

Genome-wide by environment interaction studies (GWEIS) of depressive symptoms and psychosocial stress in UK Biobank and Generation Scotland.

Running title: **GWEIS of depressive symptoms in UK**

Aleix Arnau-Soler¹, Erin Macdonald-Dunlop², Mark J. Adams³, Toni-Kim Clarke³, Donald J. MacIntyre³, Keith Milburn⁴, Lauren Navrady³, Generation Scotland⁵, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium⁶, Caroline Hayward⁷, Andrew M. McIntosh³, Pippa A. Thomson¹

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

² Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Teviot Place, Edinburgh, UK.

³ Division of Psychiatry, Deanery of Clinical Sciences, University of Edinburgh, Royal Edinburgh Hospital, Morningside Park, Edinburgh EH10 5HF, UK

⁴ Health Informatics Centre, University of Dundee, UK.

⁵ A collaboration between the University Medical School and NHS in Aberdeen, Dundee, Edinburgh and Glasgow, Scotland, UK

⁶ For a full list of MDD working group of the PGC investigators, see the Supplementary Material.

⁷ Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

Correspondence should be addressed to: P. Thomson (Pippa.Thomson@ed.ac.uk) or A. Arnau-Soler (aleix.arnau.soler@igmm.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: 250

Main body (excl. abstract): 5 000

No. Tables: 1

No. Figures: 3

No. References: 100

Supplementary Material: 2 files

Supplementary Figures: 17

Supplementary Tables: 17

43 ABSTRACT

44 Stress is associated with poorer physical and mental health. To improve our
 45 understanding of this link, we performed genome-wide association studies (GWAS) of
 46 depressive symptoms and genome-wide by environment interaction studies (GWEIS)
 47 of depressive symptoms and stressful life events (SLE) in two UK population cohorts
 48 (Generation Scotland and UK Biobank). No SNP was individually significant in either
 49 GWAS, but gene-based tests identified six genes associated with depressive
 50 symptoms in UK Biobank (*DCC*, *ACSS3*, *DRD2*, *STAG1*, *FOXP2* and *KYNU*; $p < 2.77 \times 10^{-6}$).
 51 Two SNPs with genome-wide significant GxE effects were identified by GWEIS in
 52 Generation Scotland: rs12789145 (53kb downstream *PIWIL4*; $p = 4.95 \times 10^{-9}$; total SLE)
 53 and rs17070072 (intronic to *ZCCHC2*; $p = 1.46 \times 10^{-8}$; dependent SLE). A third locus
 54 upstream *CYLC2* (rs12000047 and rs12005200, $p < 2.00 \times 10^{-8}$; dependent SLE) when
 55 the joint effect of the SNP main and GxE effects was considered. GWEIS gene-based
 56 tests identified: *MTNR1B* with GxE effect with dependent SLE in Generation Scotland;
 57 and *PHF2* with the joint effect in UK Biobank ($p < 2.77 \times 10^{-6}$). Polygenic risk scores
 58 (PRS) analyses incorporating GxE effects improved the prediction of depressive
 59 symptom scores, when using weights derived from either the UK Biobank GWAS of
 60 depressive symptoms ($p = 0.01$) or the PGC GWAS of major depressive disorder ($p =$
 61 5.91×10^{-3}). Using an independent sample, PRS derived using GWEIS GxE effects
 62 provided evidence of shared aetiologies between depressive symptoms and
 63 schizotypal personality, heart disease and COPD. Further such studies are required
 64 and may result in improved treatments for depression and other stress-related
 65 conditions.

66

67 INTRODUCTION

68 Mental illness results from the interplay between genetic susceptibility and
 69 environmental risk factors^{1,2}. Previous studies have shown that the effects of
 70 environmental factors on traits may be partially heritable³ and moderated by
 71 genetics^{4,5}. Major depressive disorder (MDD) is the most common psychiatric
 72 disorder with a lifetime prevalence of approximately 14% globally⁶ and with a
 73 heritability of approximately 37%⁷. There is strong evidence for the role of stressful
 74 life events (SLE) as risk factor and trigger for depression⁸⁻¹². Genetic control of
 75 sensitivity to stress may vary between individuals, resulting in individual differences
 76 in the depressogenic effects of SLE, i.e., genotype-by-environment interaction
 77 (GxE)^{4,13-16}. Significant evidence of GxE has been reported for common respiratory
 78 diseases and some forms of cancer¹⁷⁻²², and GxE studies have identified genetic risk
 79 variants not found by genome-wide association studies (GWAS)²³⁻²⁷.

80 Interaction between polygenic risk of MDD and recent SLE are reported to increase
 81 liability to depressive symptoms^{4,16}; validating the implementation of genome-wide
 82 approaches to study GxE in depression. Most GxE studies for MDD have been
 83 conducted on candidate genes, or using polygenic approaches to a wide range of
 84 environmental risk factors, with some contradictory findings²⁸⁻³². Incorporating
 85 knowledge about recent SLE into GWAS may improve our ability to detect risk
 86 variants in depression otherwise missed in GWAS³³. To date, four studies have
 87 performed genome-wide by environment interaction studies (GWEIS) of MDD and
 88 SLE³⁴⁻³⁷, but this is the first study to perform GWEIS of depressive symptoms using
 89 adult SLE in cohorts of relatively homogeneous European ancestry.

90 Interpretation of GxE effects may be hindered by gene-environment correlation.
 91 Gene-environment correlation denotes a genetic mediation of associations through
 92 genetic influences on exposure to, or reporting of, environments^{2,38}. Genetic factors
 93 predisposing to MDD may contribute to exposure and/or reporting of SLE³⁹. To tackle
 94 this limitation, measures of SLE can be broken down into SLE likely to be independent
 95 of a respondent's own behaviour and symptoms, or into dependent SLE, in which
 96 participants may played an active role exposure to SLE^{40,41}. Different genetic
 97 influences with a higher heritability for reported dependent SLE than independent
 98 SLE^{39,42-45} suggest that whereas GxE driven by independent SLE is likely to reflect a
 99 genetic moderation of associations between SLE and depression, GxE driven by
 100 dependent SLE may result from a genetic mediation of the association through
 101 genetically driven personality or behavioural traits. To test this we analysed
 102 dependent and independent SLE scores separately in Generation Scotland.
 103 Stress contributes to many human conditions, with evidence of genetic vulnerability
 104 to the effect of SLE⁴⁶. Therefore, genetic stress-response factors in MDD may also
 105 underlie the aetiology of other stress-linked disorders, with which MDD is often co-
 106 morbid^{47,48} (e.g. cardiovascular diseases⁴⁹, diabetes,⁵⁰ chronic pain⁵¹ and
 107 inflammation⁵²). We tested the hypothesis that pleiotropy and shared aetiology
 108 between mental and physical health conditions may be due in part to genetic variants
 109 underlying SLE effects in depression.
 110 In this study we conduct GWEIS of depressive symptoms incorporating data on SLE in
 111 two independent UK-based cohorts. We aimed to: i) identify loci associated with
 112 depressive symptoms through genetic response to SLE; ii) study dependent and
 113 independent SLE to support a contribution of genetically mediated exposure to

114 stress; iii) assess whether GxE effects improve the proportion of phenotypic variance
115 in depressive symptoms explained by genetic additive main effects alone; and iv) test
116 for a significant overlap in the genetic aetiology of the response to SLE and mental
117 and physical stress-related phenotypes.
118

119 MATERIALS & METHODS

120 The core workflow of this study is summarized at Figure 1.

121

122 COHORT DESCRIPTIONS

123 **Generation Scotland (GS).** Generation Scotland is a family-based population cohort
124 representative of the Scottish population⁵³. At baseline, blood and salivary DNA
125 samples were collected, stored and genotyped at the Wellcome Trust Clinical
126 Research Facility, Edinburgh. Genome-wide genotype data was generated using the
127 Illumina HumanOmniExpressExome-8 v1.0 DNA Analysis BeadChip (San Diego, CA,
128 USA) and Infinium chemistry⁵⁴. The procedures and details for DNA extraction and
129 genotyping have been extensively described elsewhere^{55,56}. 21 525 participants were
130 re-contacted to participate in a follow-up mental health study (Stratifying Resilience
131 and Depression Longitudinally, STRADL), of which 8 541 participants responded
132 providing updated measures in psychiatric symptoms and SLE through self-reported
133 mental health questionnaires⁵⁷. Samples were excluded if: they were duplicate
134 samples, had diagnoses of bipolar disorder, no SLE data (non-respondents), were
135 population outliers (mainly non-Caucasians and Italian ancestry subgroup), had sex
136 mismatches, or were missing more than 2% of genotypes. SNPs were excluded if:
137 missing more than 2% of genotypes, Hardy-Weinberg Equilibrium test $p < 1 \times 10^{-6}$, or
138 minor allele frequency less than 1%. Further details of the GS and STRADL cohort are
139 available elsewhere^{53,57-59}. All components of GS and STRADL obtained ethical
140 approval from the Tayside Committee on Medical Research Ethics on behalf of the
141 NHS (reference 05/s1401/89). After quality control, individuals were filtered by
142 degree of relatedness (π -hat < 0.05), maximizing retention of those individuals

143 reporting a higher number of SLE. The final dataset comprised data on 4 919
144 unrelated individuals (1 929 men; 2 990 women) and 560 351 SNPs.

145 ***Independent GS datasets.*** Additional datasets for a range of stress-linked medical
146 conditions and personality traits were created from GS (N = 21 525) excluding
147 respondents and their relatives (N = 5 724). Following the same quality control
148 criteria detailed above, we maximized unrelated non-respondents for retention of
149 cases, or proxy cases (see below), to maximize the information available for each
150 phenotype. This resulted in independent datasets with unrelated individuals for each
151 trait. Differences between respondents and non-respondents are noted in the figure
152 legend of Table 1.

153 ***UK Biobank (UKB).*** This study used data from 99 057 unrelated individuals (47 558
154 men; 51 499 women) from the initial release of UKB genotyped data (released 2015;
155 under UK Biobank project 4844.). Briefly, participants were removed based on UKB
156 genomic analysis exclusion, non-white British ancestry, high missingness, genetic
157 relatedness (kinship coefficient > 0.0442), QC failure in UK BiLEVE study, and gender
158 mismatch. GS participants and their relatives were excluded and GS SNPs imputed to
159 a reference set combining the UK10K haplotype and 1000 Genomes Phase 3
160 reference panels⁶⁰. After quality control, 1 009 208 SNPs remained. UK Biobank
161 received ethical approval from the NHS National Research Ethics Service North West
162 (reference: 11/NW/0382). Further details on UKB cohort description, genotyping,
163 imputation and quality control are available elsewhere⁶¹⁻⁶³.

164 All participants provided informed consent.

165

166 **PHENOTYPE ASSESSMENT**

167 ***Stressful life events (SLE).*** GS participants reported SLE experienced over the
 168 preceding 6 months through a self-reported brief life events questionnaire based on
 169 the 12-item List of Threatening Experiences^{40,64,65} (Supplementary Table 1a). The total
 170 number of SLE reported (TSLE) consisted of the number of ‘yes’ responses. TSLE were
 171 subdivided into SLE potentially dependent or secondary to an individual’s own
 172 behaviour (DSLE, questions 6-11 in Supplementary Table 1a), and independent SLE
 173 (ISLE, questions 1-5 in Supplementary Table 1a; pregnancy item removed) following
 174 Brugha *et al.*^{40,41}. Thus, 3 SLE measures (TSLE, DSLE and ISLE) were constructed for GS.
 175 UKB participants were screened for “*illness, injury, bereavement and stress*”
 176 (Supplementary Table 1b) over the previous 2 years using 6 items included in the UKB
 177 Touchscreen questionnaire. A score reflecting SLE reported in UKB (TSLE_{UKB}) was
 178 constructed by summing the number of ‘yes’ responses.

179 ***Psychological assessment.*** GS participants reported whether their current mental
 180 state over the preceding 2 weeks differed from their typical state using a self-
 181 administered 28-item scaled version of The General Health Questionnaire (GHQ)⁶⁶⁻⁶⁸.
 182 Participants rated the degree and severity of their current symptoms with a four-
 183 point Likert scale (following Goldberg *et al.*, 1997⁶⁸). A final log-transformed GHQ was
 184 used to detect altered psychopathology and thus, assess depressive symptoms as
 185 results of SLE. In UKB participants, current depressive symptoms over the preceding 2
 186 weeks were evaluated using 4 psychometric screening items (Supplementary Table
 187 2), including two validated and reliable questions for screening depression⁶⁹, from the
 188 Patient Health Questionnaire (PHQ) validated to screen mental illness^{70,71}. Each
 189 question was rated in a four-point Likert scale to assess impairment/severity of

190 symptoms. Due to its skewed distribution, a four-point PHQ score was formed from
 191 PHQ (0 = 0; 1 = 1-2; 2 = 3-5; 3 = 6 or more) to create a more normal distribution.
 192 ***Stress-related traits.*** Targeted GS stress-related phenotypes and sample sizes are
 193 shown in Table 1 and detailed elsewhere⁵³. These conditions were selected from
 194 literature review based on previous evidence of a link with stress⁴⁶ (see also
 195 Supplementary Material: third section). Furthermore, we created additional
 196 independent samples using mapping by proxy, where individuals with a self-reported
 197 first-degree relative with a selected phenotype were included as proxy cases. This
 198 approach provides greater power to detect susceptibility variants in traits with low
 199 prevalence⁷².

200

201 **STATISTICAL ANALYSES**

202 ***SNP-heritability and genetic correlation.*** Restricted maximum likelihood approach
 203 was applied to estimate SNP-heritability (h^2_{SNP}) of depressive symptoms and self-
 204 reported SLE measures, and within samples bivariate genetic correlation between
 205 depressive symptoms and SLE measures using GCTA⁷³.

206 ***GWAS analyses.*** GWAS were conducted in PLINK⁷⁴. In GS, age, sex and 20 principal
 207 components (PCs) were fitted as covariates. In UKB, age, sex, and 15 PCs
 208 recommended by UKB were fitted as covariates. The genome-wide significance
 209 threshold was $p = 5 \times 10^{-8}$.

210 ***GWEIS analyses.*** GWEIS were conducted on GHQ (the dependent variable) for TSLE,
 211 DSLE and ISLE in GS and on PHQ for TSLE_{UKB} in UKB fitting the same covariates
 212 detailed above to reduce error variance. GWEIS were conducted using an R plugin for
 213 PLINK⁷⁴ developed by Almli *et al.*⁷⁵ (<https://epstein-software.github.io/robust-joint->

214 interaction). This method implements a robust test, that jointly considers SNP and
 215 SNP-environment interaction effects from a full model ($Y \sim \beta_0 + \beta_{SNP} + \beta_{SLE} +$
 216 $\beta_{SNP \times SLE} + \beta_{Covariates}$) against a null model where both the SNP and SNPxSLE
 217 effects equal 0, to assess the joint effect (the combined additive main and GxE
 218 genetic effect at a SNP) using a nonlinear statistical approach that applies Huber-
 219 White estimates of variance to correct possible inflation due to heteroscedasticity
 220 (unequal variances across exposure levels). This robust test should reduce
 221 confounding due to differences in variance induced by covariate interaction effects⁷⁶
 222 if present. Additional code was added (courtesy of Prof. Michael Epstein⁷⁵;
 223 Supplementary Material) to generate beta-coefficients and the p -value of the GxE
 224 term alone. In UKB, correcting for 1 009 208 SNPs and 1 exposure we established a
 225 Bonferroni-adjusted threshold for significance at $p = 2.47 \times 10^{-8}$ for both joint and GxE
 226 effects. In GS, correcting for 560 351 SNPs and 3 measures of SLE we established a
 227 genome-wide significance threshold of $p = 2.97 \times 10^{-8}$.

228 **Post-GWAS/GWEIS analyses.** GWAS and GWEIS summary statistics were analysed
 229 using FUMA⁷⁷ including: gene-based tests, functional annotation, gene prioritization
 230 and pathway enrichment (Supplementary Material).

231 **Polygenic profiling & prediction.** Polygenic risk scores (PRS) weighting by GxE effects
 232 ($PR_{S_{GxE}}$) were generated using PRSice-2⁷⁸ (Supplementary Material) in GS using GxE
 233 effects from UKB-GWEIS. In UKB, $PR_{S_{GxE}}$ were constructed using GxE effects from all
 234 three GS-GWEIS (TSLE, DSLE and ISLE as exposures) independently. PRS were also
 235 weighted in both samples using either UKB-GWAS or GS-GWAS statistics (PR_{S_D}), and
 236 summary statistics from Psychiatric Genetic Consortium (PGC) MDD-GWAS (released
 237 2016; $PR_{S_{MDD}}$) that excluded GS and UKB individuals when required ($N_{noGS} = 155\,866$;

238 $N_{\text{noUKB}} = 138\,884$). Furthermore, we calculated PRS weighted by the joint effects (the
 239 combined additive main and GxE genetic effects; $\text{PRS}_{\text{Joint}}$) from either the UKB-GWEIS
 240 or GS-GWEIS. PRS predictions of depressive symptoms were permuted 10 000 times.
 241 Multiple regression models fitting PRS_{GxE} and PRS_{MDD} , and both PRS_{GxE} and PRS_{D} were
 242 tested. All models were adjusted by same covariates used in GWAS/GWEIS. Null
 243 models were estimated from the direct effects of covariates alone. The predictive
 244 improvement of combining PRS_{GxE} and $\text{PRS}_{\text{MDD}}/\text{PRS}_{\text{D}}$ effects over $\text{PRS}_{\text{MDD}}/\text{PRS}_{\text{D}}$ effect
 245 alone was tested for significance using the likelihood-ratio test (LRT).
 246 Prediction of PRS_{D} , PRS_{GxE} and $\text{PRS}_{\text{Joint}}$ on stress-linked traits were adjusted by age, sex
 247 and 20 PCs; and permuted 10 000 times. Empirical- p -values after permutations were
 248 further adjusted by false discovery rate (conservative threshold at *Empirical- p* =
 249 6.16×10^{-3}). The predictive improvement of fitting PRS_{GxE} combined with PRS_{D} and
 250 covariates over prediction of a phenotype using the PRS_{D} effect alone with covariates
 251 was assessed using LRT, and *LRT- p* -values adjusted by FDR (conservative threshold at
 252 *LRT- p* = 8.35×10^{-4}).
 253

254 RESULTS

255 **Phenotypic and genetic correlations.** Depressive symptoms scores and SLE measures
256 were positively correlated in both UKB ($r^2 = 0.22$, $p < 2.2 \times 10^{-16}$) and GS (TSLE- $r^2 = 0.21$,
257 $p = 1.69 \times 10^{-52}$; DSLE- $r^2 = 0.21$, $p = 8.59 \times 10^{-51}$; ISLE- $r^2 = 0.17$, $p = 2.33 \times 10^{-33}$). Significant
258 bivariate genetic correlation between depression and SLE scores was identified in
259 UKB ($r_G = 0.72$; $p < 1 \times 10^{-5}$, $N = 50\,000$), but not in GS ($r_G = 1$, $p \geq 0.056$, $N = 4\,919$;
260 Supplementary Table 3a).

261 **SNP-heritability (h^2_{SNP}).** In UKB, a significant h^2_{SNP} of PHQ was identified ($h^2_{\text{SNP}} =$
262 0.090 ; $p < 0.001$; $N = 99\,057$). This estimate remained significant after adjusting by
263 TSLE_{UKB} effect ($h^2_{\text{SNP}} = 0.079$; $p < 0.001$), suggesting a genetic contribution unique of
264 depressive symptoms. The h^2_{SNP} of TSLE_{UKB} was also significant ($h^2_{\text{SNP}} = 0.040$, $p <$
265 0.001 ; Supplementary Table 3b). In GS, h^2_{SNP} was not significant for GHQ ($h^2_{\text{SNP}} =$
266 0.071 , $p = 0.165$; $N = 4\,919$). However, in an ad hoc estimation from the baseline
267 sample of 6 751 unrelated GS participants (details in Supplementary Table 3b) we
268 detected a significant h^2_{SNP} for GHQ ($h^2_{\text{SNP}} = 0.135$; $p < 5.15 \times 10^{-3}$), suggesting that the
269 power to estimate h^2_{SNP} in GS may be limited by sample size. Estimates were not
270 significant for neither TSLE ($h^2_{\text{SNP}} = 0.061$, $p = 0.189$; Supplementary Table 3b) nor ISLE
271 ($h^2_{\text{SNP}} = 0.000$, $p = 0.5$), but h^2_{SNP} was significant for DSLE ($h^2_{\text{SNP}} = 0.131$, $p = 0.029$),
272 supporting a potential genetic mediation and gene-environment correlation.

273 **GWAS of depressive symptoms.** No genome-wide significant SNPs were detected by
274 GWAS in either cohort. Top findings ($p < 1 \times 10^{-5}$) are summarized in Supplementary
275 Table 4. Manhattan and QQ plots are shown in Supplementary Figures 1-4. There was
276 no evidence of genomic inflation (all $\lambda_{1000} < 1.01$).

277 **Post-GWAS analyses.** Gene-based test identified six genes associated with PHQ using
 278 UKB-GWAS statistics at genome-wide significance (Bonferroni-corrected $p = 2.77 \times 10^{-6}$;
 279 DCC , $p = 7.53 \times 10^{-8}$; $ACSS3$, $p = 6.51 \times 10^{-7}$; $DRD2$, $p = 6.55 \times 10^{-7}$; $STAG1$, $p = 1.63 \times 10^{-6}$;
 280 $FOXP2$, $p = 2.09 \times 10^{-6}$; $KYNU$, $p = 2.24 \times 10^{-6}$; Supplementary Figure 8). Prioritized genes
 281 based on position, eQTL and chromatin interaction mapping are detailed in
 282 Supplementary Table 5. No genes were detected in GS-GWAS gene-based test
 283 (Supplementary Figures 9). No tissue enrichment was detected from GWAS in either
 284 cohort. Significant gene-sets and GWAS catalog associations for UKB-GWAS are
 285 reported in Supplementary Table 6. These included the *biological process*: positive
 286 regulation of long term synaptic potentiation, and *GWAS catalog associations*: brain
 287 structure, schizophrenia, response to amphetamines, age-related cataracts (age at
 288 onset), fibrinogen, acne (severe), fibrinogen levels, and educational attainment; all
 289 adjusted- $p < 0.01$. There was no significant gene-set enrichment from GS-GWAS.

290 **GWEIS of depressive symptoms.** Manhattan and QQ plots are shown in
 291 Supplementary Figures 1-4. There was no evidence of GWEIS inflation for either UKB
 292 or GS (all $\lambda_{1000} < 1.01$). No genome-wide significant GWEIS associations were detected
 293 for SLE in UKB. GS-GWEIS using TSLE identified a significant GxE effect ($p < 2.97 \times 10^{-8}$)
 294 at an intragenic SNP on chromosome 11 (rs12789145, $p = 4.95 \times 10^{-9}$, $\beta = 0.06$, closest
 295 gene: *PIWIL4*; Supplementary Figure 5), and using DSLE at an intronic SNP in *ZCCHC2*
 296 on chromosome 18 (rs17070072, $p = 1.46 \times 10^{-8}$, $\beta = -0.08$; Supplementary Figure 6). In
 297 their corresponding joint effect tests both rs12789145 ($p = 2.77 \times 10^{-8}$) and rs17070072
 298 ($p = 1.96 \times 10^{-8}$) were significant. GWEIS for joint effect using DSLE identified two
 299 further significant SNPs in on chromosome 9 (rs12000047, $p = 2.00 \times 10^{-8}$, $\beta = -0.23$;
 300 rs12005200, $p = 2.09 \times 10^{-8}$, $\beta = -0.23$, LD $r^2 > 0.8$, closest gene: *CYLC2*; Supplementary

Figure 7). None of these associations replicated in UKB ($p > 0.05$), although the effect direction was consistent between cohorts for the SNP close to *PIWL1* and SNPs at *CYLC2*. No SNP achieved genome-wide significant association in GS-GWEIS using ISLE as exposure. Top GWEIS results ($p < 1 \times 10^{-5}$) are summarized in Supplementary Tables 7-10.

Post-GWEIS analyses: gene-based tests. All results are shown in Supplementary Figures 10-17. Two genes were associated with PHQ using the joint effect from UKB-GWEIS (*ACSS3* $p = 1.61 \times 10^{-6}$; *PHF2*, $p = 2.28 \times 10^{-6}$; Supplementary Figure 11). *ACSS3* was previously identified using the additive main effects, whereas *PHF2* was only significantly associated using the joint effects. Gene-based tests identified *MTNR1B* as significantly associated with GHQ on GS-GWEIS using DSLE in both GxE ($p = 1.53 \times 10^{-6}$) and joint effects ($p = 2.38 \times 10^{-6}$; Supplementary Figures 14-15).

Post-GWEIS analyses: tissue enrichment. We prioritized genes based on position, eQTL and chromatin interaction mapping in brain tissues and regions. In UKB, prioritized genes with GxE effect were enriched for up-regulated differentially expressed genes from adrenal gland (adjusted- $p = 3.58 \times 10^{-2}$). Using joint effects, prioritized genes were enriched on up-regulated differentially expressed genes from artery tibial (adjusted- $p = 4.34 \times 10^{-2}$). In GS, prioritized genes were enriched: in up-regulated differentially expressed genes from artery coronary (adjusted- $p = 4.55 \times 10^{-2}$) using GxE effects with DSLE; in down-regulated differentially expressed genes from artery aorta tissue (adjusted- $p = 4.71 \times 10^{-2}$) using GxE effects with ISLE; in up-regulated differentially expressed genes from artery coronary (adjusted- $p = 5.97 \times 10^{-3}$, adjusted- $p = 9.57 \times 10^{-3}$) and artery tibial (adjusted- $p = 1.05 \times 10^{-2}$, adjusted- $p = 1.55 \times 10^{-2}$) tissues using joint effects with both TSLE and DSLE; and in down-regulated

325 differentially expressed genes from lung tissue (adjusted- $p = 3.98 \times 10^{-2}$) and in up- and
326 down-regulated differentially expressed genes from the spleen (adjusted- $p = 4.71 \times 10^{-2}$)
327 ²) using joint effects with ISLE. There was no enrichment using GxE effect with TSLE.

328 **Post-GWEIS analyses: gene-sets enrichment.** Significant gene-sets and GWAS catalog
329 hits from GWEIS are detailed in Supplementary Tables 11-14, including for UKB
330 *Biocarta*: GPCR pathway; *Reactome*: opioid signalling, neurotransmitter receptor
331 binding and downstream transmission in the postsynaptic cell, transmission across
332 chemical synapses, gastrin CREB signalling pathway via PKC and MAPK; *GWAS*
333 *catalog*: post bronchodilator FEV1/FVC ratio, migraine and body mass index. In GS,
334 enrichment was seen using TSLE and DLSE for *GWAS catalog*: age-related macular
335 degeneration, myopia, urate levels and Heschl's gyrus morphology; and using ISLE for
336 *biological process*: regulation of transporter activity. All adjusted- $p < 0.01$.

337 **Cross-cohort prediction.** In GS, PRS_D weighted by UKB-GWAS of PHQ significantly
338 explained 0.56% of GHQ variance (*Empirical-p* $< 1.10^{-4}$), similar to PRS_{MDD} weighted by
339 PGC MDD-GWAS ($R^2 = 0.78\%$, *Empirical-p* $< 1.10^{-4}$). PRS_{GxE} weighted by UKB-GWEIS
340 GxE effects explained 0.15% of GHQ variance (*Empirical-p* = 0.03, Supplementary
341 Table 15). PRS_{GxE} fitted jointly with PRS_{MDD} significantly improved prediction of GHQ
342 ($R^2 = 0.93\%$, model $p = 6.12 \times 10^{-11}$; predictive improvement of 19%, *LRT-p* $= 5.91 \times 10^{-3}$)
343 compared to PRS_{MDD} alone. Similar to PRS_{GxE} with PRS_D ($R^2 = 0.69\%$, model $p =$
344 2.72×10^{-8} ; predictive improvement of 23%, *LRT-p* = 0.01). PRS_{Joint} weighted by UKB-
345 GWEIS also predicted GHQ ($R^2 = 0.58\%$, *Empirical-p* $< 1.10^{-4}$), although the variance
346 explained was significantly reduced compared to the model fitting PRS_{GxE} and PRS_D
347 together (*LRT-p* $= 4.69 \times 10^{-7}$), suggesting that additive and GxE effects should be
348 modelled independently for polygenic approaches (Figure 2a).

In UKB (Figure 2b), both PRS_D weighted by GS-GWAS of GHQ and PRS_{MDD} significantly explained 0.04% and 0.45% of PHQ variance, respectively (both *Empirical-p* < 1.10^{-4} ; Supplementary Table 15). PRS_{GxE} derived from GS-GWEIS GxE effect did not significantly predicted PHQ (TSLE- PRS_{GxE} *Empirical-p* = 0.382; DSLE- PRS_{GxE} *Empirical-p* = 0.642; ISLE- PRS_{GxE} *Empirical-p* = 0.748). Predictive improvements by PRS_{GxE} effect fitted jointly with PRS_{MDD} or PRS_D were not significant (all *LRT-p* > 0.08). PRS_{Joint} significantly predicted PHQ (TSLE- PRS_{Joint} : R^2 = 0.032%, *Empirical-p* < 1.10^{-4} ; DSLE- PRS_{Joint} : R^2 = 0.012%, *Empirical-p* = 4.3×10^{-3} ; ISLE- PRS_{Joint} : R^2 = 0.032%, *Empirical-p* < 1.10^{-4}), although the variances explained were significantly reduced compared to the models fitting PRS_{GxE} and PRS_D together (all *LRT-p* < 1.48×10^{-3}).

Prediction of stress-related traits. Prediction of stress-related traits in independent samples using PRS_D , PRS_{GxE} and PRS_{Joint} are summarized in Figure 3a and Supplementary Table 16. Significant trait prediction after FDR adjustment (*Empirical-p* < 6.16×10^{-3} , FDR-adjusted *Empirical-p* < 0.05) using both UKB and GS PRS_D was seen for: depression status, neuroticism and schizotypal personality. PRS_{GxE} weighted by GS-GWEIS GxE effect using ISLE significantly predicted depression status mapping by proxy (*Empirical-p* = 7.00×10^{-4} , FDR-adjusted *Empirical-p* = 9.54×10^{-3}).

Nominally significant predictive improvements (*LRT-p* < 0.05) of fitting PRS_{GxE} over the PRS_D effect alone using summary statistics generated from both UKB and GS were detected for schizotypal personality, heart diseases and COPD by proxy (Figure 3b). PRS_{GxE} weighted by GS-GWEIS GxE effect using ISLE significantly improved prediction over PRS_D effect alone of depression status mapping by proxy after FDR adjustment (*LRT-p* = 1.96×10^{-4} , FDR-adjusted *LRT-p* = 2.35×10^{-2}).

372

373 DISCUSSION

374 This study performs GWAS and incorporates data on recent adult stressful life events
375 (SLE) into GWEIS of depressive symptoms, identifies new loci and candidate genes for
376 the modulation of genetic response to SLE; and provides insights to help disentangle
377 the underlying aetiological mechanisms increasing genetic liability through SLE to
378 both depressive symptoms and stress-related traits.

379 SNP-heritability of depressive symptoms ($h^2_{\text{SNP}} = 9\text{-}13\%$), were slightly higher than
380 estimates from African American populations³⁴, and over a third larger than estimates
381 in MDD from European samples⁷⁹. h^2_{SNP} for PHQ in UKB (9.0%) remained significant
382 after adjusting for SLE (7.9%). Thus, although some genetic contributions may be
383 partially shared between depressive symptoms and reporting of SLE, there is still a
384 relatively large genetic contribution unique to depressive symptoms. Significant h^2_{SNP}
385 of DSLE in GS (13%) and TSLE_{UKB} in UKB (4%), which is mainly composed of dependent
386 SLE items, were detected similar to previous studies (8% and 29%)^{34,43}. Conversely,
387 there was no evidence for heritability of independent SLE. A significant bivariate
388 genetic correlation between depressive symptoms and SLE ($r_G = 0.72$) was detected
389 in UKB after adjusting for covariates, suggesting that there are shared common
390 variants underlying self-reported depressive symptoms and SLE. This bivariate genetic
391 correlation was smaller than that estimated from African American populations ($r_G =$
392 0.97 ; $p = 0.04$; $N = 7\,179$)³⁴. Genetic correlations between SLE measures and GHQ
393 were not significant in GS ($N = 4\,919$; $r_G = 1$; all $p > 0.056$), perhaps due to a lack of
394 power in this smaller sample.

395 Post-GWAS gene-based tests detected six genes significantly associated with PHQ
396 (*DCC*, *ACSS3*, *DRD2*, *STAG1*, *FOXP2* and *KYNU*). Previous studies have implicated these

genes in liability to depression (see Supplementary Table 17), and three of them are genome-wide significant in gene-based tests from the latest meta-analysis of major depression that includes UKB (*DCC*, $p = 2.57 \times 10^{-14}$; *DRD2*, $p = 5.35 \times 10^{-14}$; and *KYNU*, $p = 2.38 \times 10^{-6}$; $N = 807\,553$)⁸⁰. This supports the implementation of quantitative measures such as PHQ to detect genes underlying lifetime depression status⁸¹. For example, significant gene ontology analysis of the UKB-GWAS identified enrichment for positive regulation of long-term synaptic potentiation, and for previous GWAS findings of brain structure⁸², schizophrenia⁸³ and response to amphetamines⁸⁴.

The key element of this study was to conduct GWEIS of depressive symptoms and recent SLE. We identified two loci with significant GxE effect in GS. However, none of these associations replicated in UKB ($p > 0.05$). The strongest association was using TSLE at 53kb down-stream of *PIWIL4* (rs12789145). *PIWIL4* is brain-expressed and involved in chromatin-modification⁸⁵, suggesting it may moderate the effects of stress on depression. It encodes HIWI2, a protein thought to regulate OTX2, which is critical for the development of forebrain and for coordinating critical periods of plasticity disrupting the integration of cortical circuits^{86,87}. Indeed, an intronic SNP in *PIWIL4* was identified as the strongest GxE signal in ADHD using mother's warmth as environmental exposure⁸⁸. The other significant GxE identified in our study was in *ZCCHC2* using DSLE. This zinc finger protein is expressed in blood CD4+ T-cells and is down regulated in individuals with MDD⁸⁹ and in those resistant to treatment with citalopram⁹⁰. No GxE effect was seen using ISLE as exposure.

No significant locus or gene with GxE effect was detected in UKB-GWEIS. Nevertheless, joint effects (combined additive main and GxE genetic effects) identified two genes significantly associated with PHQ (*ACSS3* and *PHF2*; see

Supplementary Table 17). *PHF2* was recently detected as genome-wide significant at the latest meta-analysis of depression⁸⁰. Notably, *PHF2* paralogs have already been link with MDD through stress-response in three other studies⁹¹⁻⁹³. Joint effects in GS also detected an additional significant association upstream *CYLC2*, a gene nominally associated ($p < 1 \times 10^{-5}$) with obsessive-compulsive disorder and Tourette's Syndrome⁹⁴. Gene-based test from GS-GWEIS identified a significant association with *MTNR1B*, a melatonin receptor gene, using DSLE (both GxE and joint effect; Supplementary Table 17). Prioritized genes using GxE effects were enriched in differentially expressed genes from several tissues including the adrenal gland, which releases cortisol into the blood-stream in response to stress, thus playing a key role in the stress-response system, reinforcing a potential role of GxE in stress-related conditions.

The different instruments and sample sizes available make it hard to compare results between cohorts. Whereas GS contains deeper phenotyping measurements of stress and depressive symptoms than UKB, the sample size is much smaller, which may be reflected in the statistical power required to detect reliable GxE effects. Furthermore, the presence and size of GxE are dependent on their parameterization (i.e. the measurement, scale and distribution of the instruments used to test such interaction)⁹⁵. Thus, GxE may be incomparable across GWEIS due to differences in both phenotype assessment and stressors tested. Although our results suggest that both depressive symptom measures are correlated with lifetime depression status, different influences on depressive symptoms from the SLE covered across studies may contribute to lack of stronger replication. Instruments in GS cover a wider range of SLE and are more likely to capture changes in depressive symptoms as

consequence of their short-term effects. Conversely, UKB could capture more marked long-term effects, as SLE were captured over 2 years compared to 6 months in GS. New mental health questionnaires covering a wide range of psychiatric symptoms and SLE in the last release of UKB data provides the opportunity to create more similar measures to GS in the near future. Further replication in independent studies with equivalent instruments is required to validate our GWEIS findings. Despite these limitations and a lack of overlap in the individual genes prioritised from the two GWEIS, replication was seen in the predictive improvement of using PRS_{GxE} derived from the GWEIS GxE effects to predict stress-related phenotypes.

The third aim of this study was to test whether GxE effect could improve predictive genetic models, and thus help to explain deviation from additive models and missing heritability of MDD⁹⁶. Multiple regression models suggested that inclusion of PRS_{GxE} weighted by GxE effects could improve prediction of an individual's depressive symptoms over use of PRS_{MDD} or PRS_D weighted by additive effects alone. In GS, we detected a predictive gain of 19% over PRS_{MDD} weighted by PGC MDD-GWAS, and a gain of 23% over PRS_D weighted by UKB-GWAS (Figure 2a). However, these findings did not surpass stringent Bonferroni-correction and could not be validated in UKB. This may reflect in the poor predictive power of the PRS generated from the much smaller GS discovery sample. The results show a noticeably reduced prediction using PRS_{Joint} weighted by joint effects, which suggests that the genetic architecture of stress-response is at least partially independent and differs from genetic additive main effects. Therefore, our results from multiple regression models suggest that for polygenic approaches main and GxE effects should be modelled independently.

SLE effects are not limited to mental illness⁴⁶.

Our final aim was to investigate shared aetiology between GxE for depressive symptoms and stress-related traits. Despite the differences between the respondents and non-respondents (Table 1 legend), a significant improvement was seen predicting depressive status mapping by proxy cases using GxE effect from GS-GWEIS with independent SLE (*FDR-adjusted LRT-p* = 0.013), but not with dependent SLE. GxE effects using statistics generated from both discovery samples, despite the differences in measures, nominally improved the phenotypic prediction of schizotypal personality, heart disease and the proxy of COPD (*LRT-p* < 0.05). Other studies have found evidence supporting a link between stress and depression in these phenotypes that support our results (see Supplementary Material for extended review) and suggest, for instance, potential pleiotropy between schizotypal personality and stress-response. Our findings point to a potential genetic component underlying a stress-response-depression-comorbidities link due, at least in part, to shared stress-response mechanisms. A relationship between SLE, depression and coping strategies such as smoking suggests that perhaps, genetic stress-response may modulate adaptive behaviours such as smoking, fatty diet intake, alcohol consumption and substance abuse. This is discussed further in the Supplementary Material.

In this study, evidence for SNPs with significant GxE effects came primarily from the analyses of dependent SLE and not from independent SLE. This supports a genetic effect on probability of exposure to, or reporting of SLE, endorsing a gene-environment correlation. Chronic stress may influence cognition, decision-making and behaviour eventually leading to higher risk-taking⁹⁷. These conditions may also increase sensitivity to stress amongst vulnerable individuals, including those with depression, who also have a higher propensity to report SLE, particularly dependent

493 SLE³⁹. A potential reporting bias in dependent SLE may be mediated as well by
 494 heritable behavioural, anxiety or psychological traits such as risk-taking^{43,98}.
 495 Furthermore, individuals vulnerable to MDD may expose themselves into
 496 environments of higher risk and stress¹⁴. This complex interplay, reflected in the form
 497 of a gene-environment correlation effect, would hinder the interpretation of GxE
 498 effects from GWEIS as pure interactions. A mediation of associations between SLE
 499 and depressive symptoms through genetically driven sensitivity to stress, personality
 500 or behavioural traits would support the possibility of subtle genotype-by-genotype
 501 (GxG) interactions, or genotype-by-genotype-by-environment (GxGxE) interactions
 502 contributing to depression^{99,100}. In contrast, PRS prediction of the stress-related traits:
 503 schizotypal personality, heart disease and COPD, was primarily from derived weights
 504 using independent SLE, suggesting that a common set of variants moderate the
 505 effects of SLE across stress-related traits and that larger sample sizes will be required
 506 to detect the individual SNPs contributing to this. Thus, our finding supports the
 507 inclusion of environmental information into GWAS to enhance the detection of
 508 relevant genes. Results of studying dependent and independent SLE support a
 509 contribution of genetically mediated exposure to and/or reporting of SLE, perhaps
 510 through sensitivity to stress as mediator.
 511 This study emphasises the relevance of GxE in depression and human health in
 512 general and provides the basis for future lines of research.

513

514 **ACKNOWLEDGMENTS**

515 See Supplementary Material: acknowledgments.

516

517 **FINANCIAL DISCLOSURE**

518 The authors declare no conflict of interest.

519

520 Supplementary Material is available at Translational Psychiatry's website.

521

522 REFERENCES

- 523 1. Plomin, R., Owen, M. & McGuffin, P. The genetic basis of complex human
524 behaviors. *Science* **264**, 1733-1739 (1994).
- 525 2. Kendler, K.S. & Eaves, L.J. Models for the joint effect of genotype and
526 environment on liability to psychiatric illness. *Am J Psychiatry* **143**, 279-89
527 (1986).
- 528 3. Kendler, K.S. & Baker, J.H. Genetic influences on measures of the
529 environment: a systematic review. *Psychol Med* **37**, 615-26 (2007).
- 530 4. Colodro-Conde, L. *et al.* A direct test of the diathesis-stress model for
531 depression. *Mol Psychiatry* (2017).
- 532 5. Luciano, M. *et al.* Shared genetic aetiology between cognitive ability and
533 cardiovascular disease risk factors: Generation Scotland's Scottish family
534 health study. *Intelligence* **38**, 304-313 (2010).
- 535 6. Kessler, R.C. & Bromet, E.J. The epidemiology of depression across
536 cultures. *Annu Rev Public Health* **34**, 119-38 (2013).
- 537 7. Sullivan, P.F., Neale, M.C. & Kendler, K.S. Genetic epidemiology of major
538 depression: review and meta-analysis. *Am J Psychiatry* **157**, 1552-62
539 (2000).
- 540 8. Hammen, C. Stress and depression. *Annu Rev Clin Psychol* **1**, 293-319
541 (2005).
- 542 9. Kessler, R.C. The effects of stressful life events on depression. *Annu Rev*
543 *Psychol* **48**, 191-214 (1997).
- 544 10. Kendler, K.S., Karkowski, L.M. & Prescott, C.A. Causal relationship between
545 stressful life events and the onset of major depression. *Am J Psychiatry*
546 **156**, 837-41 (1999).
- 547 11. Stroud, C.B., Davila, J. & Moyer, A. The relationship between stress and
548 depression in first onsets versus recurrences: a meta-analytic review. *J*
549 *Abnorm Psychol* **117**, 206-13 (2008).
- 550 12. Kendler, K.S., Karkowski, L.M. & Prescott, C.A. Stressful life events and
551 major depression: risk period, long-term contextual threat, and diagnostic
552 specificity. *J Nerv Ment Dis* **186**, 661-9 (1998).
- 553 13. Silberg, J., Rutter, M., Neale, M. & Eaves, L. Genetic moderation of
554 environmental risk for depression and anxiety in adolescent girls. *Br J*
555 *Psychiatry* **179**, 116-21 (2001).
- 556 14. Kendler, K.S. *et al.* Stressful life events, genetic liability, and onset of an
557 episode of major depression in women. *Am J Psychiatry* **152**, 833-42
558 (1995).
- 559 15. Arnau-Soler, A., Adams, M., Hayward, C. & Thomson, P. Genome-wide
560 interaction study of a proxy for stress-sensitivity and its prediction of
561 major depressive disorder. *bioRxiv* (2018).
- 562 16. Arnau-Soler, A. *et al.* A validation of the diathesis-stress model for
563 depression in Generation Scotland. *bioRxiv* (2018).
- 564 17. Garantziotis, S. & Schwartz, D.A. Ecogenomics of respiratory diseases of
565 public health significance. *Annu Rev Public Health* **31**, 37-51 1 p following
566 51 (2010).
- 567 18. Aschard, H. *et al.* Evidence for large-scale gene-by-smoking interaction
568 effects on pulmonary function. *Int J Epidemiol* **46**, 894-904 (2017).

- 569 19. Molfino, N.A. & Coyle, A.J. Gene-environment interactions in chronic
570 obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis* **3**, 491-7
571 (2008).
- 572 20. Polonikov, A.V., Ivanov, V.P. & Solodilova, M.A. Genetic variation of genes
573 for xenobiotic-metabolizing enzymes and risk of bronchial asthma: the
574 importance of gene-gene and gene-environment interactions for disease
575 susceptibility. *J Hum Genet* **54**, 440-9 (2009).
- 576 21. Haiman, C.A. *et al.* Ethnic and racial differences in the smoking-related risk
577 of lung cancer. *N Engl J Med* **354**, 333-42 (2006).
- 578 22. Han, J., Hankinson, S.E., Colditz, G.A. & Hunter, D.J. Genetic variation in
579 XRCC1, sun exposure, and risk of skin cancer. *Br J Cancer* **91**, 1604-9
580 (2004).
- 581 23. Manning, A.K. *et al.* A genome-wide approach accounting for body mass
582 index identifies genetic variants influencing fasting glycemic traits and
583 insulin resistance. *Nat Genet* **44**, 659-69 (2012).
- 584 24. Wang, L., Murk, W. & DeWan, A.T. Genome-Wide Gene by Environment
585 Interaction Analysis Identifies Common SNPs at 17q21.2 that Are
586 Associated with Increased Body Mass Index Only among Asthmatics. *PLoS*
587 *One* **10**, e0144114 (2015).
- 588 25. Siegert, S. *et al.* Genome-wide investigation of gene-environment
589 interactions in colorectal cancer. *Hum Genet* **132**, 219-31 (2013).
- 590 26. Gong, J. *et al.* Genome-Wide Interaction Analyses between Genetic Variants
591 and Alcohol Consumption and Smoking for Risk of Colorectal Cancer. *PLoS*
592 *Genet* **12**, e1006296 (2016).
- 593 27. Polfus, L.M. *et al.* Genome-wide association study of gene by smoking
594 interactions in coronary artery calcification. *PLoS One* **8**, e74642 (2013).
- 595 28. Duncan, L.E. & Keller, M.C. A critical review of the first 10 years of
596 candidate gene-by-environment interaction research in psychiatry. *Am J*
597 *Psychiatry* **168**, 1041-9 (2011).
- 598 29. Risch, N. *et al.* Interaction between the serotonin transporter gene (5-
599 HTTLPR), stressful life events, and risk of depression: a meta-analysis.
600 *JAMA* **301**, 2462-71 (2009).
- 601 30. Karg, K., Burmeister, M., Shedden, K. & Sen, S. The serotonin transporter
602 promoter variant (5-HTTLPR), stress, and depression meta-analysis
603 revisited: evidence of genetic moderation. *Arch Gen Psychiatry* **68**, 444-54
604 (2011).
- 605 31. Bleys, D., Luyten, P., Soenens, B. & Claes, S. Gene-environment interactions
606 between stress and 5-HTTLPR in depression: A meta-analytic update.
607 *Journal of Affective Disorders* **226**, 339-345 (2018).
- 608 32. Peyrot, W.J. *et al.* Does Childhood Trauma Moderate Polygenic Risk for
609 Depression? A Meta-analysis of 5765 Subjects From the Psychiatric
610 Genomics Consortium. *Biol Psychiatry* (2017).
- 611 33. Kraft, P., Yen, Y.C., Stram, D.O., Morrison, J. & Gauderman, W.J. Exploiting
612 gene-environment interaction to detect genetic associations. *Hum Hered*
613 **63**, 111-9 (2007).
- 614 34. Dunn, E.C. *et al.* Genome-Wide Association Study (Gwas) and Genome-
615 Wide by Environment Interaction Study (Gweis) of Depressive Symptoms
616 in African American and Hispanic/Latina Women. *Depress Anxiety* **33**, 265-
617 80 (2016).

- 618 35. Otowa, T. *et al.* The First Pilot Genome-Wide Gene-Environment Study of
619 Depression in the Japanese Population. *PLoS One* **11**, e0160823 (2016).
- 620 36. Ikeda, M. *et al.* Genome-wide environment interaction between depressive
621 state and stressful life events. *J Clin Psychiatry* **77**, e29-30 (2016).
- 622 37. Coleman, J.R.I., Eley, T.C. & Breen, G. Genome-wide gene-environment
623 analyses of depression and reported lifetime traumatic experiences in UK
624 Biobank. *bioRxiv* (2018).
- 625 38. Plomin, R., DeFries, J.C. & Loehlin, J.C. Genotype-environment interaction
626 and correlation in the analysis of human behavior. *Psychol Bull* **84**, 309-22
627 (1977).
- 628 39. Clarke, T. *et al.* Genetic and environmental determinants of stressful life
629 events and their overlap with depression and neuroticism [version 1;
630 referees: 3 approved with reservations]. *Wellcome Open Research* **3**(2018).
- 631 40. Brugha, T., Bebbington, P., Tennant, C. & Hurry, J. The List of Threatening
632 Experiences: a subset of 12 life event categories with considerable long-
633 term contextual threat. *Psychol Med* **15**, 189-94 (1985).
- 634 41. Kendler, K.S., Karkowski, L.M. & Prescott, C.A. The assessment of
635 dependence in the study of stressful life events: validation using a twin
636 design. *Psychol Med* **29**, 1455-60 (1999).
- 637 42. Plomin, R., Lichtenstein, P., Pedersen, N.L., McClearn, G.E. & Nesselroade,
638 J.R. Genetic influence on life events during the last half of the life span.
639 *Psychol Aging* **5**, 25-30 (1990).
- 640 43. Power, R.A. *et al.* Estimating the heritability of reporting stressful life
641 events captured by common genetic variants. *Psychol Med* **43**, 1965-71
642 (2013).
- 643 44. Bemmels, H.R., Burt, S.A., Legrand, L.N., Iacono, W.G. & McGue, M. The
644 heritability of life events: an adolescent twin and adoption study. *Twin Res*
645 *Hum Genet* **11**, 257-65 (2008).
- 646 45. Boardman, J.D., Alexander, K.B. & Stallings, M.C. Stressful life events and
647 depression among adolescent twin pairs. *Biodemography Soc Biol* **57**, 53-
648 66 (2011).
- 649 46. Salleh, M.R. Life event, stress and illness. *Malays J Med Sci* **15**, 9-18 (2008).
- 650 47. Thaipisuttikul, P., Ittasakul, P., Waleeprakhon, P., Wisajun, P. & Jullagate, S.
651 Psychiatric comorbidities in patients with major depressive disorder.
652 *Neuropsychiatr Dis Treat* **10**, 2097-103 (2014).
- 653 48. Moussavi, S. *et al.* Depression, chronic diseases, and decrements in health:
654 results from the World Health Surveys. *Lancet* **370**, 851-8 (2007).
- 655 49. Topic, R. *et al.* Somatic comorbidity, metabolic syndrome, cardiovascular
656 risk, and CRP in patients with recurrent depressive disorders. *Croat Med J*
657 **54**, 453-9 (2013).
- 658 50. Lloyd, C.E., Roy, T., Nouwen, A. & Chauhan, A.M. Epidemiology of
659 depression in diabetes: international and cross-cultural issues. *J Affect*
660 *Disord* **142 Suppl**, S22-9 (2012).
- 661 51. Ohayon, M.M. & Schatzberg, A.F. Using chronic pain to predict depressive
662 morbidity in the general population. *Arch Gen Psychiatry* **60**, 39-47 (2003).
- 663 52. Slavich, G.M. & Irwin, M.R. From stress to inflammation and major
664 depressive disorder: a social signal transduction theory of depression.
665 *Psychol Bull* **140**, 774-815 (2014).

- 666 53. Smith, B.H. *et al.* Cohort Profile: Generation Scotland: Scottish Family
667 Health Study (GS:SFHS). The study, its participants and their potential for
668 genetic research on health and illness. *Int J Epidemiol* **42**, 689-700 (2013).
- 669 54. Gunderson, K.L. Whole-Genome Genotyping on Bead Arrays. in *DNA*
670 *Microarrays for Biomedical Research: Methods and Protocols* (ed. Dufva, M.)
671 197-213 (Humana Press, Totowa, NJ, 2009).
- 672 55. Kerr, S.M. *et al.* Pedigree and genotyping quality analyses of over 10,000
673 DNA samples from the Generation Scotland: Scottish Family Health Study.
674 *BMC Med Genet* **14**, 38 (2013).
- 675 56. Nagy, R. *et al.* Exploration of haplotype research consortium imputation
676 for genome-wide association studies in 20,032 Generation Scotland
677 participants. *Genome Med* **9**, 23 (2017).
- 678 57. Navrady, L.B. *et al.* Cohort Profile: Stratifying Resilience and Depression
679 Longitudinally (STRADL): a questionnaire follow-up of Generation
680 Scotland: Scottish Family Health Study (GS:SFHS). *Int J Epidemiol* (2017).
- 681 58. Smith, B.H. *et al.* Generation Scotland: the Scottish Family Health Study; a
682 new resource for researching genes and heritability. *BMC Med Genet* **7**, 74
683 (2006).
- 684 59. Fernandez-Pujals, A.M. *et al.* Epidemiology and Heritability of Major
685 Depressive Disorder, Stratified by Age of Onset, Sex, and Illness Course in
686 Generation Scotland: Scottish Family Health Study (GS:SFHS). *PLoS One* **10**,
687 e0142197 (2015).
- 688 60. Huang, J. *et al.* Improved imputation of low-frequency and rare variants
689 using the UK10K haplotype reference panel. *Nat Commun* **6**, 8111 (2015).
- 690 61. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the
691 causes of a wide range of complex diseases of middle and old age. *PLoS*
692 *Med* **12**, e1001779 (2015).
- 693 62. Biobank, U. UK Biobank: protocol for a large-scale prospective
694 epidemiological resource. (2010).
- 695 63. Biobank, U. UK Biobank ethics and governance framework, version 3.0
696 (2007).
- 697 64. Brugha, T.S. & Cragg, D. The List of Threatening Experiences: the reliability
698 and validity of a brief life events questionnaire. *Acta Psychiatr Scand* **82**,
699 77-81 (1990).
- 700 65. Motrico, E. *et al.* Psychometric properties of the List of Threatening
701 Experiences--LTE and its association with psychosocial factors and mental
702 disorders according to different scoring methods. *J Affect Disord* **150**, 931-
703 40 (2013).
- 704 66. Goldberg, D.P. & Hillier, V.F. A scaled version of the General Health
705 Questionnaire. *Psychol Med* **9**, 139-45 (1979).
- 706 67. Sterling, M. General Health Questionnaire - 28 (GHQ-28). *J Physiother* **57**,
707 259 (2011).
- 708 68. Goldberg, D.P. *et al.* The validity of two versions of the GHQ in the WHO
709 study of mental illness in general health care. *Psychol Med* **27**, 191-7
710 (1997).
- 711 69. Wang, L. *et al.* [Value of patient health questionnaires (PHQ)-9 and PHQ-2
712 for screening depression disorders in cardiovascular outpatients].
713 *Zhonghua Xin Xue Guan Bing Za Zhi* **43**, 428-31 (2015).
- 714 70. Spitzer, R.L., Kroenke, K. & Williams, J.B. Validation and utility of a self-
715 report version of PRIME-MD: the PHQ primary care study. Primary Care

716 Evaluation of Mental Disorders. Patient Health Questionnaire. *JAMA* **282**,
717 1737-44 (1999).

718 71. Smith, D.J. *et al.* Prevalence and characteristics of probable major
719 depression and bipolar disorder within UK biobank: cross-sectional study
720 of 172,751 participants. *PLoS One* **8**, e75362 (2013).

721 72. Liu, J.Z., Erlich, Y. & Pickrell, J.K. Case-control association mapping by
722 proxy using family history of disease. *Nat Genet* **49**, 325-331 (2017).

723 73. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-
724 wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).

725 74. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and
726 population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).

727 75. Almli, L.M. *et al.* Correcting systematic inflation in genetic association tests
728 that consider interaction effects: application to a genome-wide association
729 study of posttraumatic stress disorder. *JAMA Psychiatry* **71**, 1392-9 (2014).

730 76. Keller, M.C. Gene x environment interaction studies have not properly
731 controlled for potential confounders: the problem and the (simple)
732 solution. *Biol Psychiatry* **75**, 18-24 (2014).

733 77. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. FUMA:
734 Functional mapping and annotation of genetic associations. *bioRxiv*
735 (2017).

736 78. Euesden, J., Lewis, C.M. & O'Reilly, P.F. PRSice: Polygenic Risk Score
737 software. *Bioinformatics* **31**, 1466-8 (2015).

738 79. Cross-Disorder Group of the Psychiatric Genomics, C. *et al.* Genetic
739 relationship between five psychiatric disorders estimated from genome-
740 wide SNPs. *Nat Genet* **45**, 984-94 (2013).

741 80. Howard, D.M. *et al.* Genome-wide meta-analysis of depression in 807,553
742 individuals identifies 102 independent variants with replication in a
743 further 1,507,153 individuals. *bioRxiv* (2018).

744 81. Altman, D.G. & Royston, P. The cost of dichotomising continuous variables.
745 *BMJ* **332**, 1080 (2006).

746 82. Stein, J.L. *et al.* Voxelwise genome-wide association study (vGWAS).
747 *Neuroimage* **53**, 1160-74 (2010).

748 83. Li, Z. *et al.* Genome-wide association analysis identifies 30 new
749 susceptibility loci for schizophrenia. *Nat Genet* **49**, 1576-1583 (2017).

750 84. Hart, A.B. *et al.* Genome-wide association study of d-amphetamine
751 response in healthy volunteers identifies putative associations, including
752 cadherin 13 (CDH13). *PLoS One* **7**, e42646 (2012).

753 85. Sugimoto, K. *et al.* The induction of H3K9 methylation by PIWIL4 at the
754 p16Ink4a locus. *Biochem Biophys Res Commun* **359**, 497-502 (2007).

755 86. Sivagurunathan, S., Arunachalam, J.P. & Chidambaram, S. PIWI-like protein,
756 HIWI2 is aberrantly expressed in retinoblastoma cells and affects cell-
757 cycle potentially through OTX2. *Cell Mol Biol Lett* **22**, 17 (2017).

758 87. Lee, H.H.C. *et al.* Genetic Otx2 mis-localization delays critical period
759 plasticity across brain regions. *Mol Psychiatry* **22**, 680-688 (2017).

760 88. Sonuga-Barke, E.J. *et al.* Does parental expressed emotion moderate
761 genetic effects in ADHD? An exploration using a genome wide association
762 scan. *Am J Med Genet B Neuropsychiatr Genet* **147B**, 1359-68 (2008).

763 89. Belzeaux, R. *et al.* Responder and nonresponder patients exhibit different
764 peripheral transcriptional signatures during major depressive episode.
765 *Transl Psychiatry* **2**, e185 (2012).

- 766 90. Mamdani, F., Berlim, M.T., Beaulieu, M.M. & Turecki, G. Pharmacogenomic
767 predictors of citalopram treatment outcome in major depressive disorder.
768 *World J Biol Psychiatry* **15**, 135-44 (2014).
- 769 91. Wong, M.L. *et al.* The PHF21B gene is associated with major depression
770 and modulates the stress response. *Mol Psychiatry* **22**, 1015-1025 (2017).
- 771 92. Walsh, R.M. *et al.* Phf8 loss confers resistance to depression-like and
772 anxiety-like behaviors in mice. *Nat Commun* **8**, 15142 (2017).
- 773 93. Shi, G. *et al.* PHD finger protein 2 (PHF2) represses ribosomal RNA gene
774 transcription by antagonizing PHF finger protein 8 (PHF8) and recruiting
775 methyltransferase SUV39H1. *J Biol Chem* **289**, 29691-700 (2014).
- 776 94. Yu, D. *et al.* Cross-disorder genome-wide analyses suggest a complex
777 genetic relationship between Tourette's syndrome and OCD. *Am J*
778 *Psychiatry* **172**, 82-93 (2015).
- 779 95. Eaves, L.J., Last, K., Martin, N.G. & Jinks, J.L. A progressive approach to non-
780 additivity and genotype-environmental covariance in the analysis of
781 human differences. *British Journal of Mathematical and Statistical*
782 *Psychology* **30**, 1-42 (1977).
- 783 96. Polderman, T.J. *et al.* Meta-analysis of the heritability of human traits
784 based on fifty years of twin studies. *Nat Genet* **47**, 702-9 (2015).
- 785 97. Ceccato, S., Kudielka, B.M. & Schwieren, C. Increased Risk Taking in
786 Relation to Chronic Stress in Adults. *Front Psychol* **6**, 2036 (2015).
- 787 98. Kandler, C., Bleidorn, W., Riemann, R., Angleitner, A. & Spinath, F.M. Life
788 events as environmental States and genetic traits and the role of
789 personality: a longitudinal twin study. *Behav Genet* **42**, 57-72 (2012).
- 790 99. Conway, C.C., Hammen, C., Brennan, P.A., Lind, P.A. & Najman, J.M.
791 Interaction of chronic stress with serotonin transporter and catechol-O-
792 methyltransferase polymorphisms in predicting youth depression. *Depress*
793 *Anxiety* **27**, 737-45 (2010).
- 794 100. Cicchetti, D. & Rogosch, F.A. Genetic moderation of child maltreatment
795 effects on depression and internalizing symptoms by serotonin
796 transporter linked polymorphic region (5-HTTLPR), brain-derived
797 neurotrophic factor (BDNF), norepinephrine transporter (NET), and
798 corticotropin releasing hormone receptor 1 (CRHR1) genes in African
799 American children. *Dev Psychopathol* **26**, 1219-39 (2014).

801

802 **FIGURE LEGENDS**

803

804 **Figure 1. Study flowchart.** Overview of analysis conducted on this study to achieve
 805 our aims: i) identify loci associated with depressive symptoms through genetic
 806 response to SLE; ii) test whether results of studying dependent and independent SLE
 807 support a contribution of genetically mediated exposure to stress; iii) assess whether
 808 GxE effects improve the proportion of phenotypic variance in depressive symptoms
 809 explained by genetic additive main effects alone and iv) test whether there is
 810 significant overlap in the genetic aetiology of the response to SLE and mental and
 811 physical stress-related phenotypes. Two core cohorts are used, Generation Scotland
 812 (GS) and UK Biobank (UKB). Summary statistics from Genome-Wide Association
 813 Studies (GWAS) and Genome-Wide by Environment Interaction Studies (GWEIS) are
 814 used to generate Polygenic Risk Scores (PRS). Summary statistics from Psychiatric
 815 Genetic Consortium (PGC) Major Depressive Disorder (MDD) GWAS are also used to
 816 generate PRS (PRS_{MDD}). PRS weighted by: additive effects (PRS_D and PRS_{MDD}), GxE
 817 effects (PRS_{GxE}) and joint effects (the combined additive and GxE effect; PRS_{Joint}), are
 818 used for phenotypic prediction. TSLE stands for Total number of SLE reported. DSLE
 819 stands for SLE dependent on an individual's own behaviour. Conversely, ISLE stands for
 820 independent SLE. N stands for sample size. N_{noGS} stands for sample size with GS
 821 individuals removed. N_{noUKB} stands for sample size with UKB individuals removed.

822

823 **Figure 2. Prediction of depression scores by PRS_{GxE} , PRS_D , PRS_{MDD} & PRS_{Joint} .** Variance
 824 of depression score explained by PRS_{GxE} , PRS_D , PRS_{MDD} and PRS_{Joint} as single effect; and
 825 combining both PRS_D and PRS_{MDD} with PRS_{GxE} in single models. Prediction was
 826 conducted using **2a)** Generation Scotland and **2b)** UK Biobank as target sample.
 827 PRS_{GxE} were weighted by cross sample GWEIS using GxE effect. PRS_D were weighted
 828 by cross sample GWAS of depressive symptoms effect. PRS_{MDD} was weighted by PGC
 829 MDD-GWAS summary statistics. PRS_{Joint} were weighted by cross sample GWEIS using
 830 joint effect. A nominally significant gain in variance explained of GHQ of about 23%
 831 was seen in Generation Scotland when PRS_{GxE} was incorporated into a multiple
 832 regression model along with PRS_D ; and of about 19% when PRS_{GxE} was incorporated
 833 into a multiple regression model along with PRS_{MDD} . Such gain was not seen in UK
 834 Biobank, but it must be noted that both PRS_D and PRS_{MDD} also explains much less

835 variance of PHQ in UK Biobank than of GHQ in Generation Scotland. To note a
836 noticeably reduction of variance explained by PRS_{Joint} compared to combined
837 PRS/effects.

838

839 **Figure 3. a) PRS prediction in independent GS datasets.** Heatmap illustrating PRS
840 prediction of a wide range of traits from GS listed in the x-axis (Table 1). (R) refers to
841 traits using mapping by proxy approach (i.e. where first-degree relatives of individuals
842 with the disease are considered proxy cases and included into the group of cases). Y-
843 axis shows the discovery sample and the effect used to weight PRS. Numbers in cells
844 indicate the % of variance explained, also represented by colour scale. Significance is
845 represented by “*” according to the following significance codes: p -values **** <
846 1×10^{-4} < *** < 0.001 < ** < 0.01 * < 0.05; in grey *Empirical-p-values* after permutation
847 (10 000 times) and in yellow *FDR-adjusted Empirical-p-values*. **b) Predictive**
848 **improvement by GxE effect in independent GS datasets.** Heatmap illustrating the
849 predictive improvement as a result of incorporating PRS_{GxE} into a multiple model
850 along with PRS_D and covariates (full model), over a model fitting PRS_D alone with
851 covariates (null model); predicting a wide range of traits from GS listed in the x-axis
852 (Table 1). Covariates: age, sex and 20 PCs. (R) refers to traits using mapping by proxy
853 approach (i.e. where first-degree relatives of individuals with the disease are consider
854 proxy cases and included into the group of cases). PRS_{GxE} are weighted by GWEIS
855 using GxE effects. PRS_D were weighted by the GWAS of depressive symptoms additive
856 main effects. The Y-axis shows the discovery sample used to weight PRS. Numbers in
857 cells indicate the % of variance explained by the PRS_{GxE} , also represented by colour
858 scale. Notice that those correspond to the PRS_{GxE} predictions in Figure 2 when PRS_{GxE}
859 are weighted by GxE effects. Significance was tested by Likelihood ratio tests (LRT):
860 full model including $PRS_D + PRS_{GxE}$ vs. null model with PRS_D alone (covariates
861 adjusted). Significance is represented by “*” according to the following significance
862 codes: p -values *** < 0.001 < ** < 0.01 * < 0.05; in grey *LRT-p-values* and in yellow
863 *FDR-adjusted LRT-p-values*.

Table 1. GS samples with stress-related phenotypes.

Trait	N	Males/Females	N SNPs	N Cases	N Controls
Alzheimer (R)	3377	1475/1903	560622	655	2722
Asthma	3390	1500/1890	560569	555	2835
Asthma (R)	3375	1470/1905	560432	910	2465
Bowel cancer (R)	3386	1495/1891	560630	672	2714
Breast cancer	3388	1486/1902	560611	83	3305
Breast cancer (R)	3386	1482/1904	560579	564	2822
Chronic obstructive pulmonary disease	3387	1496/1891	560591	73	3314
Chronic obstructive pulmonary disease (R)	3387	1474/1913	560620	553	2834
Depression	3385	1495/1890	560584	483	2902
Depression (R)	3382	1506/1876	560514	731	2651
Diabetes	3388	1497/1891	560469	185	3203
Diabetes (R)	3389	1481/1908	560584	1144	2245
Heart disease	3392	1504/1888	560526	212	3180
Heart disease (R)	3377	1483/1894	560479	2254	1123
High blood pressure	3402	1501/1901	560508	729	2673
High blood pressure (R)	3372	1464/1908	560569	1901	1471
Hip fracture (R)	3388	1489/1899	560572	421	2967
Lung cancer (R)	3379	1492/1887	560600	798	2581
Osteoarthritis	3395	1486/1909	560640	411	2984
Osteoarthritis (R)	3383	1466/1917	560516	961	2422
Parkinson (R)	3388	1488/1900	560590	236	3152
Prostate cancer (R)	3381	1495/1886	560570	329	3052
Rheumatoid arthritis	3387	1490/1897	560618	93	3294
Rheumatoid arthritis (R)	3380	1487/1893	560543	765	2615
Stroke	3387	1492/1895	560613	81	3306
Stroke (R)	3385	1463/1922	560478	1506	1879
Neuroticism*	3421	1521/1900	560484	-	-
Extraversion*	3420	1520/1900	560476	-	-
Schizotypal personality*	2386	1065/1321	560369	-	-
Mood disorder*	2307	1040/1267	560318	-	-

Samples were maximized for retention of cases to maximize the information available for each trait. There was no preferential selection of relatives in pairs for quantitative phenotypes, in order to retain the underlying distribution. All individuals involved in the datasets listed above were non-respondents to the GS follow-up study. Compared to individuals included at GS GWEIS (respondents in GS follow-up), non-respondents were significantly: younger, from more socioeconomically deprived areas, generally less healthier and wealthier. Non-respondents were more likely to smoke, and less likely to drink alcohol, although they consumed more units per week, compared with respondents. At GS baseline, non-respondents experienced more psychological distress and reported higher scores in symptoms of GHQ-depression and GHQ-anxiety than respondents⁵⁷. The total target sample size (N), number of males and females in N, number of SNPs (N SNPs) in target sample size N: the number of SNPs used as predictors after clumping step range between 90650 - 91000. The number of cases and controls in the independent target sample is indicated for binary phenotypes only. Samples were mapping by proxy approach was used (i.e. where first-degree relatives of individuals with the disease were considered proxy cases and included into the group of cases) are indicated by (R). *Assessed through self-reported questionnaires.

Figure 1.

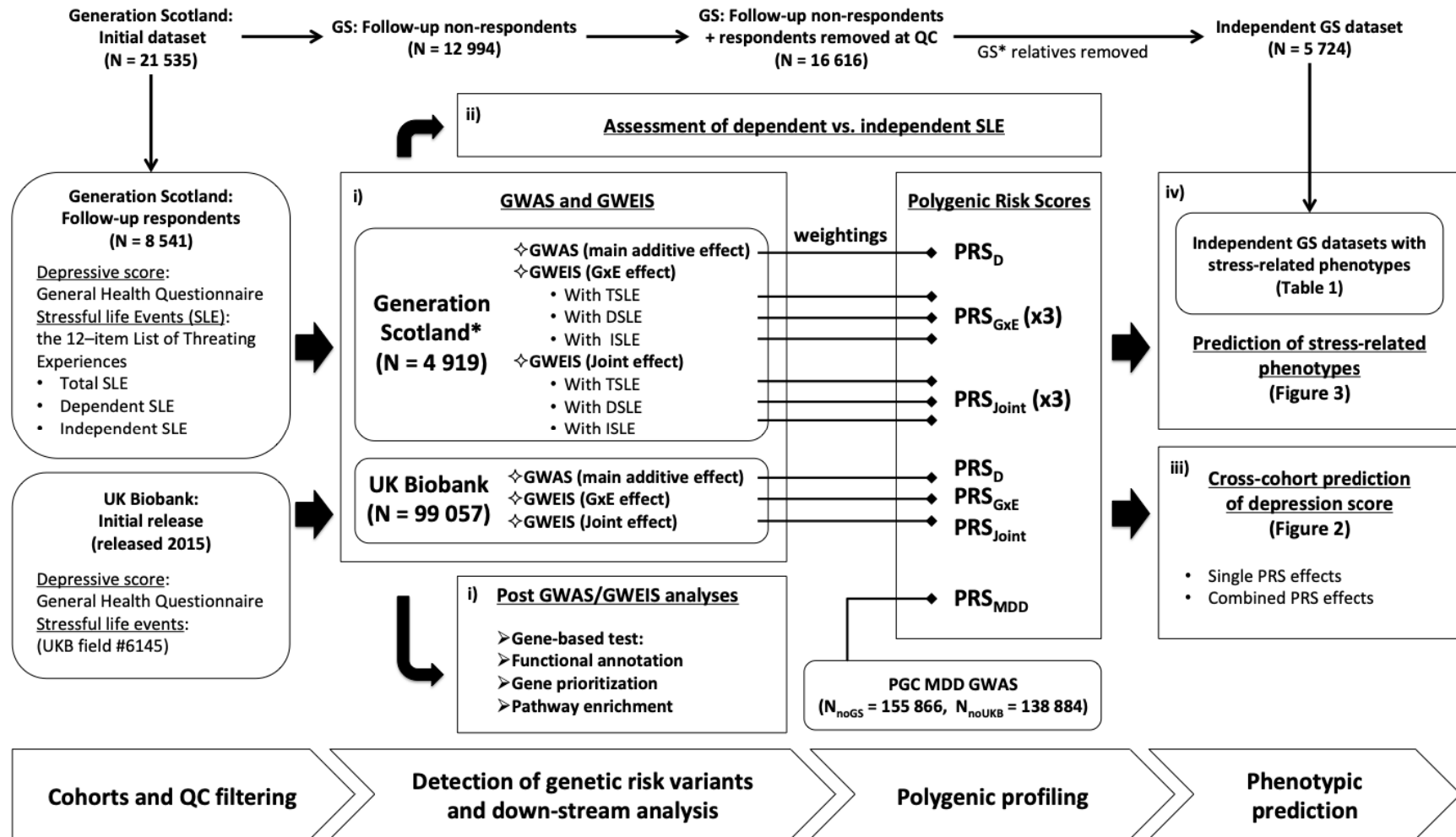


Figure 2.

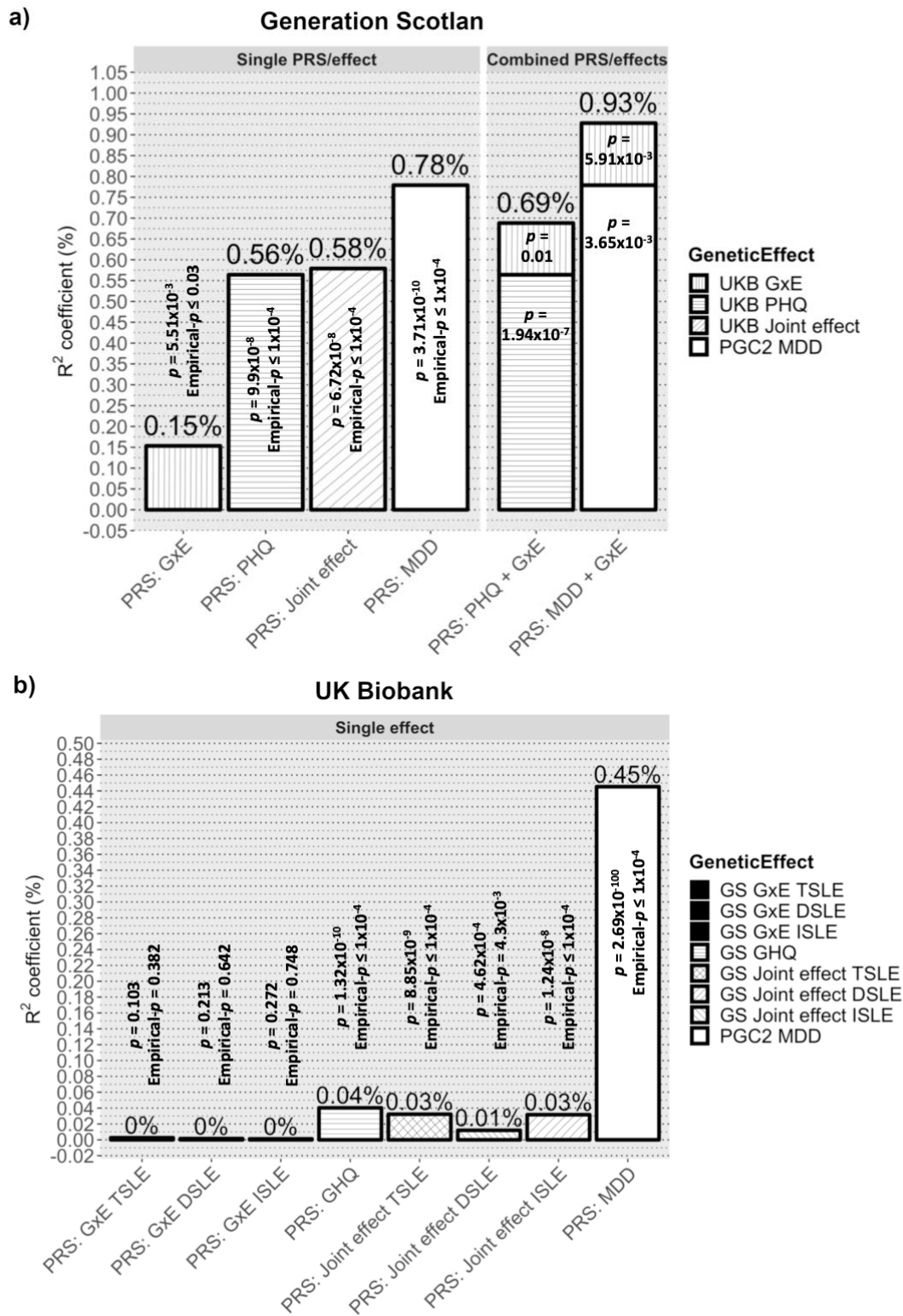


Figure 3.

