Genome-wide interaction study of a proxy for stress-sensitivity and its prediction of major depressive disorder

Running title: **Stress-sensitivity proxy and depression**

Aleix Arnau-Soler MSc[1], Mark J. Adams PhD[2], Generation Scotland[3], Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium[4], Caroline Hayward PhD[5], Pippa A. Thomson PhD[1,6]

[1] Medical Genetics Section, University of Edinburgh, Centre for Genomic and Experimental Medicine and MRC Institute of Genetics and Molecular Medicine, Edinburgh, UK

[2] Division of Psychiatry, Deanery of Clinical Sciences, University of Edinburgh, Royal Edinburgh Hospital, Morningside Park, Edinburgh EH10 5HF, UK

[3] A collaboration between the University Medical School and NHS in Aberdeen, Dundee, Edinburgh and Glasgow, Scotland, UK

[4] For a full list of MDD working group of the PGC investigators, see the Supplemental Information

[5] Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

[6] Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK

Correspondence should be addressed to: P. Thomson (Pippa.Thomson@ed.ac.uk)

Centre for Genomic and Experimental Medicine
Institute of Genetics and Molecular Medicine
University of Edinburgh
Western General Hospital
Crewe Road South
Edinburgh, EH4 2XU, UK
Phone: +44(0)1316518756

ABSTRACT

Individual response to stress is correlated with neuroticism and is an important predictor of both neuroticism and the onset of major depressive disorder (MDD). Identification of the genetics underpinning in individual differences in response to negative events (stress-sensitivity) may improve our understanding of the molecular pathways involved, and its association with stress-related illnesses. We sought to generate a proxy for stress-sensitivity through modelling the interaction between SNP allele and MDD status on neuroticism score in order to identify genetic variants that contribute to the higher neuroticism seen in individuals with a lifetime diagnosis of depression compared to unaffected individuals. Meta-analysis of genome-wide interaction studies (GWIS) in UK Biobank (N = 23 092; UKB) and Generation Scotland: Scottish Family Health Study (N = 7 155; GS:SFHS) identified no genome-wide significance SNP interaction. However, gene-based tests identified a genome-wide significant gene, *ZNF366*, a negative regulator of glucocorticoid receptor function implicated in alcohol dependence ($p = 1.48 \times 10^{-7}$; Bonferroni-corrected significance threshold $p < 2.79 \times 10^{-6}$). Using summary statistics from the stress-sensitivity term of the GWIS, SNP heritability for stress-sensitivity was estimated at 5.0%. In models fitting polygenic risk scores of both MDD and neuroticism derived from independent GWAS, we show that polygenic risk scores derived from the stress-sensitivity GWIS significantly improved the prediction of MDD. This study may improve interpretation of larger genome-wide association studies of MDD and other stress-related illnesses, and the understanding of the etiological mechanisms underpinning stress-sensitivity.

53 **INTRODUCTION**

54 Stressful life events are known to increase liability to mental illness and disease-related
55 traits[1] including neuroticism[2-4], major depressive disorder (MDD)[5-7], autoimmune diseases[8]
56 and some cancers[9,10]. A greater understanding of the causal mechanism by which negative
57 events affect disease risk or outcome may be beneficial in identifying individuals for targeted
58 support. However, it has been proposed that sensitivity to stress may be an important
59 predictor of response to stress[11,12]. In particular, the effect on an individual may result more
60 from the perceived stress than the event itself, and be dependent on individual differences in
61 stress-sensitivity[13-18]. Studies of 5-HTT and twin studies suggest that stress-sensitivity may, at
62 least in part, be heritable[19-21]. Despite a complex interaction between MDD, neuroticism and
63 stress, multivariate structural equation models have confirmed a genetic effect on perceived
64 stress, overlapping that on MDD or neuroticism, but with a specific genetic component[20]. The
65 inter-relatedness of these traits may offer an approach to identify the genetic variation that
66 affects an individual's stress-sensitivity, and improve genetic prediction of an individual's
67 liability to negative outcomes. By modelling the interaction between SNP allele and MDD
68 status on neuroticism score through genome-wide interaction studies (GWIS), we sought to
69 investigate the genetics of stress-sensitivity.

70 The personality trait neuroticism is highly heritable (30–50% estimates from twin
71 studies)[22-25], is higher in individuals with depression compared to controls[26,27] and is
72 known to have shared genetic aetiology with depression[28-31]. Neuroticism is strongly
73 correlated with measures of sensitivity to punishment but not reward[32], positively
74 correlated with perceived personal relevance of a stressor[33,34] and has been used
75 previously as a proxy measure of stress-sensitivity[35]. Neuroticism is thought to mediate
76 or interact with the effects of adverse life events on risk of depression[5,36]. It has a

77 substantial stable component[37], however, there is evidence for change, as well as
78 stability, across the life span[2-4,38]. Whereas the stable component of neuroticism is
79 strongly determined by genetics, change in neuroticism score is attributed to the effects
80 of unshared environment[38]. Persistent change in neuroticism score has been shown in
81 response to life events[2-4].
82 Major depressive disorder (MDD) is a complex disorder influenced by both genetic
83 contributions and environmental risk factors, with heritability estimates from twin and
84 family studies of between 31-42%[39,40]. Confirmed environmental risk factors for MDD
85 include maternal infections, childhood maltreatment and negative life events[5-7,41,42].
86 However, few genetic studies have such information and even fewer prospective studies
87 exist. Incorporation of stressful life events has been shown to improve the ability to
88 predict MDD[43,44] and, although stress is an environmental risk factor, it may have an
89 independent genetic contribution to risk of depression[44-48].
90 These studies suggest that a genetic variable derived from the difference in neuroticism
91 levels seen in individuals with MDD compared to controls may allow us to identify
92 genetic loci important for stress-sensitivity. Further, polygenic risk scores (PRS) derived
93 from this stress-sensitivity variable may improve prediction of MDD over that based on
94 MDD or neuroticism PRS alone.
95 Using unrelated individuals from two large population-based samples, UK Biobank (UKB;
96 N = 23 092) and Generation Scotland: Scottish Family Health Study (GS:SFHS; N = 7 155),
97 we sought to identify genes involved in stress-sensitivity by performing GWIS for the
98 interaction between MDD status and SNP allele on neuroticism score. We identified a
99 gene significantly associated with stress-sensitivity and show that a PRS derived from
100 the interaction term of the GWIS, significantly predicts liability to depression
101 independently of the PRS for MDD and/or neuroticism.

102 **MATERIALS AND METHODS**

103 **Cohorts' description**

104 Study ethics are detailed in Supplemental Information: Ethics statements.

105 ***UK Biobank (UKB).*** UKB is a major national health resource that aims to improve the
106 prevention, diagnosis and treatment of a wide range of illnesses. It recruited more than
107 500 000 participants aged from middle to older age who visited 22 assessment centres
108 across the UK between 2006 and 2010. Data were collected on background and lifestyle,
109 cognitive and physical assessments, sociodemographic factors and medical history. The
110 scientific rationale, study design, ethical approval, survey methods, and limitations are
111 reported elsewhere[49,50]. The present study was conducted on genome-wide genotyping
112 data available from the first release of UKB data. Details of sample processing specific to
113 UKB project are available at http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=155583
114 and the Axiom array at http://media.affymetrix.com/support/downloads/
115 manuals/axiom_2_assay_auto_workflow_user_guide.pdf. UKB genotyping and the
116 stringent QC protocol applied to UKB data before it was released can be found at
117 http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=155580. Imputed UKB genotype data[51]
118 (imputed by UKB using a merged panel of the UK10K haplotype reference panel and the
119 1000 Genomes Phase 3 reference panel) with quality > 0.9 was hard-called using PLINK
120 v1.9[52] and SNPs genotyped on GS:SFHS extracted. Individuals were removed based on
121 UKB genomic analysis exclusion (UKB Data Dictionary item #22010), non-white British
122 ancestry (#22006: genetic ethnic grouping; from those individuals who self-identified as
123 British, PCs analysis was used to remove outliers), high genotype missingness (#22005),
124 genetic relatedness (#22012; no pair of individuals have a KING-estimated kinship
125 coefficient > 0.0442), QC failure in UK BiLEVE study (#22050 and #22051: UK BiLEVE
126 Affymetrix and UK BiLEVE genotype quality controls for samples) and gender mismatch

127 (#22001: genetic sex). Further, individuals in UKB who were also participants of
128 GS:SFHS, and their relatives, were excluded to remove any overlap of individuals
129 between discovery and target samples. A dataset of 109 283 individuals with 557 813
130 imputed SNPs remained for further analysis, aged 40-79 (57 328 female, 51 954 male;
131 mean age = 57.1 years, s.d. = 7.99), of which 109 282 had data available for neuroticism
132 score and 23 092 had data available on MDD status (n_{cases} = 7 834, $n_{controls}$ = 15 258, n_{female}
133 = 11 510, n_{male} = 11 582; mean age = 57.7 years, s.d. = 8.04). Thus, the final dataset
134 comprised 23 092 unrelated individuals.
135
136 ***Generation Scotland Scottish Family Health Study (GS:SFHS).*** GS:SFHS is a family-
137 based genetic epidemiology study which includes 23 960 participants from ~ 7 000
138 Scottish family groups collected by a cross-disciplinary collaboration of Scottish medical
139 schools and the National Health Service (NHS) from Feb 2006 to Mar 2011. Participants
140 were interviewed and clinically assessed for a wide range of health-related traits
141 (including high-fidelity phenotyping for Major Depressive Disorder and related
142 endophenotypes), environmental covariates and linked to routine health records[53,54].
143 The protocol for recruitment is described in detail in previous publications[55,56]. GS:SFHS
144 genotyping and quality control is detailed elsewhere[57]. Individuals with more than 2%
145 missing genotypes and gender discrepancies were removed, as well as population
146 outliers. SNPs with genotype missingness > 2%, minor allele frequency < 1% and a
147 Hardy-Weinberg Equilibrium test $p < 1x10^{-6}$ were exclude. Finally, individuals were
148 removed based on relatedness (pi-hat < 0.05), maximizing retention of case individuals,
149 using PLINK v1.9[52]. Genome-wide SNP data comprised 7 233 unrelated individuals
150 genotyped for 560,698 SNPs (n_{female} = 3 476, n_{male} = 3 757; PLINK v1.9[52]), aged 18-92
151 (mean age = 50.4 years, s.d. = 12.06) of which: 7 190 had clinical data on MDD; 7 196

152 individuals had data on neuroticism; and 7 155 had data on both neuroticism and MDD,
153 remained for further analysis.
154
155 **Phenotype assessment**
156 ***Neuroticism score (EPQN).*** Participants in both UKB and GS:SFHS cohorts were assessed
157 for neuroticism using 12 questions from the Eysenck Personality Questionnaire-Revised
158 Short Form's Neuroticism Scale (EPQN)[58-61]. Neuroticism can be scored by adding up the
159 number of "Yes" responses on EPQN. This short scale has a reliability of more than 0.8[62].
160 EPQN distributions were found to be sufficiently "normal" after assessment for skewness
161 and kurtosis to be analysed using linear regression (both coefficients were between -1
162 and 1).
163
164 ***MDD status.*** In UKB depressive symptoms were assessed by items relating to the
165 lifetime experience of minor and major depression[58], items from the Patient Health
166 Questionnaire[63] and items on help-seeking for mental health[61]. Controls were defined as
167 those who a) answered "No" to the question "Ever depressed for a whole week" or b)
168 who reported symptoms but duration was too short (<2 weeks) or who did not report
169 having seen a GP or psychiatrist for nerves, anxiety, tension or depression. In our
170 unrelated sample: 7 834 participants were diagnosed with MDD (with single, moderate
171 or recurrent episodes) and 15 258 were controls (N = 23 092).
172 In GS:SFHS, participants who answered yes to either of two screening questions were
173 assessed for psychiatric, mood state/psychological distress, personality and cognitive
174 assessment by trained researchers using the Structured Clinical Interview for DSM-IV
175 Non-Patient Version (SCID)[64]. In our unrelated sample, of those individuals with data

176 available for both MDD and EPQN (N = 7 155), used for the GWIS, 2 010 had a lifetime
177 diagnosis of MDD and 5 145 were controls.
178
179 **GWIS and derivation of a genetic stress-sensitivity effect**
180 The effect size of an stress-sensitivity effect (β_{SS}) was derived by performing a GWIS for
181 the effect of the MDD status and SNP allele on EPQN (dependent variable) in both UKB
182 and GS:SFHS cohorts using PLINK 1.90 (PLINK-command --gxe; fitting MDD diagnosis as
183 a binary "group" effect)[52]. PLINK-command --gxe estimates the difference in allelic
184 association with a quantitative trait (EPQN) between two groups (MDD cases vs.
185 controls) producing effect estimates on each group and a test of significance for the
186 interaction between SNP allele and MDD status. The interaction p value reflects the
187 difference between the regression coefficient of the allelic effect in a linear model for
188 EPQN in MDD cases (β_A) and the same regression coefficient in a linear model for EPQN
189 in controls (β_B). The stress-sensitivity interaction effect was defined as the difference in
190 allele effect between MDD cases and control groups.
191 Considering one SNP, the effect it confers to EPQN can be modelled by MDD status
192 (control = 0, MDD case = 1) as follows:

$$\begin{cases} MDD = 0;\ \ EPQN = \beta_0 + \beta_B SNP + \beta_{0c} COV + \varepsilon \\ MDD = 1;\ EPQN = \beta_1 + \beta_A SNP + \beta_{1c} COV + \varepsilon \end{cases}$$

193 This is equivalent to modelling the effect on MDD cases as follows:

$$\begin{cases} MDD = 0;\ \ EPQN = \beta_0 + \beta_B SNP + \beta_{0c} COV + \varepsilon \\ MDD = 1;\ EPQN = \beta_1 + \beta_B SNP + (\beta_A - \beta_B) SNP + \beta_{1c} COV + \varepsilon \end{cases}$$

194 Or, it can be modelled as a whole as:

$$EPQN = \beta_0 + \beta_2 MDD + \beta_B SNP + (\beta_A - \beta_B) SNP * MDD + \beta_{0c} COV + \beta_{2C} COV * MDD + \varepsilon$$

195 Where COV stands for covariates, β_2 stands for $\beta_1 - \beta_0$, and β_{2c} stands for $\beta_{1c} - \beta_{0c}$.

196 Thus, the interaction effect (β_{SS}) can be estimated as the difference in allelic effect on
197 EPQN between MDD cases (β_A) and controls (β_B) as follows,

$$\hat{\beta}_{SS} = \hat{\beta}_A - \hat{\beta}_B$$

198 $\hat{\beta}_{SS}$ is therefore defined as the effect size reflecting the genetic stress-sensitivity effect on
199 MDD cases compared to controls (see Supplemental Figure S1).
200

201 **Stress-sensitivity GWIS, main additive effect GWASs, meta-analysis and gene-set**
202 **analysis.**

203 For GWIS and subsequent analyses, sample specific covariates were applied as follows:
204 *UKB.* All phenotypes were adjusted for centre, array and batch as random effects prior to
205 analyses. Analyses were adjusted for age, sex and 15 informative principal components
206 (PCs; UKB Data Dictionary items #22009.01 to #22009.15) as fixed effects to take
207 account of possible population stratification. *GS:SFHS.* All the analyses were adjusted for
208 age, sex and 20 PCs.
209 GWAS for MDD and neuroticism, using logistic and linear models of additive allelic
210 effects respectively, were conducted on the same sample sets for comparison and
211 generation of matched PRS using PRSice[65].
212 Results from the GWIS of UKB and GS:SFHS were combined in a sample size weighted
213 meta-analysis performed using METAL[66]. While the use of standard error weighting is
214 more common, the different diagnostic schema and MDD prevalence between the two
215 cohorts (GS:SFHS; 12.2%, UKB: 25.8%)[55,61] may indicate systematic differences in the
216 measurement of MDD. Generalized gene-based analysis of the meta-analysis was
217 performed using MAGMA[67] implemented through FUMA[68] (http://fuma.ctglab.nl).
218 Briefly, SNP summary statistics were mapped to 17 931 protein-coding genes. Individual
219 SNP *p* values from a gene were combined into a gene test-statistic using a SNP-wise

220 model and a known approximation of the sampling distribution used to obtain a gene-
221 based *p* value. Genome-wide significance was defined at $p = 0.05/17\,931 = 2.79\text{x}10^{-6}$.
222
223 **LD Score regression**
224 The summary statistics from the meta-analysis were used to examine the genetic overlap
225 between the polygenic architecture of stress-sensitivity, MDD and neuroticism. LD score
226 regression was used to derive the genetic correlations (r_G) between these traits[69,70] using
227 meta-analysed GWAS and GWIS summary statistics. SNP-based heritability was also
228 estimated using LD score regression, which only requires summary statistics from
229 single-SNP analyses.
230
231 **Pathway, functional and gene expression analyses**
232 Lead SNPs, independently associated with the phenotype, were identified using PLINK
233 1.90 by clumping (*p* threshold < $2\text{x}10^{-5}$; LD r2 > 0.1; physical kb threshold = 500kb; 1000
234 Genomes Project Phase 1 CEU, GBR, TSI genotype data), and analysed using DEPICT[71].
235 Further detail is given in the Supplemental Information: DEPICT analyses.
236 Genes associated with lead SNPs were investigated for evidence of: phenotypic
237 association in the NCBI dbGaP database of genotypes and phenotypes[72]
238 (https://www.ncbi.nlm.nih.gov/gap/phegeni), regulatory DNA elements in normal cell
239 lines and association with expression quantitative trait loci (eQTLs) using the
240 RegulomeDB database[73] (http://www.regulomedb.org) and the Genotype-Tissue
241 Expression (GTEx) Portal[74] (http://www.gtexportal.org).
242
243 **Polygenic profiling**

244 PRS were produced using PRSice[65] and standardized to a mean of 0 and a standard
245 deviation of 1. PRS for stress-sensitivity (PRS_{SS}) were created by weighting the sum of
246 the reference alleles in an individual by the stress-sensitivity effect, β_{SS} using GWIS
247 summary statistics. Additional PRS were generated weighting by MDD main additive
248 effects (PRS_D) and neuroticism main additive effects (PRS_N) using GWAS summary
249 statistics from GS:SFHS or UKB. In addition, PRS_D and PRS_N were also generated using
250 summary statistics from the most recent Psychiatric Genetic Consortium (PGC) MDD
251 meta-analysis[40] (excluding GS:SFHS, and UKB individuals when required; N = 155 866 &
252 138 884) and the Genetics of Personality Consortium (GPC) neuroticism meta-
253 analysis[23,75] (N = 63 661). Generalized linear models were implemented in R 3.1.3[76]. The
254 direct effect of PRS_{SS} (model 1), PRS_D (model 2) and PRS_N (model 3) on MDD risk were
255 assessed in independent logistic regression models on GS:SFHS (target cohort) using
256 GWAS and GWIS statistics from UKB (the largest cohort) as the discovery sample to
257 weight PRS. Multiple regression models fitting both PRS_D and PRS_N (model 4) and fitting
258 each of them separately with PRS_{SS} (models 5 and 6) were also calculated. Finally, full
259 additive multiple regression models fitting PRS weighted by all three effects (full model)
260 was assessed using both PRS_{SS}, PRS_D and PRS_N at their best-fit in independent models.
261 Further, results were also assessed using PRS_D and PRS_N weighted by PGC2 MDD[40] and
262 GPC neuroticism[75] summary statistics. Further detail is given in the Supplemental
263 Information: Polygenic Profiling. All models were adjusted by sex, age and 20 PCs. A null
264 model was estimated from the direct effects of all covariates on MDD. Likelihood ratio
265 tests were used to assess significance of each PRS.
266 Cross-validation was performed using UKB as target sample and GS:SFHS as discovery
267 sample. Additional analyses using PRS_D and PRS_N weighted by PGC2 MDD[40] and GPC
268 neuroticism[75] summary statistics were also tested. MDD status on UKB was adjusted by

269 centre, array and genotyping batch as random effects and scaled (between 0 and 1) prior
270 to analysis, giving a quasi-binomial distribution of MDD status on UKB. Models
271 implemented on UKB (quasi-binomial regression) were adjusted by sex, age and 15 PCs.
272 Nagelkerke's R^2 coefficients were estimated to quantify the proportion of MDD liability
273 explained at the observed scale by each model and converted into R^2 coefficients at the
274 liability scale (prevalence: 12.2% in GS:SFHS[55] and 25.8% in UKB[61]) using Hong Lee's
275 transformation[77] available from GEAR: GEnetic Analysis Repository[78].
276
277 **Using stress-sensitivity to stratify depression**
278 GS:SFHS MDD cases (n_{cases} = 2 016; n_{female} = 1 345, n_{male} = 671) have data available on
279 MDD course (single or recurrent), age of onset (n = 1 964) and episode count (n = 2 016),
280 as well as on neuroticism (n = 2 010). In addition, a subset were evaluated by Mood
281 Disorder Questionnaire[79] (MDQ; n = 1 022) and Schizotypal Personality Questionnaire[80]
282 (SPQ; n = 1 093). The reduced sample number of MDQ and SPQ reflects the later addition
283 of these questionnaires to the study and does not reflect a particular subgroup of
284 GS:SFHS.
285 Difference in PRS_{SS} and PRS_D between MDD cases and controls on GS:SFHS were tested
286 using a Student's two sample t-test (two tailed). Cases of MDD on GS:SFHS with data
287 available on each trait analyzed were stratified by quintiles based on PRS_{SS} and PRS_D
288 (5x5 groups). Post hoc, the effects on each trait of quintiles based on PRS_{SS} and its
289 interaction effect with quintiles based on PRS_D were assessed using linear regression
290 models adjusting by sex and age in an attempt to identify a characteristic subtype of MDD
291 patients with differential stress-sensitivity levels. The same analysis was reproduced
292 using PRSs as continuous variables.
293

294 **RESULTS**

295 We confirmed the elevated neuroticism score in MDD cases in our samples. Individuals
296 with a diagnosis of MDD had significantly higher EPQN scores compared to healthy
297 controls (all $p < 1.9.x10^{-279}$) in both GS:SFHS ($mean_{controls}$ = 3.16; $mean_{cases}$ = 6.42) and
298 UKB ($mean_{controls}$ = 2.79; $mean_{cases}$ = 5.64). This difference remained significant after
299 matching cases with controls for PGC PRS_D, sex and age. (GS:SFHS: $mean_{controls}$ = 3.51;
300 UKB: $mean_{controls}$ = 2.97; all $p < 2.7x10^{-158}$; see Supplemental Table S1).

301

302 **Meta-analysis of stress-sensitivity in UKB and GS:SFHS**

303 No SNPs were associated with stress-sensitivity at the genome-wide significant
304 threshold ($p < 5x10^{-8}$, Figure 1). However, 14 SNPs from 8 loci achieved suggestive *p*
305 value ($p < 1x10^{-5}$) ranging between $p = 8.9x10^{-6}$-$5.1x10^{-7}$ (Meta-analysis: Table 1; UKB
306 and GS:SFHS: Supplemental Tables S2 & S3; Meta-analysis QQ-plot with λ: Supplemental
307 Figure S2; UKB and GS:SFHS QQ-plots: Supplemental Figure S3). Traits with prior
308 evidence of association with the nearest genes to the 8 lead SNPs were identified using
309 dbGap and are shown in Supplemental Table S4. Comparison between the SNP
310 association profile along the genome between stress-sensitivity GWIS and MDD GWAS
311 meta-analyses is shown in Miami plots filtering for the most significant stress-sensitivity
312 or MDD SNPs ($p < 0.001$; Meta-analysis: Figure 2; UKB and GS:SFHS: Supplemental Figure
313 S4). No SNP with a *p*-value < 0.01 had a corresponding p-value in the alternate trait,
314 suggesting that different variants contribute to depression and stress-sensitivity. Gene-
315 based test identified *ZNF366* as the only gene achieving genome-wide significance (p =
316 $1.48x10^{-7}$; Bonferroni-corrected significance threshold $p < 2.79x10^{-6}$; Supplemental Table
317 S5 & Supplemental Figure S5). Using summary statistics from meta-analysis GWIS
318 results, stress-sensitivity SNP-based heritability was estimated at 5.0% (h^2 = 0.0499, s.e.

319 = 0.017, p = 1.67x10^{-3}). Conversely, matched SNP-based heritabilities for MDD and
320 neuroticism were estimated at 9.6% (h^2 = 0.0962, s.e. = 0.0179, p = 3.87x10^{-8}) and 10.1%
321 (h^2 = 0.1006, s.e. = 0.0076, p = 3.47x10^{-40}) respectively, using summary statistics from
322 the meta-analysed GWAS of UKB and GS:SFHS.
323 LD score regression was performed to obtain genetic correlations between stress-
324 sensitivity, MDD and neuroticism. As previously shown, there was a significant genetic
325 correlation between MDD and neuroticism (r_G = 0.637, s.e. = 0.0704, p = 1.39x10^{-19}).
326 However, we found no evidence for a genetic correlation between stress-sensitivity and
327 MDD (rG = -0.099, s.e. = 0.182, p = 0.585) or between stress-sensitivity and neuroticism
328 (r_G = 0.114, s.e. = 0.107, p = 0.285).
329
330 **Pathway enrichment, functional annotation and gene expression analyses**
331 Lead SNPs from the GWIS meta-analysis were investigated using DEPICT. No gene
332 showed statistically significant links to stress-sensitivity at a DEPICT false discovery rate
333 (FDR) < 0.05. No significant result was found for either gene set analysis or tissue
334 enrichment analysis at FDR < 0.05. Evidence of regulatory elements on normal cell
335 lines/tissues was identified for 5 of the 12 lead SNPs (i.e. rs3762096, rs10987199,
336 rs2221540, rs246565, rs319924). Two lead SNPs were associated with eQTLs: rs319924
337 (an intronic SNP in *EYS*) and rs9509508 (an intronic SNP in *LATS2*) and potentially
338 regulate *LGSN/RP3-407E4.3* (p = 6.31x10^{-12}/ p = 1.15x10^{-5}) and *LATS2* (p = 3.74x10^{-8}),
339 respectively.
340
341 **Polygenic risk scores for stress-sensitivity predict MDD liability.**
342 PRS were used to investigate whether common variants affecting stress-sensitivity
343 predict MDD risk. We generated PRS (PRS_{SS}) based on the summary statistics from the

344 GWIS for stress-sensitivity. PRS_{SS} significantly predicted MDD risk in GS:SFHS using
345 weights from the larger UKB summary data (p = 1.8x10^{-3}; β = 0.087, s.e. = 0.028; best-fit
346 p threshold = 0.004). This association was cross-validated (nominal significance $p < 0.05$)
347 in UKB using summary data from the smaller GS:SFHS GWIS (p = 0.04; β = 0.027, s.e. =
348 0.013; best-fit p threshold = 1.15x10^{-4}; Table 2 and Supplemental Figure S6 & S7). On the
349 liability scale, the MDD variance explained by PRS_{SS} was modest, 0.243% in GS:SFHS and
350 0.037% in UKB. This was less than predicted by PRS weighted by the genetic main effects
351 of MDD or neuroticism (PRS_D: 0.386% in GS:SFHS and 0.211% in UKB; PRS_N: 0.34% in
352 GS:SFHS and 0.178% in UKB; Table 2 and Supplemental Figure S6).
353 Due to the known genetic correlations between MDD, neuroticism and stressful life
354 events[20], models jointly fitting the effects of multiple PRS were analysed. Multiple
355 regression analyses showed that, compared to PRS_D effects alone, stress-sensitivity effect
356 significantly explains an additional 0.248% in GS:SFHS (a predictive improvement of
357 64.3%; p = 1.6x10^{-3}; PRS_D: β = 0.113, s.e. = 0.028; PRS_{SS} : β = 0.088, s.e. = 0.028) and an
358 additional 0.036% in UKB (a predictive improvement of 17.1%; p = 0.044; PRS_D: β =
359 0.064, s.e. = 0.013; PRS_{SS} : β = 0.027, s.e. = 0.013; Figure 3). The inclusion of PRS_{SS} in the
360 full model, where PRS_{SS} was fitted along with both PRS_D and PRS_N weighted by GWAS
361 summary statistics derived from the same datasets, remained significant; explaining an
362 additional 0.245% (a predictive improvement of 35.8%: full model p = 1.7x10^{-3}; PRS_D: β =
363 0.107, s.e. = 0.029; PRS_N: β = 0.123, s.e. = 0.036; PRS_{SS}: β = 0.088, s.e. = 0.028) and an
364 additional 0.036% (a predictive improvement of 11.2%; full model p = 0.044; PRS_D: β =
365 0.054, s.e. = 0.014; PRS_N: β = 0.048, s.e. = 0.014; PRS_{SS} : β = 0.027 and s.e. = 0.013) in
366 GS:SFHS and UKB, respectively.
367 In models fitting PRS_D and PRS_N, the variances explained were non-additive,
368 demonstrating the partial overlap between MDD risk prediction from PRS_D and PRS_N

369 main additive effects. This overlap was not seen between variance explained by stress-
370 sensitivity effect and the variance explained by either PRS_D and/or PRS_N.
371 Multiple regression analyses fitting PRS_D and PRS_N derived from worldwide consortiums
372 (Figure 3) showed that the increased sample size from GWAS used to derive PRS_D
373 resulted in an increment of MDD variance explained by PRS_D (GS:SFHS: from 0.386% to
374 1.233%; UKB: from 0.21% to 1.281%). However, there was no change in the proportion
375 of the variance explained by the PRS_{SS} in the full model (PRS_{SS} $p = 1.29x10^{-3}$ and $p = 0.042$
376 in GS:SFHS and UKB, respectively). These results suggest that PRS_{SS} explains a
377 proportion of MDD risk not accounted for by PRS_D or PRS_N at current sample sizes.
378
379 **Using stress-sensitivity to stratify MDD in GS:SFHS**
380 MDD cases show significantly higher PRS_{SS} ($p = 2x10^{-3}$) and PRS_D ($p = 1.8x10^{-4}$) than
381 controls. Association between MDD-related traits and stress-sensitivity risk quintiles
382 was assessed on MDD cases in order to identify a subgroup of MDD patients, perhaps
383 defining a characteristic aetiological subtype of MDD. However, stratification analysis
384 failed, and no quintile based on PRS_{SS} nor its interaction with quintiles based on PRS_D
385 showed statistically significant effects on any trait analyzed. Individuals with high PRS_{SS}
386 were not significantly different from other cases for sex, MDD course, age of onset or
387 episode count, nor neuroticism, mood disorder or schizotypal personality scores ($p >$
388 0.05; see Supplemental Table S6). Results remained non significant when PRSs were
389 used as continuous variables ($p > 0.05$).

390 **DISCUSSION**

391 The existence of genetic variants affecting an individual's risk of depression in response
392 to stress has been predicted previously[44,47,48] and is consistent with the departure from a
393 simple additive genetic model seen in twin-studies of recurrent depressive disorder[81].
394 Through international research efforts such as the PGC and UK Biobank, there are ever-
395 increasing sample sizes available for understanding the genetics of MDD. These
396 resources are beginning, and will continue to, identify genome-wide significant loci[82,83
397 40]. However, the lack of environmental data and/or their reliability, makes the study of
398 genetic individual's response to their negative effects, and their contribution to the onset
399 of MDD and other stress-related disorders, difficult. As a way to address this limitation,
400 we generated a proxy for stress-sensitivity through modelling the interaction between
401 SNP allele and MDD status on neuroticism score in a GWIS approach. Thus, we sought to
402 identify the genetic underpinnings of individual's sensitivity to stress response (stress-
403 sensitivity) through those variants that contribute to higher neuroticism levels only in
404 individuals with a lifetime diagnosis of MDD but not in healthy controls.

405 We performed a GWIS to identify loci showing differential effects on neuroticism scores
406 in individuals with and without MDD (so called stress-sensitivity proxy). No SNPs
407 reached genome-wide significance, but 14 SNPs from 8 loci reached suggestive
408 significance levels. The closest genes to these SNPs for stress-sensitivity have prior
409 evidence of association in dbGAP with a wide range of neuropsychiatric traits such as
410 schizophrenia, bipolar disorder, attention deficit disorder with hyperactivity, mental
411 competency, intuition, sleep or alcohol drinking (Supplemental Table S4: Traits with
412 significant evidence of association with suggestive stress-sensitivity genes). However,
413 enrichment analysis showed no evidence for enrichment of specific pathways or tissues.
414 The top two loci, *PTP4A1-PHF3-EYS* and *ZNF366* have been previously associated with

415 alcohol dependence[84-88], alcohol intake (dbGaP: phs000342) and glucocorticoid receptor
416 function[89-91]. The most significant SNP in this study, rs319924, is an intronic variant in
417 *EYS* that is a potential eQTL for *LGSN*[74], a gene previously associated with male-specific
418 depression[92]. This is of particular interest given previous studies linking alcohol
419 consumption, stress and the risk of depression[93-98]. Evidence of an eQTL effect was also
420 predicted for a lead SNP in *LATS2*, a positive regulator of histone methyltransferase
421 activity[99] a process important in anxiety-related behaviours[100]. The prior association of
422 the top two loci in this study with alcohol related-phenotypes suggests that genes
423 involved in the sensitivity to stress may mediate the effects of stress on alcohol
424 consumption. Some *PHF3* paralogs have been shown to be linked with depression and
425 modulate stress response [101,102].
426 Gene-based analysis identified a genome-wide significant association between *ZNF366*
427 and stress-sensitivity. *ZNF366* (also known as *DC-SCRIPT*) is a corepressor of
428 transcription found in nuclear receptor complexes including the glucocorticoid receptor.
429 *ZNF366* represses glucocorticoid receptor-mediated transcription in monocyte-derived
430 dendritic cells[89]; and may act through histone deacetylases to modulate immune
431 response[90]. There is evidence from a large-scale mRNA display study that *PHF3*, in the
432 region underlying the most significant peak in the single SNP analysis, may also interact,
433 directly or indirectly, with the glucocorticoid receptor (IntAct database[91]) but this has
434 not been confirmed. These results reinforce the hypothesis that our proxy for stress-
435 sensitivity truly reflects the genetic architecture of sensitivity to respond to stress.
436 We estimated a significant lower bound on heritability for stress-sensitivity of 5%.
437 Whilst the known genetic overlap between MDD and neuroticism was detectable, the
438 lack of genetic correlation with stress-sensitivity indicated that there is no significant
439 overlap between stress-sensitivity and MDD or neuroticism. This analysis may be limited

440 by our sample size, however using the largest available meta-analyses of MDD and
441 neuroticism[40,75] did not decrease the proportion of liability explained by the PRS_{SS}. We
442 note, however, that as such meta-analyses increase in size it is likely, as with the effects
443 of smoking in schizophrenia[103,104], that the indirect genetic effects of the environment on
444 the risk of depression will be detected by GWAS. However, it is only through studies such
445 as ours that the mechanism for the effect of the risk alleles may be clarified.
446 Further, we show that such genetic information in stress-sensitivity could significantly
447 improve the proportion of liability to MDD predicted by PRS based only on additive
448 genetic effects on MDD identified by large GWAS. The summary results from the GWIS
449 were used to derive a PRS reflecting the genetic difference in stress-sensitivity. Although
450 not robust to multiple testing correction, this variable significantly predicted liability to
451 MDD in GS:SFHS (p = $1.8x10^{-3}$) and this was replicated in UKB (p = 0.04). This is
452 consistent with the expectation that the larger the discovery sample (i.e. UKB), the
453 greater the accuracy of the weighting and the more predictive the PRS[105]. Multiple
454 regression models suggest that inclusion of PRS weighted by stress-sensitivity
455 significantly improves MDD prediction over use of either MDD and/or neuroticism
456 weighted PRS alone (GS:SFHS full model p = $1.7x10^{-3}$; UKB full model p = 0.04). However,
457 we were unable to identify a subgroup of MDD cases with higher PRS_{ss}. The polygenic
458 interaction approach used in our study may, therefore, improve the interpretation of
459 both positive and negative findings from GWAS studies (i.e. pathways and mechanisms
460 involved, lack of replication, or negative findings in variants mediating environmental
461 effects). Added to paralleling recent developments in GWAS analyses, it may maximize
462 our power to detect gene-by-environment effects in this heterogeneous disorder
463 Future studies will be required to further investigate the effects of adverse life events in
464 individuals with high or low polygenic risk scores for stress-sensitivity. However, the

465 methodology presented allows addressing the genetic response to negative outcomes via
466 proxy in the absence of prospective environmental data.
467 Here we identify an independent set of risk variants for an individual's response to
468 negative outcomes and show that incorporating information across many loci provides
469 clear and replicable evidence for a genetic effect of stress-sensitivity on MDD risk;
470 identifying a potential genetic link with alcohol intake. These results require further
471 study, but may inform treatment of comorbid alcohol dependency and depression.

472 **ACKNOWLEDGMENTS**

473 The 1st author AAS is funded by University of Edinburgh (www.ed.ac.uk) and Medical
474 Research Council for his PhD study at the University of Edinburgh Institute of Genetics
475 and Molecular Medicine (www.ed.ac.uk/igmm). MA is supported by STRADL through a
476 Wellcome Trust Strategic Award (reference 104036/Z/14/Z). Generation Scotland
477 received core support from the Chief Scientist Office of the Scottish Government Health
478 Directorates [CZD/16/6] and the Scottish Funding Council [HR03006]. Genotyping of the
479 GS:SFHS samples was carried out by the Genetics Core Laboratory at the Wellcome Trust
480 Clinical Research Facility, Edinburgh, Scotland and was funded by the Medical Research
481 Council UK and the Wellcome Trust (Wellcome Trust Strategic Award "STratifying
482 Resilience and Depression Longitudinally" (STRADL), Reference 104036/Z/14/Z). The
483 Psychiatric Genomics Consortium MDD Working Group depends on the contributions of
484 many parties.

485

486 **FINANCIAL DISCLOSURE**

487 The authors declare no conflict of interest.

488 **REFERENCES**

489 1. Monroe SM. Modern approaches to conceptualizing and measuring human life
490 stress. *Annu Rev Clin Psychol* 2008; **4:** 33-52.
491
492 2. Jeronimus BF, Ormel J, Aleman A, Penninx BW, Riese H. Negative and positive life
493 events are associated with small but lasting change in neuroticism. *Psychol Med* 2013;
494 **43**(11): 2403-2415.
495
496 3. Jeronimus BF, Riese H, Sanderman R, Ormel J. Mutual reinforcement between
497 neuroticism and life experiences: a five-wave, 16-year study to test reciprocal causation.
498 *J Pers Soc Psychol* 2014; **107**(4): 751-764.
499
500 4. Riese H, Snieder H, Jeronimus BF, Korhonen T, Rose RJ, Kaprio J, *et al*. Timing of
501 Stressful Life Events Affects Stability and Change of Neuroticism. *European Journal of*
502 *Personality* 2014; **28**(2): 193-200.
503
504 5. Kendler KS, Kuhn J, Prescott CA. The interrelationship of neuroticism, sex, and
505 stressful life events in the prediction of episodes of major depression. *Am J Psychiatry*
506 2004; **161**(4): 631-636.
507
508 6. Kessler RC. The effects of stressful life events on depression. *Annu Rev Psychol*
509 1997; **48:** 191-214.
510
511 7. Tennant C. Life events, stress and depression: a review of recent findings. *Aust N Z*
512 *J Psychiatry* 2002; **36**(2): 173-182.
513
514 8. Porcelli B, Pozza A, Bizzaro N, Fagiolini A, Costantini MC, Terzuoli L, *et al*.
515 Association between stressful life events and autoimmune diseases: A systematic review
516 and meta-analysis of retrospective case-control studies. *Autoimmun Rev* 2016; **15**(4):
517 325-334.
518
519 9. Lin Y, Wang C, Zhong Y, Huang X, Peng L, Shan G, *et al*. Striking life events
520 associated with primary breast cancer susceptibility in women: a meta-analysis study. *J*
521 *Exp Clin Cancer Res* 2013; **32**(1): 53.
522
523 10. Jafri SHR, Ali F, Mollaeian A, Hasan SM, Hussain R, Akkanti BH, *et al*. Major
524 stressful life events and risk of developing lung cancer. *Journal of Clinical Oncology* 2017;
525 **35**(15_suppl): 1575-1575.
526
527 11. Hasler G, Drevets WC, Manji HK, Charney DS. Discovering endophenotypes for
528 major depression. *Neuropsychopharmacology* 2004; **29**(10): 1765-1781.
529
530 12. Bogdan R, Nikolova YS, Pizzagalli DA. Neurogenetics of depression: a focus on
531 reward processing and stress sensitivity. *Neurobiol Dis* 2013; **52:** 12-23.
532

533 13. Lazarus RS, Folkman S. Transactional theory and research on emotions and
534 coping. *European Journal of Personality* 1987; **1**(3): 141-169.
535
536 14. Smith MA, Riccalton VC, Kelly-Hughes DH, Craw OA, Allen SF, O'Connor DB, *et al*.
537 The relationship between Type D personality and physical health complaints is
538 mediated by perceived stress and anxiety but not diurnal cortisol secretion. *Stress* 2018:
539 1-8.
540
541 15. Herr RM, Barrech A, Riedel N, Gundel H, Angerer P, Li J. Long-Term Effectiveness
542 of Stress Management at Work: Effects of the Changes in Perceived Stress Reactivity on
543 Mental Health and Sleep Problems Seven Years Later. *Int J Environ Res Public Health*
544 2018; **15**(2).
545
546 16. Moore RC, Eyler LT, Mausbach BT, Zlatar ZZ, Thompson WK, Peavy G, *et al*.
547 Complex interplay between health and successful aging: role of perceived stress,
548 resilience, and social support. *Am J Geriatr Psychiatry* 2015; **23**(6): 622-632.
549
550 17. Hayman LW, Jr., Lucas T, Porcerelli JH. Cognitive appraisal vs. exposure-based
551 stress measures: links to perceived mental and physical health in low-income black
552 women. *J Nerv Ment Dis* 2014; **202**(11): 807-812.
553
554 18. Harris ML, Loxton D, Sibbritt DW, Byles JE. The influence of perceived stress on
555 the onset of arthritis in women: findings from the Australian Longitudinal Study on
556 women's health. *Ann Behav Med* 2013; **46**(1): 9-18.
557
558 19. Caspi A, Hariri AR, Holmes A, Uher R, Moffitt TE. Genetic sensitivity to the
559 environment: the case of the serotonin transporter gene and its implications for
560 studying complex diseases and traits. *Am J Psychiatry* 2010; **167**(5): 509-527.
561
562 20. Rietschel L, Zhu G, Kirschbaum C, Strohmaier J, Wust S, Rietschel M, *et al*.
563 Perceived stress has genetic influences distinct from neuroticism and depression. *Behav*
564 *Genet* 2014; **44**(6): 639-645.
565
566 21. Rietschel L, Streit F, Zhu G, McAloney K, Frank J, Couvy-Duchesne B, *et al*. Hair
567 Cortisol in Twins: Heritability and Genetic Overlap with Psychological Variables and
568 Stress-System Genes. *Sci Rep* 2017; **7**(1): 15351.
569
570 22. Glahn DC, Curran JE, Winkler AM, Carless MA, Kent JW, Jr., Charlesworth JC, *et al*.
571 High dimensional endophenotype ranking in the search for major depression risk genes.
572 *Biol Psychiatry* 2012; **71**(1): 6-14.
573
574 23. van den Berg SM, de Moor MH, McGue M, Pettersson E, Terracciano A, Verweij KJ,
575 *et al*. Harmonization of Neuroticism and Extraversion phenotypes across inventories
576 and cohorts in the Genetics of Personality Consortium: an application of Item Response
577 Theory. *Behav Genet* 2014; **44**(4): 295-313.
578

24. Wray NR, Birley AJ, Sullivan PF, Visscher PM, Martin NG. Genetic and phenotypic stability of measures of neuroticism over 22 years. *Twin Res Hum Genet* 2007; **10**(5): 695-702.

25. Vukasovic T, Bratko D. Heritability of personality: A meta-analysis of behavior genetic studies. *Psychol Bull* 2015; **141**(4): 769-785.

26. Hirschfeld RM, Klerman GL. Personality attributes and affective disorders. *Am J Psychiatry* 1979; **136**(1): 67-70.

27. Hirschfeld RM, Klerman GL, Clayton PJ, Keller MB, McDonald-Scott P, Larkin BH. Assessing personality: effects of the depressive state on trait measurement. *Am J Psychiatry* 1983; **140**(6): 695-699.

28. Levinson DF. The genetics of depression: a review. *Biol Psychiatry* 2006; **60**(2): 84-92.

29. Middeldorp CM, Cath DC, Van Dyck R, Boomsma DI. The co-morbidity of anxiety and depression in the perspective of genetic epidemiology. A review of twin and family studies. *Psychol Med* 2005; **35**(5): 611-624.

30. Smith DJ, Escott-Price V, Davies G, Bailey ME, Colodro-Conde L, Ward J, *et al*. Genome-wide analysis of over 106 000 individuals identifies 9 neuroticism-associated loci. *Mol Psychiatry* 2016; **21**(6): 749-757.

31. Luciano M, Hagenaars SP, Davies G, Hill WD, Clarke TK, Shirali M, *et al*. Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. *Nat Genet* 2018; **50**(1): 6-11.

32. Smits DJM, Boeck PD. From BIS/BAS to the big five. *European Journal of Personality* 2006; **20**(4): 255-270.

33. Schneider TR, Rench TA, Lyons JB, Riffle RR. The influence of neuroticism, extraversion and openness on stress responses. *Stress Health* 2012; **28**(2): 102-110.

34. Kim SE, Kim HN, Cho J, Kwon MJ, Chang Y, Ryu S, *et al*. Direct and Indirect Effects of Five Factor Personality and Gender on Depressive Symptoms Mediated by Perceived Stress. *PLoS One* 2016; **11**(4): e0154140.

35. Tak LM, Kingma EM, van Ockenburg SL, Ormel J, Rosmalen JG. Age- and sex-specific associations between adverse life events and functional bodily symptoms in the general population. *J Psychosom Res* 2015; **79**(2): 112-116.

36. Hovens JG, Giltay EJ, van Hemert AM, Penninx BW. Childhood Maltreatment and the Course of Depressive and Anxiety Disorders: The Contribution of Personality Characteristics. *Depress Anxiety* 2016; **33**(1): 27-34.

37. Bazana PG, Stelmack RM. Chapter 8 - Stability of Personality Across the Life Span: A Meta-Analysis. *On the Psychobiology of Personality*. Elsevier: Oxford, 2004, pp 113-144.

38. McGue M, Bacon S, Lykken DT. Personality Stability and Change in Early Adulthood. *Developmental Psychology* 1993; **29**(1): 96-109.

39. Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry* 2000; **157**(10): 1552-1562.

40. Wray NR, Sullivan PF. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *bioRxiv* 2017.

41. Loret de Mola C, de Franca GV, Quevedo Lde A, Horta BL. Low birth weight, preterm birth and small for gestational age association with adult depression: systematic review and meta-analysis. *Br J Psychiatry* 2014; **205**(5): 340-347.

42. Peyrot WJ, Milaneschi Y, Abdellaoui A, Sullivan PF, Hottenga JJ, Boomsma DI, *et al*. Effect of polygenic risk scores on depression in childhood trauma. *Br J Psychiatry* 2014; **205**(2): 113-119.

43. Mullins N, Power RA, Fisher HL, Hanscombe KB, Euesden J, Iniesta R, *et al*. Polygenic interactions with environmental adversity in the aetiology of major depressive disorder. *Psychol Med* 2016; **46**(4): 759-770.

44. Colodro-Conde L, Couvy-Duchesne B, Zhu G, Coventry WL, Byrne EM, Gordon S, *et al*. A direct test of the diathesis-stress model for depression. *Mol Psychiatry* 2017.

45. Kendler KS, Baker JH. Genetic influences on measures of the environment: a systematic review. *Psychol Med* 2007; **37**(5): 615-626.

46. Luciano M, Batty GD, McGilchrist M, Linksted P, Fitzpatrick B, Jackson C, *et al*. Shared genetic aetiology between cognitive ability and cardiovascular disease risk factors: Generation Scotland's Scottish family health study. *Intelligence* 2010; **38**(3): 304-313.

47. Silberg J, Rutter M, Neale M, Eaves L. Genetic moderation of environmental risk for depression and anxiety in adolescent girls. *Br J Psychiatry* 2001; **179**: 116-121.

48. Kendler KS, Kessler RC, Walters EE, MacLean C, Neale MC, Heath AC, *et al*. Stressful life events, genetic liability, and onset of an episode of major depression in women. *Am J Psychiatry* 1995; **152**(6): 833-842.

49. Biobank U (2010). UK Biobank: protocol for a large-scale prospective epidemiological resource.

50. Biobank U (2007). UK Biobank ethics and governance framework, version 3.0

51. Biobank U. Genotype imputation and genetic association studies of UK Biobank. 2015.

52. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**(3): 559-575.

53. Amador C, Huffman J, Trochet H, Campbell A, Porteous D, Generation S, *et al*. Recent genomic heritage in Scotland. *BMC Genomics* 2015; **16:** 437.

54. Smith BH, Campbell A, Linksted P, Fitzpatrick B, Jackson C, Kerr SM, *et al*. Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *Int J Epidemiol* 2013; **42**(3): 689-700.

55. Fernandez-Pujals AM, Adams MJ, Thomson P, McKechanie AG, Blackwood DH, Smith BH, *et al*. Epidemiology and Heritability of Major Depressive Disorder, Stratified by Age of Onset, Sex, and Illness Course in Generation Scotland: Scottish Family Health Study (GS:SFHS). *PLoS One* 2015; **10**(11): e0142197.

56. Smith BH, Campbell H, Blackwood D, Connell J, Connor M, Deary IJ, *et al*. Generation Scotland: the Scottish Family Health Study; a new resource for researching genes and heritability. *BMC medical genetics* 2006; **7:** 74.

57. Kerr SM, Campbell A, Murphy L, Hayward C, Jackson C, Wain LV, *et al*. Pedigree and genotyping quality analyses of over 10,000 DNA samples from the Generation Scotland: Scottish Family Health Study. *BMC Med Genet* 2013; **14:** 38.

58. Biobank U (2012). Touchscreen questionnaire.

59. Eysenck HJ ES (1975). Manual of the Eysenck Personality Questionnaire. Hodder and Stoughton: London.

60. Kendler KS, Myers J. The genetic and environmental relationship between major depression and the five-factor model of personality. *Psychol Med* 2010; **40**(5): 801-806.

61. Smith DJ, Nicholl BI, Cullen B, Martin D, Ul-Haq Z, Evans J, *et al*. Prevalence and characteristics of probable major depression and bipolar disorder within UK biobank: cross-sectional study of 172,751 participants. *PLoS One* 2013; **8**(11): e75362.

62. Eysenck SBG, Eysenck, H.J. & Barrett, P. A revised version of the psychoticism scale. *Personality and Individual Differences* 1985; **6:** 21-29.

63. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. *JAMA* 1999; **282**(18): 1737-1744.

64. First MB SR, Gibbon M, Williams JB (2002). Structured Clinical Interview for DSM-IV-TR Axis I Disorders. New York State Psychiatric Institute: New York.

65. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. *Bioinformatics* 2015; **31**(9): 1466-1468.

66. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010; **26**(17): 2190-2191.

67. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* 2015; **11**(4): e1004219.

68. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. FUMA: Functional mapping and annotation of genetic associations. *bioRxiv* 2017.

69. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics C, *et al*. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 2015; **47**(3): 291-295.

70. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, *et al*. An atlas of genetic correlations across human diseases and traits. *Nat Genet* 2015; **47**(11): 1236-1241.

71. Pers TH, Karjalainen JM, Chan Y, Westra HJ, Wood AR, Yang J, *et al*. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* 2015; **6**: 5890.

72. Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, *et al*. The NCBI dbGaP database of genotypes and phenotypes. *Nat Genet* 2007; **39**(10): 1181-1186.

73. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, *et al*. Annotation of functional variation in personal genomes using RegulomeDB. *Genome research* 2012; **22**(9): 1790-1797.

74. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 2015; **348**(6235): 648-660.

75. Genetics of Personality C, de Moor MH, van den Berg SM, Verweij KJ, Krueger RF, Luciano M, *et al*. Meta-analysis of Genome-wide Association Studies for Neuroticism, and

760 the Polygenic Association With Major Depressive Disorder. *JAMA Psychiatry* 2015;
761 **72**(7): 642-650.
762
763 76. Team RC. *R: A language and environment for statistical computing*. R Foundation
764 for Statistical Computing: Vienna, Austria, 2015.
765
766 77. Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for
767 disease from genome-wide association studies. *Am J Hum Genet* 2011; **88**(3): 294-305.
768
769 78. Chen GB. Estimating heritability of complex traits from genome-wide association
770 studies using IBS-based Haseman-Elston regression. *Front Genet* 2014; **5**: 107.
771
772 79. Hirschfeld RM, Williams JB, Spitzer RL, Calabrese JR, Flynn L, Keck PE, Jr., *et al*.
773 Development and validation of a screening instrument for bipolar spectrum disorder:
774 the Mood Disorder Questionnaire. *Am J Psychiatry* 2000; **157**(11): 1873-1875.
775
776 80. Raine A. The SPQ: a scale for the assessment of schizotypal personality based on
777 DSM-III-R criteria. *Schizophr Bull* 1991; **17**(4): 555-564.
778
779 81. Polderman TJ, Benyamin B, de Leeuw CA, Sullivan PF, van Bochoven A, Visscher
780 PM, *et al*. Meta-analysis of the heritability of human traits based on fifty years of twin
781 studies. *Nat Genet* 2015; **47**(7): 702-709.
782
783 82. consortium C. Sparse whole-genome sequencing identifies two loci for major
784 depressive disorder. *Nature* 2015; **523**(7562): 588-591.
785
786 83. Hyde CL, Nagle MW, Tian C, Chen X, Paciga SA, Wendland JR, *et al*. Identification of
787 15 genetic loci associated with risk of major depression in individuals of European
788 descent. *Nat Genet* 2016; **48**(9): 1031-1036.
789
790 84. Zuo L, Tan Y, Zhang X, Wang X, Krystal J, Tabakoff B, *et al*. A New Genomewide
791 Association Meta-Analysis of Alcohol Dependence. *Alcohol Clin Exp Res* 2015; **39**(8):
792 1388-1395.
793
794 85. Cook MN, Baker JA, Heldt SA, Williams RW, Hamre KM, Lu L. Identification of
795 candidate genes that underlie the QTL on chromosome 1 that mediates genetic
796 differences in stress-ethanol interactions. *Physiol Genomics* 2015; **47**(8): 308-317.
797
798 86. Zuo L, Wang K, Wang G, Pan X, Zhang X, Zhang H, *et al*. Common PTP4A1-PHF3-
799 EYS variants are specific for alcohol dependence. *Am J Addict* 2014; **23**(4): 411-414.
800
801 87. Zuo L, Zhang X, Deng HW, Luo X. Association of rare PTP4A1-PHF3-EYS variants
802 with alcohol dependence. *J Hum Genet* 2013; **58**(3): 178-179.
803

804 88. Zuo L, Zhang CK, Wang F, Li CS, Zhao H, Lu L, *et al*. A novel, functional and
805 replicable risk gene region for alcohol dependence identified by genome-wide
806 association study. *PLoS One* 2011; **6**(11): e26726.
807
808 89. Hontelez S, Karthaus N, Looman MW, Ansems M, Adema GJ. DC-SCRIPT regulates
809 glucocorticoid receptor function and expression of its target GILZ in dendritic cells. *J*
810 *Immunol* 2013; **190**(7): 3172-3179.
811
812 90. Lopez-Garcia J, Periyasamy M, Thomas RS, Christian M, Leao M, Jat P, *et al*.
813 ZNF366 is an estrogen receptor corepressor that acts through CtBP and histone
814 deacetylases. *Nucleic Acids Res* 2006; **34**(21): 6126-6136.
815
816 91. Miyamoto-Sato E, Fujimori S, Ishizaka M, Hirai N, Masuoka K, Saito R, *et al*. A
817 comprehensive resource of interacting protein regions for refining human transcription
818 factor networks. *PLoS One* 2010; **5**(2): e9289.
819
820 92. Aragam N, Wang KS, Pan Y. Genome-wide association analysis of gender
821 differences in major depressive disorder in the Netherlands NESDA and NTR
822 population-based samples. *J Affect Disord* 2011; **133**(3): 516-521.
823
824 93. Fidalgo TM, da Silveira ED, da Silveira DX. Psychiatric comorbidity related to
825 alcohol use among adolescents. *Am J Drug Alcohol Abuse* 2008; **34**(1): 83-89.
826
827 94. Lopez B, Turner RJ, Saavedra LM. Anxiety and risk for substance dependence
828 among late adolescents/young adults. *J Anxiety Disord* 2005; **19**(3): 275-294.
829
830 95. Low NC, Lee SS, Johnson JG, Williams JB, Harris ES. The association between
831 anxiety and alcohol versus cannabis abuse disorders among adolescents in primary care
832 settings. *Fam Pract* 2008; **25**(5): 321-327.
833
834 96. Rutledge PC, Sher KJ. Heavy drinking from the freshman year into early young
835 adulthood: the roles of stress, tension-reduction drinking motives, gender and
836 personality. *J Stud Alcohol* 2001; **62**(4): 457-466.
837
838 97. Schmidt NB, Buckner JD, Keough ME. Anxiety sensitivity as a prospective
839 predictor of alcohol use disorders. *Behav Modif* 2007; **31**(2): 202-219.
840
841 98. Varlinskaya EI, Kim EU, Spear LP. Chronic intermittent ethanol exposure during
842 adolescence: Effects on stress-induced social alterations and social drinking in
843 adulthood. *Brain Res* 2017; **1654**(Pt B): 145-156.
844
845 99. Torigata K, Daisuke O, Mukai S, Hatanaka A, Ohka F, Motooka D, *et al*. LATS2
846 Positively Regulates Polycomb Repressive Complex 2. *PLoS One* 2016; **11**(7): e0158562.
847
848 100. Shen EY, Jiang Y, Javidfar B, Kassim B, Loh YE, Ma Q, *et al*. Neuronal Deletion of
849 Kmt2a/Mll1 Histone Methyltransferase in Ventral Striatum is Associated with Defective

850 Spike-Timing-Dependent Striatal Synaptic Plasticity, Altered Response to Dopaminergic
851 Drugs, and Increased Anxiety. *Neuropsychopharmacology* 2016; **41**(13): 3103-3113.
852
853 101. Wong ML, Arcos-Burgos M, Liu S, Velez JI, Yu C, Baune BT, *et al*. The PHF21B gene
854 is associated with major depression and modulates the stress response. *Mol Psychiatry*
855 2017; **22**(7): 1015-1025.
856
857 102. Walsh RM, Shen EY, Bagot RC, Anselmo A, Jiang Y, Javidfar B, *et al*. Phf8 loss
858 confers resistance to depression-like and anxiety-like behaviors in mice. *Nat Commun*
859 2017; **8:** 15142.
860
861 103. Gage SH, Jones HJ, Taylor AE, Burgess S, Zammit S, Munafo MR. Investigating
862 causality in associations between smoking initiation and schizophrenia using Mendelian
863 randomization. *Sci Rep* 2017; **7:** 40653.
864
865 104. Hartz SM, Horton AC, Hancock DB, Baker TB, Caporaso NE, Chen LS, *et al*. Genetic
866 correlation between smoking behaviors and schizophrenia. *Schizophr Res* 2017.
867
868 105. Dudbridge F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet*
869 2013; **9**(3): e1003348.
870
871

872 **FIGURE LEGENDS**

873

874 **Figure 1. Manhattan plots showing stress-sensitivity associations.** Manhattan plots
875 of the GWIS from **A)** UKB, **B)** GS:SFHS and **C)** sample size weighted meta-analysis of UKB
876 and GS:SFHS. The x-axis is chromosomal position and y-axis is the *p* value (-log10 *p*
877 value) of association with stress-sensitivity effect. Suggestive genome-wide significance
878 threshold ($p = 1x10^{-5}$) is shown by solid line at y = 5. Genes or closest gene up- and
879 down-stream from SNP position (/) are annotated. "-": No gene within 100kb of the SNP.

880

881 **Figure 2. Miami plots showing comparison between association profile between**
882 **stress-sensitivity GWIS and MDD GWAS.** Miami plots from meta-analysis filter at $p =$
883 $1x10^{-3}$: **A)** filtering for stress-sensitivity *p* values (•), **B)** filtering for MDD *p* values (×).
884 The x-axis is chromosomal position and y-axis is the *p* value (-log10 *p* value) of
885 association with stress-sensitivity (up; red dots) and MDD *p* value (down; blue crosses).
886 Dot line: genome-wide suggestive threshold ($p = 1x10^{-5}$) at the filtered effect; dashed
887 lines: $p = 0.01$ and 0.05 at unfiltered effect.

888

889 **Figure 3. MDD is best predicted using multiple PRS.** MDD risk explained (R^2
890 coefficient (%); top bar values) on the liability scale by each PRS; weighted by GWAS
891 main additive and GWIS stress-sensitivity effects independently and combined. **A)**
892 GS:SFHS: there is an increment on MDD risk prediction from adding PRS_{SS} to PRS_D model
893 of 64.3% and 35.8% when adding PRS_{SS} to both MDD and neuroticism PRS together. **B)**
894 UKB: there is an increment on MDD risk prediction from adding PRS_{SS} to PRS_D model of
895 17.1% and 11.2% when combining PRS_{SS} with both MDD and neuroticism PRS. **C & D)**
896 Replication of A & B fitting PRS_D and PRS_N using summary statistics from worldwide
897 consortiums (i.e. PGC &GPC). Significance codes: *p* values *** < 0.001 < ** < 0.01 < * <
898 0.05; derived from likelihood ratio tests. *SS* stands for stress-sensitivity.

Table 1. Top 25 SNPs from meta-analysis of GWISs.

Rank	CHR	SNP	BP	A1	Z-score	Effect	*p*	*p* (EPQN)	*p* (MDD)	GENE	POSITION
1	6	rs319924	64487247	A	5.024	++	$5.05x10^{-7}$	0.376	0.637	EYS	Intronic
2	5	rs246565	71809247	A	-4.684	--	$2.82x10^{-6}$	0.248	0.589	ZNF366	5998bp 5'
3	10	rs2265265	24854876	A	4.604	++	$4.15x10^{-6}$	0.035	0.084	KIAA1217 / ARHGAP21	18104bp 3' / 17662bp 3'
4	6	rs1057530	64427095	A	-4.556	--	$5.21x10^{-6}$	0.636	0.840	PHF3/EYS	1677bp 3' / 2781bp 3'
5	16	rs7199110	78790765	A	-4.553	--	$5.29x10^{-6}$	0.661	0.741	WWOX	Intronic
6	6	rs10485358	64386060	A	-4.546	--	$5.46x10^{-6}$	0.390	0.902	PHF3	Intronic
7	12	rs10778077	101193988	A	4.54	++	$5.62x10^{-6}$	0.614	0.430	ANO4	Intronic
8	5	rs13358894	71803446	A	4.527	++	$5.99x10^{-6}$	0.257	0.651	ZNF366	197bp 5'
9	10	rs2256220	24856314	A	-4.524	--	$6.06x10^{-6}$	0.134	0.129	KIAA1217 / ARHGAP21	19542bp 3' / 16224bp 3'
10	10	rs3762096	98136250	A	-4.521	--	$6.15x10^{-6}$	0.437	0.149	TLL2	Intronic
11	11	rs2221540	132716369	A	-4.492	--	$7.05x10^{-6}$	0.468	0.364	OPCML	Intronic
12	5	rs10043659	71781839	A	4.483	++	$7.37x10^{-6}$	0.339	0.808	ZNF366	Intronic
13	12	rs10778078	101195088	A	-4.45	--	$8.58x10^{-6}$	0.599	0.456	ANO4	Intronic
14	9	rs10987199	128968987	A	-4.442	--	$8.91x10^{-6}$	0.199	0.026	LOC101929116	63416bp 3'
15	5	rs10042132	71789021	A	-4.416	--	$1.01x10^{-5}$	0.418	0.538	ZNF366	Intronic
16	11	rs10894606	132671611	A	-4.404	--	$1.06x10^{-5}$	0.438	0.587	OPCML	Intronic
17	12	rs7295089	2440464	A	4.372	++	$1.23x10^{-5}$	0.266	0.212	CACNA1C	Intronic
18	5	rs9293292	71696942	A	-4.351	--	$1.36x10^{-5}$	0.126	0.731	PTCD2/ZNF366	41762bp 3' / 42292bp 3'
19	15	rs3097437	27872136	A	4.346	++	$1.38x10^{-5}$	0.970	0.226	GABRG3	93762bp 3'
20	9	rs1999377	11919732	A	4.344	++	$1.40x10^{-5}$	0.436	0.064	-	Intragenic
21	5	rs6862221	71754962	A	4.342	++	$1.41x10^{-5}$	0.543	0.823	ZNF366	Intronic
22	5	rs9293289	71683885	A	-4.323	--	$1.54x10^{-5}$	0.395	0.510	PTCD2/ZNF366	28705bp 3' / 55349bp 3'
23	11	rs4575282	132719646	A	-4.313	--	$1.61x10^{-5}$	0.598	0.514	OPCML	Intronic
24	9	rs2417008	128970219	A	-4.3	--	$1.71x10^{-5}$	0.208	0.026	LOC101929116	62184bp 3'
25	9	rs7021461	128972210	A	4.299	++	$1.72x10^{-5}$	0.202	0.025	LOC101929116	60193bp 3'

Effect shows effect direction in GS:SFHS and UK Biobank, *p* reflects significance of GWIS stress-sensitivity effect, *p (EPQN)* SNP main effect on neuroticism derived from GWAS meta-analysis of EPQN between UK Biobank and Generation Scotland. *p (MDD)* SNP main effect on MDD derived from GWAS meta-analysis of MDD between UK Biobank and Generation Scotland. *Position* position of the SNP respect to closest gene transcripts within 100kb (including UTRs) from 5 prime (5') or 3prime (3').

899

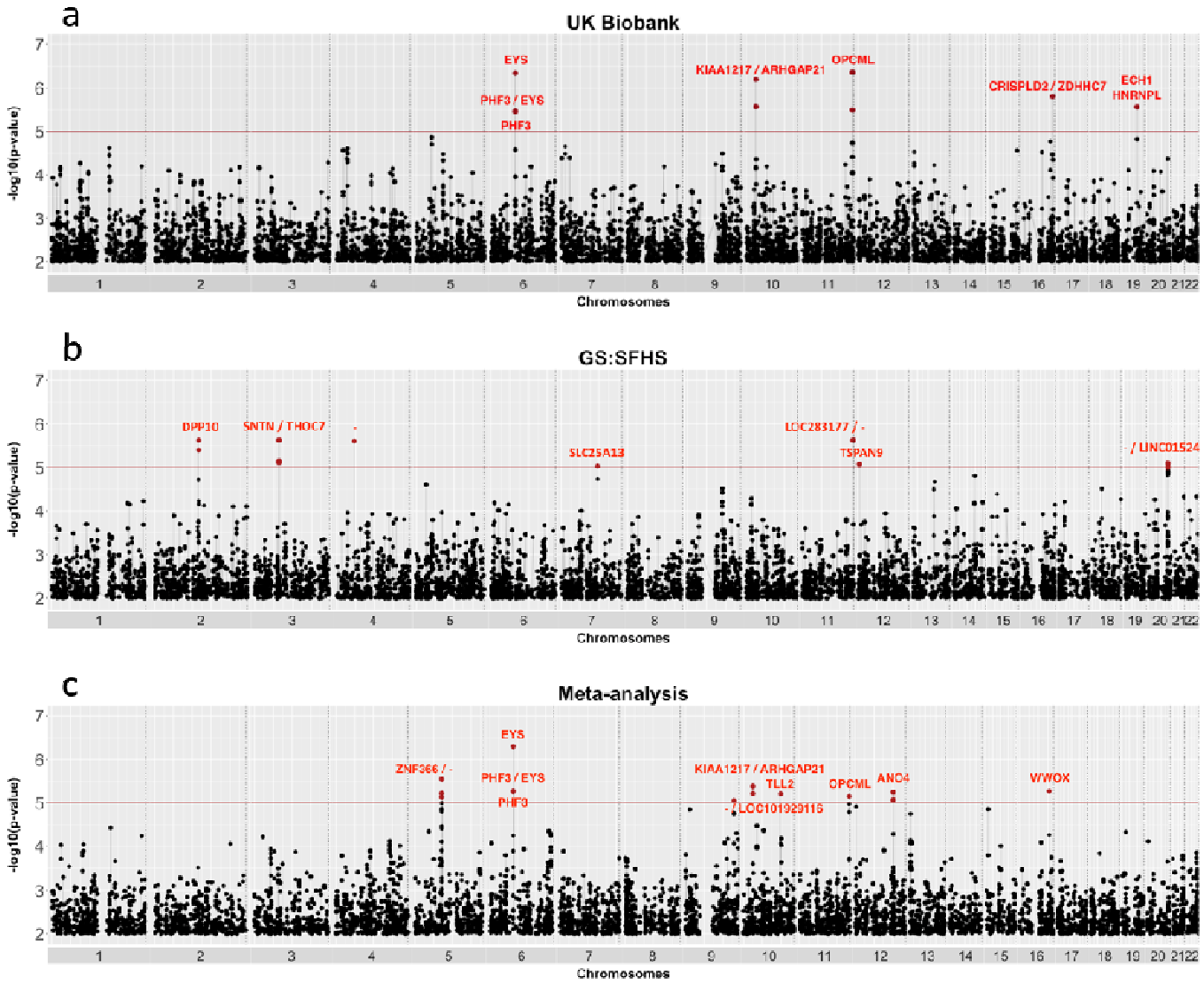
Table 2. MDD risk prediction at best-fits.

Weighted effect	Best-fit threshold	# SNPs	R^2 (%)[a]	R^2 (%)[b]	*p*
UKB predicting on GS:SFHS					
stress-sensitivity	0.004	1 313	0.175	0.243	$1.8x10^{-3}$
MDD	0.125	25 833	0.278	0.386	$8.3x10^{-5}$
EPQN	$1x10^{-3}$	863	0.245	0.340	$2.2x10^{-4}$
MDD + EPQN	-	-	0.492	0.684	$1.1x10^{-6}$
joint models	-	-	0.669	0.928	$3.9x10^{-8}$
PGC2 & GPC predicting on GS:SFHS					
PGC2 MDD	0.227	41 917	0.888	1.233	$1.9x10^{-12}$
GPC EPQN	0.013	4 234	0.101	0.140	0.018
PGC2 MDD + GPC EPQN	-	-	0.175	1.324	$2.7x10^{-12}$
joint models	-	-	1.138	1.580	$9.7x10^{-14}$

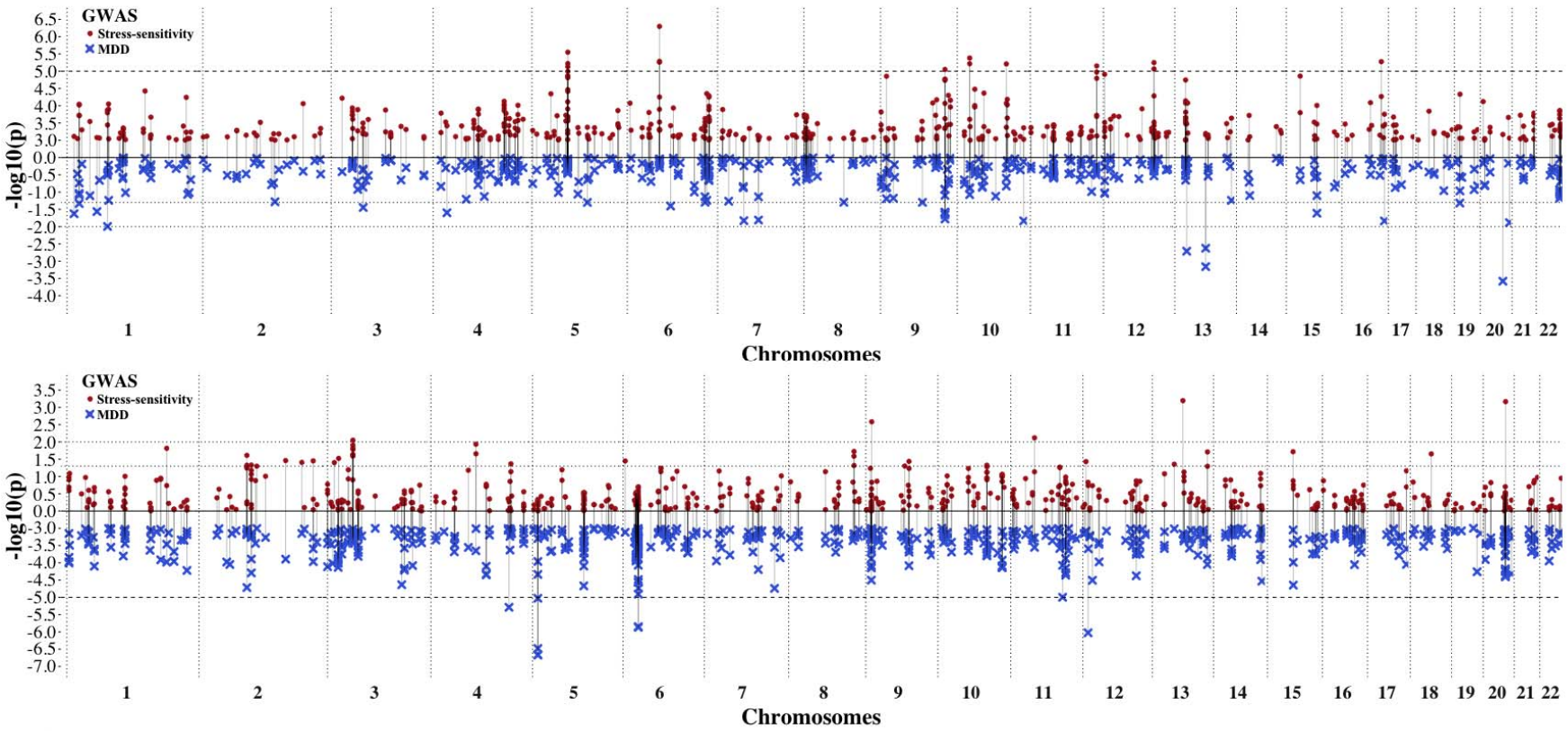
Weighted effect	Best-fit threshold	# SNPs	R^2 (%)[a]	R^2 (%)[b]	*p*
GS:SFHS predicting on UKB					
stress-sensitivity	$1.15x10^{-4}$	42	0.024	0.037	0.04
MDD	0.033	8 135	0.135	0.211	$1.1x10^{-6}$
EPQN	0.07	15 613	0.114	0.178	$7.3x10^{-6}$
MDD + EPQN	-	-	0.206	0.323	$1.3x10^{-8}$
joint models	-	-	0.229	0.359	$8.7x10^{-9}$
PGC2 & GPC predicting on UKB					
PGC2 MDD	0.095	22 005	0.818	1.281	$2.4x10^{-33}$
GPC EPQN	$6x10^{-3}$	2 182	0.067	0.105	$5.6x10^{-4}$
PGC2 MDD + GPC EPQN	-	-	0.863	1.352	$6.6x10^{-34}$
joint models	-	-	0.886	1.388	$1.8x10^{-32}$

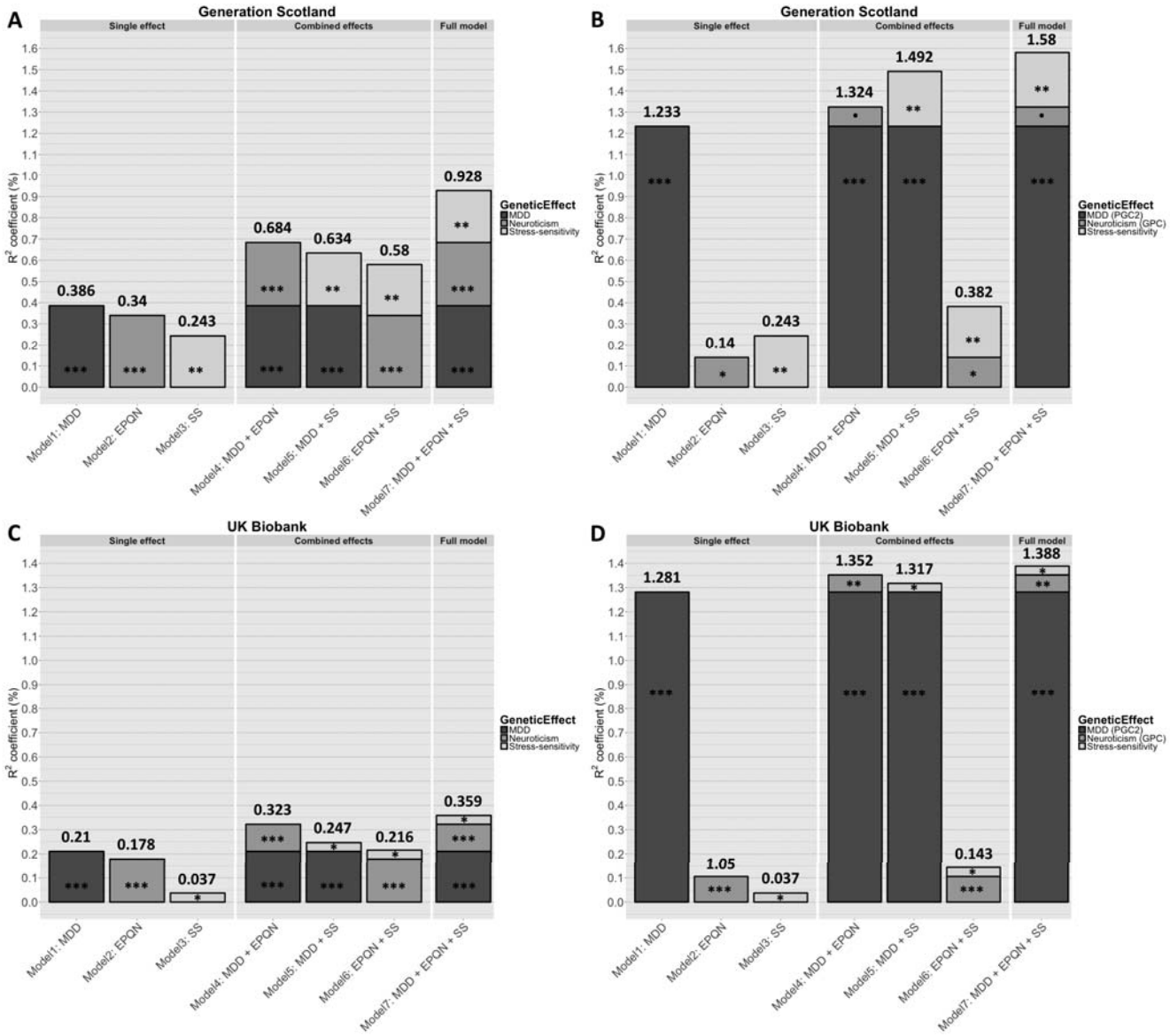
stress-sensitivity stress-sensitivity, *MDD* major depressive disorder, *EPQN* neuroticism score, *joint models* combined effect fitting all 3 PRS weighted by all the effects (i.e. stress-sensitivity, MDD and EPQN), [a]Nagelkerke's R^2 at observed scale, [b]R^2 on the liability scale

900 **Figure 1**
901



902 **Figure 2**



903

904 **Figure 3**
905