1**) and were
164 potentially underpowered to detect an effect.

165 Rheumatoid Arthritis (RA) is an autoimmune disease which, like many other chronic diseases, has
166 been shown in multiple studies to be comorbid with depression^{30,31}, with a potential subgroup of
167 depressed individuals within RA sufferers³². However, no evidence ($P \geq 0.05$) of subgroup
168 heterogeneity was found for RA within either GS:SFHS or UK Biobank MDD cases. Multiple studies
169 have suggested that morphological traits^{33,34} and type 2 diabetes³⁵ may identify subgroups of
170 individuals with MDD, but we found no evidence ($P \geq 0.05$) for subgroup heterogeneity according to
171 these traits in the current study.

172 The two cohorts used in this study reflect a subsample of the UK population, with additional steps
173 taken to ensure that there were no overlapping individuals. An MDD diagnosis was made using a structured
174 clinical interview within GS:SFHS, whereas UK Biobank cases were defined by a number of self-
175 reported measures. A broad range of traits were assessed and the selection of the summary statistics
176 used was based on the number of individuals analysed, the availability of summary statistics and
177 publication date. Buhmbox measures the correlations within cases, which is independent (or
178 orthogonal) information from the effect size (personal communication with Buhm Han). Therefore,
179 UK Biobank was able to be used to obtain the summary statistics for neuroticism 2, alcohol
180 consumption and the blood pressure traits, and then also used to assess the existence of MDD
181 subgroup heterogeneity.

182 Multiple studies have suggested the presence of aetiology subgroups within MDD and Buhmbox
183 provides a quantifiable measure of their existence. Our study has provided replicable evidence of
184 novel subgroup heterogeneity within MDD for a range of disease and quantitative traits, including
185 blood pressure, cholesterol and triglyceride levels, migraine, eczema and alcohol consumption. This

186 research underlines the potential of using genomic data for developing stratified approaches to the
187 diagnosis and treatment of depression.

188 COMPETING FINANCIAL INTERESTS

189 The authors declare that no competing financial interests exist.

190 ACKNOWLEDGEMENTS

191 Please refer to the supplementary information for full acknowledgments.

192 AUTHOR CONTRIBUTIONS

193 AMMcI, DJP, BHS, ADM, IJD and CH were involved in the acquisition of the GS:SFHS cohort.
194 Quality control of the GS:SFHS data was conducted by LSH, JDH, MJA, CH and DHM. Imputation
195 of the GS:SFHS data was conducted by TB and CH. Quality control of the UK Biobank data was
196 conducted by MJA and DHM. AMMcI and DMH conceived the initial design of the study with HCW,
197 JDH, YZ, T-KC, MJA, EMW, JG, PAT, CSH, IJD and DJP involved in the ongoing development of
198 the project. DMH conducted the analysis and wrote the paper and all authors have read and approved
199 its submission.

200

201

202

203

204

205

206

207

208

209

210

211 Methods

212 Generation Scotland: Scottish Family Health Study (GS:SFHS)

213 The family and population-based Generation Scotland: Scottish Family Health Study (GS:SFHS)
214 cohort¹² consisted of 23,960 individuals, of whom 20,195 were genotyped with the Illumina
215 OmniExpress BeadChip (706,786 SNPs). The genotypic data was uploaded to the Michigan
216 Imputation Server³⁶ and phased using SHAPE IT v2.r837³⁷ and imputed using the Haplotype
217 Reference Consortium reference panel (HRC.r1-1)³⁸. The imputation of GS:SFHS has been published
218 previously³⁹. We applied an imputation accuracy threshold (infoscore) of ≥ 0.8 and this provided us
219 with a total of 8,633,288 genome-wide variants calls for 20,032 individuals.

220 A diagnosis of MDD was made using two initial screening questions and the Structured Clinical
221 Interview for the Diagnostic and Statistical Manual of Mental Disorders (SCID)²⁹. The diagnosis of
222 MDD within GS:SFHS has been described previously⁴⁰ and in our study, MDD was defined by at
223 least one instance of a major depressive episode. Further to this, we used record linkage to the
224 Scottish Morbidity Record⁴¹ to examine the psychiatric history of both case and control individuals.
225 We identified 1,072 control individuals who had attended at least one psychiatry outpatient clinic and
226 we excluded these individuals from our study. Using the psychiatric inpatient records, we identified
227 47 MDD cases who were also diagnosed with bipolar disorder or schizophrenia and these individuals
228 were also excluded from our study. These participants provided us with prior consent for their
229 anonymised data to be linked to clinical data. As GS:SFHS was a family-based cohort, we created an
230 unrelated subsample using GCTA v.122⁴² ensuring that no two individuals shared a genomic
231 relatedness of ≥ 0.025 . A further 186 individuals who were identified as population outliers through
232 principal component analyses of their genotypic information⁴³. This left a total of 975 MDD cases and
233 5,971 controls (14.0% prevalence) in the GS:SFHS cohort.

234 UK Biobank

235 The population-based UK Biobank¹³ (provided as part of project #4844) consisted of 152,249
236 individuals with genomic data for 72,355,667 imputed variants⁴⁴. This was the standard data release
237 available to all approved researchers of UK Biobank. Detailed information regarding the imputation
238 procedure⁴⁵ and initial quality control⁴⁶ are provided elsewhere. In summary, phasing was achieved
239 using a modified version of SHAPE IT 2⁴⁷ with a combined reference panel of 1,000 genomes phase 3
240 and the UK10K haplotype reference⁴⁸ panels and the IMPUTE2 package⁴⁹ used for imputation. We
241 applied an infoscore threshold of ≥ 0.8 which left a total of 24,467,210 variants. We removed
242 individuals listed as non-white British and those individuals that had also participated in GS:SFHS
243 identified using a checksum approach⁵⁰ using genotype data.

244 Of the remaining participants, 25,035 had completed a touchscreen assessment of depressive
245 symptoms and previous treatment. We used the diagnostic definitions of Smith, et al.⁵¹ and defined
246 case status as either ‘probable single lifetime episode of major depression’ or ‘probable recurrent
247 major depression (moderate and severe)’ and with control status defined as ‘no mood disorder’. This
248 provided us with a total of 8,508 cases and 16,527 controls (34.0% prevalence) within UK Biobank,
249 which is greater than that observed within GS:SFHS.

250 Statistical Approach

251 Buhmbox v0.33⁶ was used to conduct the statistical analysis and this package requires raw genetic
252 and phenotypic data (disease A) and also summary statistics relating to the additional disease and
253 quantitative traits for testing (disease B). The disease B associated loci were drawn from either
254 published material or from personal communications and are detailed in **Supplementary Table 1**.
255 The pruning of disease B associated loci was conducted using Plink 1.90⁵² and the --indep-pairwise
256 command. A 50 variant window with a 5 variant sliding window was applied to the summary statistics
257 and pruned any variants with an $r^2 > 0.1$.

258 For GS:SFHS the first 20 principal components were derived from the genotypic data using GCTA
259 v1.22⁴² and these were fitted within Buhmbox to account for population stratification. For UK
260 Biobank the first 15 genetic principal components⁵³ were fitted. Buhmbox examines whether there is a

261 sharing of risk alleles between the disease B associated loci and the disease A cases (in our case
262 MDD). Buhmbox uses the positive correlations between risk allele dosages in disease A cases to
263 determine whether any sharing of risk alleles is driven by all individuals (pleiotropy) or by a subset of
264 individuals (heterogeneity). The likelihood of observing such positive correlations are used to
265 determine the reported *P*-values. The Buhmbox software and manual is freely downloadable from
266 <http://software.broadinstitute.org/mpg/buhmbox/>. The data that support the findings of this study are
267 available on reasonable request from the corresponding author, DMH. The data are not publicly
268 available due to participant confidentiality and the terms of the existing mutual transfer agreements
269 with the respective data repositories.

270

271

- 272 1. Sullivan, P.F., Neale, M.C. & Kendler, K.S. Genetic epidemiology of major depression: review
273 and meta-analysis. *American Journal of Psychiatry* **157**, 1552-1562 (2000).
- 274 2. World Health Organization. Mental health action plan 2013 - 2020. (2013).
- 275 3. Flint, J. & Kendler, Kenneth S. The genetics of major depression. *Neuron* **81**, 484-503 (2014).
- 276 4. Levinson, D.F. *et al.* Genetic studies of major depressive disorder: why are there no genome-
277 wide association study findings and what can we do about it? *Biological Psychiatry* **76**, 510-
278 512 (2014).
- 279 5. Lubke, G.H. *et al.* Genome-wide analyses of borderline personality features. *Molecular*
280 *Psychiatry* **19**, 923-929 (2014).
- 281 6. Han, B. *et al.* A method to decipher pleiotropy by detecting underlying heterogeneity driven
282 by hidden subgroups applied to autoimmune and neuropsychiatric diseases. *Nature Genetics*
283 **48**, 803-810 (2016).
- 284 7. Milaneschi, Y. *et al.* Polygenic dissection of major depression clinical heterogeneity.
285 *Molecular Psychiatry* **21**, 516-522 (2016).
- 286 8. Walker, J. *et al.* Prevalence, associations, and adequacy of treatment of major depression in
287 patients with cancer: a cross-sectional analysis of routinely collected clinical data. *The Lancet*
288 *Psychiatry* **1**, 343-350 (2014).
- 289 9. Hare, D.L., Toukhsati, S.R., Johansson, P. & Jaarsma, T. Depression and cardiovascular
290 disease: a clinical review. *European Heart Journal*, 1365-1372 (2013).
- 291 10. Häfner, H., Löffler, W., Maurer, K., Hambrecht, M. & an der Heiden, W. Depression, negative
292 symptoms, social stagnation and social decline in the early course of schizophrenia. *Acta*
293 *Psychiatrica Scandinavica* **100**, 105-118 (1999).
- 294 11. Avenevoli, S., Swendsen, J., He, J.-P., Burstein, M. & Merikangas, K. Major depression in the
295 national comorbidity survey- adolescent supplement: prevalence, correlates, and treatment.
296 *Journal of the American Academy of Child and Adolescent Psychiatry* **54**, 37-44. e2 (2015).
- 297 12. Smith, B.H. *et al.* Cohort profile: Generation Scotland: Scottish Family Health Study
298 (GS:SFHS). The study, its participants and their potential for genetic research on health and
299 illness. *International Journal of Epidemiology* **42**, 689-700 (2013).

- 300 13. Allen, N.E., Sudlow, C., Peakman, T. & Collins, R. UK biobank data: come and get it. *Science*
301 *Translational Medicine* **6**, 224ed4 (2014).
- 302 14. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful
303 approach to multiple testing. *Journal of the Royal Statistical Society. Series B*
304 *(Methodological)* **57**, 289-300 (1995).
- 305 15. Milaneschi, Y. *et al.* Polygenic dissection of major depression clinical heterogeneity. *Mol*
306 *Psychiatry* **21**, 516-522 (2016).
- 307 16. Grant, B.F., Stinson, F.S., Dawson, D.A. & *et al.* Prevalence and co-occurrence of substance
308 use disorders and independent mood and anxiety disorders: Results from the national
309 epidemiologic survey on alcohol and related conditions. *Archives of General Psychiatry* **61**,
310 807-816 (2004).
- 311 17. Brière, F.N., Rohde, P., Seeley, J.R., Klein, D. & Lewinsohn, P.M. Comorbidity between major
312 depression and alcohol use disorder from adolescence to adulthood. *Comprehensive*
313 *Psychiatry* **55**, 526-533 (2014).
- 314 18. Flensburg-Madsen, T. *et al.* Comorbidity and temporal ordering of alcohol use disorders and
315 other psychiatric disorders: results from a Danish register-based study. *Comprehensive*
316 *Psychiatry* **50**, 307-314 (2009).
- 317 19. Ellingson, J.M., Richmond-Rakerd, L.S., Statham, D.J., Martin, N.G. & Slutske, W.S. Most of
318 the genetic covariation between major depressive and alcohol use disorders is explained by
319 trait measures of negative emotionality and behavioral control. *Psychological Medicine* **46**,
320 2919-2930 (2016).
- 321 20. Breslau, N., Lipton, R.B., Stewart, W.F., Schultz, L.R. & Welch, K.M.A. Comorbidity of
322 migraine and depression: Investigating potential etiology and prognosis. *Neurology* **60**, 1308-
323 1312 (2003).
- 324 21. Hyde, C.L. *et al.* Identification of 15 genetic loci associated with risk of major depression in
325 individuals of European descent. *Nature Genetics* **48**, 1031-1036 (2016).
- 326 22. Miller, A.H. & Raison, C.L. The role of inflammation in depression: from evolutionary
327 imperative to modern treatment target. *Nat Rev Immunol* **16**, 22-34 (2016).
- 328 23. Klock, M., Gotestam, K.G. & Mykletun, A. Factors accounting for the association between
329 anxiety and depression, and eczema: the Hordaland health study (HUSK). *BMC Dermatology*
330 **10**, 3 (2010).
- 331 24. Millan, M.J. *et al.* Cognitive dysfunction in psychiatric disorders: characteristics, causes and
332 the quest for improved therapy. *Nat Rev Drug Discov* **11**, 141-168 (2012).
- 333 25. Dere, E., Pause, B.M. & Pietrowsky, R. Emotion and episodic memory in neuropsychiatric
334 disorders. *Behavioural Brain Research* **215**, 162-171 (2010).
- 335 26. Diniz, B.S., Butters, M.A., Albert, S.M., Dew, M.A. & Reynolds, C.F. Late-life depression and
336 risk of vascular dementia and Alzheimer's disease: systematic review and meta-analysis of
337 community-based cohort studies. *The British Journal of Psychiatry* **202**, 329-335 (2013).
- 338 27. Norton, S., Matthews, F.E., Barnes, D.E., Yaffe, K. & Brayne, C. Potential for primary
339 prevention of Alzheimer's disease: an analysis of population-based data. *The Lancet*
340 *Neurology* **13**, 788-794 (2014).
- 341 28. Smith, D.J. *et al.* Genome-wide analysis of over 106 000 individuals identifies 9 neuroticism-
342 associated loci. *Molecular Psychiatry* **21**, 749-757 (2016).
- 343 29. First, M.B., Spitzer, R.L., Gibbon Miriam. & Williams, J.B.W. Structured Clinical Interview for
344 DSM-IV-TR Axis I Disorders, Research Version, Patient Edition. (SCID-I/P) (2002).
- 345 30. Matcham, F., Rayner, L., Steer, S. & Hotopf, M. The prevalence of depression in rheumatoid
346 arthritis: a systematic review and meta-analysis. *Rheumatology* **52**, 2136-2148 (2013).
- 347 31. Rathbun, A.M., Harrold, L.R. & Reed, G.W. Temporal associations between the different
348 domains of rheumatoid arthritis disease activity and the onset of patient-reported
349 depressive symptoms. *Clinical Rheumatology* **34**, 653-663 (2015).

- 350 32. Tillmann, T., Krishnadas, R., Cavanagh, J. & Petrides, K. Possible rheumatoid arthritis
351 subtypes in terms of rheumatoid factor, depression, diagnostic delay and emotional
352 expression: an exploratory case-control study. *Arthritis Research & Therapy* **15**, R45 (2013).
- 353 33. Deaton, A. & Arora, R. Life at the top: The benefits of height. *Economics & Human Biology* **7**,
354 133-136 (2009).
- 355 34. Carey, M. *et al.* Prevalence of comorbid depression and obesity in general practice: a cross-
356 sectional survey. *British Journal of General Practice* **64**, e122-e127 (2014).
- 357 35. Rustad, J.K., Musselman, D.L. & Nemeroff, C.B. The relationship of depression and diabetes:
358 Pathophysiological and treatment implications. *Psychoneuroendocrinology* **36**, 1276-1286
359 (2011).
- 360 36. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nature Genetics* **48**,
361 1284-1287 (2016).
- 362 37. Delaneau, O., Zagury, J.-F. & Marchini, J. Improved whole-chromosome phasing for disease
363 and population genetic studies. *Nature Methods* **10**, 5-6 (2013).
- 364 38. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nature*
365 *Genetics* **48**, 1279-1283 (2016).
- 366 39. Nagy, R. *et al.* Exploration of haplotype research consortium imputation for genome-wide
367 association studies in 20,032 Generation Scotland participants. *Genome Medicine* **9**, 23
368 (2017).
- 369 40. Fernandez-Pujals, A.M. *et al.* Epidemiology and heritability of major depressive disorder,
370 stratified by age of onset, sex, and illness course in generation scotland: scottish family
371 health study (GS:SFHS). *PLoS ONE* **10**, e0142197 (2015).
- 372 41. Information Services Division. SMR Data Manual. ([http://www.ndc.scot.nhs.uk/Data-](http://www.ndc.scot.nhs.uk/Data-Dictionary/SMR-Datasets)
373 [Dictionary/SMR-Datasets](http://www.ndc.scot.nhs.uk/Data-Dictionary/SMR-Datasets), 2016).
- 374 42. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: A tool for genome-wide complex
375 trait analysis. *The American Journal of Human Genetics* **88**, 76-82 (2011).
- 376 43. Amador, C. *et al.* Recent genomic heritage in Scotland. *BMC Genomics* **16**, 1-17 (2015).
- 377 44. Marchini, J. UK Biobank phasing and imputation documentation. Version 1.2.
378 (http://biobank.ctsu.ox.ac.uk/crystal/docs/impute_ukb_v1.pdf, 2015).
- 379 45. UK Biobank. Genotype imputation and genetic association studies using UK Biobank data.
380 <http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=157020> (2015).
- 381 46. UK Biobank. Genotyping and quality control for UK Biobank.
382 <http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=155580> (2015).
- 383 47. O'Connell, J. *et al.* Haplotype estimation for biobank-scale data sets. *Nature Genetics* **48**,
384 817-820 (2016).
- 385 48. Huang, J. *et al.* Improved imputation of low-frequency and rare variants using the UK10K
386 haplotype reference panel. *Nature Communications* **6**, 8111 (2015).
- 387 49. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes.
388 *G3: Genes/Genomes/Genetics* **1**, 457-470 (2011).
- 389 50. Ripke, S. GWAS genotypic overlap test without sharing genotypes.
390 https://personal.broadinstitute.org/sripke/share_links/checksums_download/readme_0415
391 [b.txt](https://personal.broadinstitute.org/sripke/share_links/checksums_download/readme_0415) (2014).
- 392 51. Smith, D.J. *et al.* Prevalence and characteristics of probable major depression and bipolar
393 disorder within UK Biobank: cross-sectional study of 172,751 participants. *PLoS ONE* **8**,
394 e75362 (2013).
- 395 52. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer
396 datasets. *GigaScience* **4**, 7 (2015).
- 397 53. UK Biobank. Genetic principal components.
398 <http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=22009> (2015).

399