

GWAS Meta-Analysis of Neuroticism (N=449,484) Identifies Novel Genetic Loci and Pathways

Mats Nagel^{1,2†}, Philip R Jansen^{1,3†}, Sven Stringer¹, Kyoko Watanabe¹, Christiaan A de Leeuw¹, Julien Bryois⁴, Jeanne E Savage¹, Anke R Hammerschlag¹, Nathan Skene⁵, Ana B Muñoz-Manchado⁵, the 23andMe Research Team⁶, Sten Linnarsson⁵, Jens Hjerling-Leffler⁵, Tonya JH White³, Henning Tiemeier^{3,7}, Tinca JC Polderman¹, Patrick F Sullivan^{4,8,9}, Sophie van der Sluis^{1,2§}, Danielle Posthuma^{1,2§*}

Affiliations:

- ¹ Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, The Netherlands
- ² Department of Clinical Genetics, Section Complex Trait Genetics, Amsterdam Neuroscience, VU University Medical Center, Amsterdam, the Netherlands
- ³ Department of Child and Adolescent Psychiatry, Erasmus University Medical Center, Rotterdam, the Netherlands
- ⁴ Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
- ⁵ Laboratory of Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden
- ⁶ 23andMe, Inc., Mountain View, CA, USA
- ⁷ Department of Psychiatry, Erasmus University Medical Center, Rotterdam, the Netherlands
- ⁸ Department of Genetics, University of North Carolina, Chapel Hill, USA
- ⁹ Department of Psychiatry, University of North Carolina, Chapel Hill, USA

† These authors contributed equally to this work

§ These authors jointly supervised this work

*Correspondence should be addressed to: Danielle Posthuma: Department of Complex Trait Genetics, Vrije Universiteit Amsterdam, De Boelelaan 1085, 1081 HV, Amsterdam, The Netherlands. Phone: +31 20 598 2823, Fax: +31 20 5986926, d.posthuma@vu.nl

Word count: introductory paragraph: 150 main text: 1,498

Display items: 5 (Tables 1, Figures 4)

Includes **Online Methods, Supplementary Information** (Figures 14) and **Supplementary Tables** in excel 34.

1 Neuroticism is an important risk factor for psychiatric traits including depression¹, anxiety^{2,3},
2 and schizophrenia⁴⁻⁶. Previous genome-wide association studies⁷⁻¹² (GWAS) reported 16
3 genomic loci¹⁰⁻¹². Here we report the largest neuroticism GWAS meta-analysis to date
4 (N=449,484), and identify 136 independent genome-wide significant loci (124 novel),
5 implicating 599 genes. Extensive functional follow-up analyses show enrichment in several
6 brain regions and involvement of specific cell-types, including dopaminergic neuroblasts
7 ($P=3\times 10^{-8}$), medium spiny neurons ($P=4\times 10^{-8}$) and serotonergic neurons ($P=1\times 10^{-7}$). Gene-set
8 analyses implicate three specific pathways: neurogenesis ($P=4.4\times 10^{-9}$), behavioural response
9 to cocaine processes ($P=1.84\times 10^{-7}$), and axon part ($P=5.26\times 10^{-8}$). We show that neuroticism's
10 genetic signal partly originates in two genetically distinguishable subclusters¹³ (*depressed*
11 *affect* and *worry*, the former being genetically strongly related to depression, $r_g=0.84$),
12 suggesting distinct causal mechanisms for subtypes of individuals. These results vastly enhance
13 our neurobiological understanding of neuroticism, and provide specific leads for functional
14 follow-up experiments.

15

16 The neuroticism meta-analysis comprised data from the UK Biobank Study (UKB, full
17 release¹⁴; N=372,903; **Online Methods; Supplementary Fig. 1**), 23andMe, Inc.¹⁵
18 (N=59,206), and the Genetics of Personality Consortium (GPC1⁹; N=17,375; **Online**
19 **Methods**, N=449,484 in total). In all samples, neuroticism was measured through (digital)
20 questionnaires (**Online Methods; Supplementary Information**). SNP associations were
21 meta-analyzed using METAL¹⁶, weighted by sample size (**Online Methods**). The quantile-
22 quantile (Q-Q) plot of the genome-wide meta-analysis on 449,484 subjects and 14,978,477
23 SNPs showed high inflation ($\lambda=1.65$) and mean χ^2 statistic (1.91) (**Fig. 1a; Supplementary**
24 **Table 1**). The LD score regression (LDSC)^{17,18} intercept (1.02; SE=0.01) was consistent with

25 inflation due to true polygenicity and large sample size. The LDSC SNP-based heritability
26 (h^2_{SNP}) of neuroticism was 0.100 (SE=0.003).

27 The GWAS meta-analysis identified 9,745 genome-wide significant (GWS) SNPs ($P < 5 \times 10^{-8}$),
28 of which 157 and 2,414 were located in known associated inversions on chromosomes 8 and
29 17¹⁰⁻¹², respectively (**Supplementary Table 2; Fig. 1b; Supplementary Fig. 2**). FUMA¹⁹, a
30 tool to functionally map and annotate GWAS results (**Online Methods**), extracted 170
31 independent lead SNPs (158 novel; see **Supplementary information** for definition of lead
32 SNPs), which mapped to 136 independent genomic loci (124 novel; **Online Methods**;
33 **Supplementary information; Supplementary Table 3-8**). Of all lead SNPs, 4 were in exonic,
34 88 in intronic, and 52 in intergenic regions. Of the 17,794 SNPs in high LD with one of the
35 independent significant SNPs (see **Supplementary information** for definition of independent
36 significant SNPs), most were intronic (9,147: 51,4%) or intergenic (5,460: 30,7%), and 3.8%
37 was annotated as potentially having a functional impact, with 0.9% (155 SNPs) being exonic
38 (**Fig. 1c, Supplementary Table 9**; see **Supplementary Tables 10-11** for an overview of
39 chromatin state and regulatory functions of these SNPs). Of these, 37 were exonic non-
40 synonymous (ExNS) (**Table 1, Supplementary Table 12**). The highest CADD score (34) of
41 ExNS SNPs was for rs17651549, in exon 6 of *MAPT*, with a GWAS P -value of 1.11×10^{-28} , in
42 high LD with the lead SNP in that region ($r^2=0.97$). rs17651549 is a missense mutation leading
43 to an Arginine to Tryptophan change with allele frequencies matching the inversion in that
44 region. The ancestral allele C is associated with a lower neuroticism score (see **Table 1** and
45 **Supplementary Table 12** for a detailed overview of all functional variants in genomic risk
46 loci).

47 Stratified LDSC²⁰ (**Online Methods**), showed significant enrichment for h^2 of SNPs located
48 in conserved regions (enrichment=13.79, $P=5.14 \times 10^{-16}$), intronic regions (enrichment=1.27,

49 $P=1.27\times 10^{-6}$), and in H3K4me3 (enrichment=2.14, $P=1.02\times 10^{-5}$) and H3K9ac regions
50 (enrichment=2.17, $P=3.06\times 10^{-4}$) (**Fig. 1d; Supplementary Table 13**).

51 Polygenic scores (PGS) calculated using PRSice²¹ (clumping followed by P -value
52 thresholding) and LDpred²² in three randomly drawn hold-out samples (UKB only, N=3,000
53 each; **Online Methods**), explained up to 4.2% ($P=1.49\times 10^{-30}$) of the variance in neuroticism
54 (**Supplementary Fig. 3; Supplementary Table 14**).

55 We used four strategies to link our SNP results to genes: positional, eQTL, and chromatin
56 interaction mapping (**Online Methods**) and genome-wide gene-association analysis
57 (GWGAS; MAGMA²³). GWGAS evaluates the joint association effect of all SNPs within a
58 gene yielding a gene-based P -value. Based on our meta-analytic results, 283 genes were
59 implicated through positional mapping, 369 through eQTL-mapping, and 119 through
60 chromatin interaction-mapping (**Fig. 2a; Supplementary Table 15**). GWGAS identified 336
61 GWS genes ($P<2.75\times 10^{-6}$, **Figs. 2b-c; Supplementary Table 16, Supplementary**
62 **information**), of which 203 overlapped with genes implicated by FUMA, resulting in 599
63 unique neuroticism-related genes. Of these, 50 were implicated by all four methods, of which
64 49 had chromatin interaction and eQTL associations in the same tissue/cell type (**Fig. 2a,**
65 **Supplementary Table 15**).

66

67 19 of the 119 genes implicated through chromatin interaction mapping are especially
68 interesting as they are implicated via interactions between two independent GWS genomic risk
69 loci. There are several chromatin interactions in 7 tissue types (aorta, hippocampus, left
70 ventricle, right ventricle, liver, spleen, pancreas) across two risk loci on chromosome 6 (**Fig.**
71 **3a**). Two genes are located in locus 45 and are mapped by chromatin interactions from risk
72 locus 46 (*HFE* and *HIST1H4C*), and another 16 genes are coding histones in locus 46 and are
73 mapped by interactions from locus 45 (**Supplementary Table 15**). *XKRR6* is located on

74 chromosome 8 in risk locus 61, and is implicated by chromatin interactions in 5 tissue types
75 (aorta, left ventricle, liver, pancreas and spleen) including cross locus interactions from locus
76 60 (**Fig. 3b; Supplementary Table 15**). This gene is also mapped by eQTLs in blood and
77 transformed fibroblasts. Out of the 19 genes mapped by two loci, 4 are located outside of the
78 risk loci (*HIST1H2AI*, *HIST1H3H*, *HIST1H2AK* and *HIST1H4L*), and 7 are also implicated by
79 eQTLs in several tissue types (*HFE* in adipose subcutaneous, aorta, esophagus muscularis,
80 lung, tibial nerve, sub-exposed skin and thyroid; *HIST1H4J* in blood and adrenal gland;
81 and *HIST1H4K*, *HIST1H2AK*, *HIST1H2BO* and *XKR6* in blood).

82

83 Gene-based *P*-values were used for gene-set analysis in MAGMA^{23,26}, testing 7,246 pre-
84 defined gene-sets derived from MsigDB²⁴, gene expression profiles in 53 tissue types obtained
85 from the GTEx Project²⁵, and 24 cell-type specific expression profiles using RNAseq
86 information²⁶ (**Online Methods**). Neuroticism was significantly associated with genes
87 predominantly expressed in 11 brain tissue types (**Fig. 2d; Supplementary Table 17-18**) and
88 with 7 gene ontology (GO) gene-sets, with the strongest association for neurogenesis
89 ($P=0.0004$) and neuron differentiation ($P=0.002$) (**Supplementary Table 17**). Conditional
90 gene-set analyses (**Online Methods**) suggested that 3 of the 7 gene-sets (neurogenesis,
91 $P=4.4\times 10^{-9}$; behavioral response to cocaine, $P=1.84\times 10^{-7}$; axon part, $P=5.26\times 10^{-8}$) had largely
92 independent associations, implying a role in neuroticism (**Supplementary Table 19**).
93 Conditional analyses of the tissue-specific expression ascertained general involvement of
94 (frontal) cortex expressed genes (**Supplementary Table 20; Supplementary Fig. 4**).

95 Cell type specific gene-set analysis showed significant association with genes expressed in
96 multiple brain cell types (**Fig. 2e; Supplementary Table 21**), with dopaminergic neuroblasts
97 ($P=3\times 10^{-8}$), medium spiny neurons ($P=4\times 10^{-8}$) and serotonergic neurons ($P=1\times 10^{-7}$) showing

98 the strongest associations, and conditional analysis indicated that these three cell types were
99 also independently associated with neuroticism.

100 Aiming to further specify neuroticism's neurobiological interpretation, we compared the
101 genetic signal of the full neuroticism trait to that of two genetically distinguishable neuroticism
102 subclusters *depressed affect* and *worry*¹³ (**Online Methods**). As a validation of the *depressed*
103 *affect* dimension, we also compare with GWAS results for depression. GWA analyses of the
104 subclusters were conducted on the UKB-data only (dictated by item-level data availability;
105 **Online Methods**; *depressed affect*, N=357,957; *worry*, N=348,219). For depression, our meta-
106 analysis comprised data from the UKB¹⁴ (N=362,696; **Supplementary Fig. 5**), 23andMe¹⁵
107 (N=307,354), and the Psychiatric Genetics Consortium (PGC²⁷; N=18,759) (total N=688,809,
108 not previously published; r_g between samples: 0.61-0.80; **Online Methods**; **Supplementary**
109 **Table 22, Supplementary Information**). Genetic correlations of neuroticism with all three
110 phenotypes were considerable (depression: $r_g=0.79$; *depressed affect*: $r_g=0.88$, *worry*: $r_g=0.87$;
111 **Supplementary Table 23**).

112 The subclusters showed notable differences in genetic signal (e.g., exclusive GWS associations
113 on chromosomes 2 and 19 for *depressed affect*, and chromosomes 3 and 22 for *worry*;
114 **Supplementary Figs. 6-12; Supplementary Tables 24-26**). Of the 136 genetic loci associated
115 with neuroticism, 32 were also GWS for *depressed affect* (7 shared with depression) but not
116 for *worry*, and 26 were also GWS for *worry* (3 shared with depression) but not for *depressed*
117 *affect* (**Supplementary Table 27; Supplementary Fig. 12**). These results were mirrored by
118 gene-based analyses (**Supplementary information; Supplementary Tables 28-30;**
119 **Supplementary Fig. 13**), suggesting that part of neuroticism's genetic signal originates
120 specifically in one of the two subclusters, possibly implicating different causal genetic
121 mechanisms .

122 To test specificity of the gene-sets implicated in neuroticism in the conditional analyses, we
123 repeated the analyses, but now corrected for *depressed affect*, and *worry* scores, respectively
124 (**Supplementary Table 31; Supplementary Fig. 14**). The association with ‘axon-part’ was
125 markedly lower after correction for *worry* scores (uncorrected $P=5.26\times 10^{-8}$; corrected for
126 *depressed affect* $P=2.42\times 10^{-6}$; corrected for *worry* $P=.0013$), suggesting that the involvement
127 of ‘axon-part’ in neuroticism originates predominantly from the *worry*-component.

128

129 To examine the genetic correlational pattern of neuroticism, and to compare it to the patterns
130 observed for depression, *depressed affect* and *worry*, we used LDSC to calculate genetic
131 correlations with 35 traits for which large-scale GWAS summary statistics were available
132 (**Supplementary Table 32; Online Methods**). We observed 11 Bonferroni-corrected
133 significant genetic correlations for neuroticism ($\alpha=0.05/(4\times 35)$; $P<3.6\times 10^{-4}$) (**Fig. 4;**
134 **Supplementary Table 33**), covering previously reported psychiatric traits (r_g range: .20-.82)
135 and subjective well-being ($r_g = -.68$). These correlations were supported by enrichment of
136 neuroticism genes in sets of genes previously implicated in psychiatric traits (**Supplementary**
137 **Table 34**). The r_g 's of depression and *depressed affect* strongly mirrored each other (correlation
138 between their r_g 's is $r=.98$; **Supplementary information**), validating the *depressed affect*
139 cluster. The correlational patterns for *depressed affect* and *worry* were markedly different and
140 sometimes antipodal, with the genetic signal of the full neuroticism trait being a blend of both.

141

142 In conclusion, we identified 119 novel genetic loci for neuroticism. Extensive functional
143 annotations highlighted several genes being implicated through multiple routes. We
144 demonstrated the involvement of specific neuronal cell types and three independently
145 associated genetic pathways, and established the genetic multidimensionality of the
146 neuroticism phenotype, and its link with depression. The current study provides new leads, and

147 testable functional hypotheses for unraveling the neurobiology of neuroticism, its subtypes,

148 and genetically associated traits.

149

150 **References and notes**

- 151 1. Kendler, K. S. & Myers, J. The genetic and environmental relationship between major
152 depression and the five-factor model of personality. *Psychol. Med.* **40**, 801–806 (2010).
- 153 2. Middeldorp, C. M. *et al.* The Association of Personality with Anxious and Depressive
154 Psychopathology. *Biol. Personal. Individ. Differ.* 251–272 (2006).
- 155 3. Hettema, J. M., Neale, M. C., Myers, J. M., Prescott, C. A. & Kendler, K. S. A
156 Population-Based Twin Study of the Relationship Between Neuroticism and
157 Internalizing Disorders. *Am. J. Psychiatry* **163**, 857–864 (2006).
- 158 4. Hayes, J. F. *et al.* Association of Late Adolescent Personality With Risk for Subsequent
159 Serious Mental Illness Among Men in a Swedish Nationwide Cohort Study. *JAMA*
160 *Psychiatry* **54**, 948–963 (2017).
- 161 5. Smeland, O. B. *et al.* Identification of genetic loci shared between schizophrenia and the
162 Big Five personality traits. *Sci. Rep.* **7**, 1–9 (2017).
- 163 6. Van Os, J. & Jones, P. B. Neuroticism as a risk factor for schizophrenia. *Psychol. Med.*
164 **31**, 1129–1134 (2001).
- 165 7. Genetics of Personality Consortium. Meta-analysis of Genome-wide Association
166 Studies for Neuroticism, and the Polygenic Association With Major Depressive
167 Disorder. *JAMA Psychiatry* **72**, 642–650 (2015).
- 168 8. Terracciano, A. *et al.* Genome-wide association scan for five major dimensions of
169 personality. *Mol. Psychiatry* **15**, 647–656 (2010).
- 170 9. Moor, M. H. M. De *et al.* Meta-analysis of genome-wide association studies for
171 personality. *Mol. Psychiatry* **17**, 337–349 (2012).
- 172 10. Lo, M. *et al.* Genome-wide analyses for personality traits identify six genomic loci and
173 show correlations with psychiatric disorders. *Nat. Genet.* (2016). doi:10.1038/ng.3736
- 174 11. Smith, D. J. *et al.* Genome-wide analysis of over 106 000 individuals identifies 9

- 175 neuroticism-associated loci. *Mol. Psychiatry* **21**, 1–9 (2016).
- 176 12. Okbay, A. *et al.* Genetic variants associated with subjective well-being, depressive
177 symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* **48**,
178 624–636 (2016).
- 179 13. Nagel, M., Watanabe, K., Stringer, S., Posthuma, D. & Van der Sluis, S. Item-level
180 Analyses Reveal Genetic Heterogeneity in Neuroticism. Manuscript submitted for
181 publication (2017).
- 182 14. Bycroft, C. *et al.* Genome-wide genetic data on ~500,000 UK Biobank participants.
183 *bioRxiv* (2017). at <<http://biorxiv.org/content/early/2017/07/20/166298.abstract>>
- 184 15. Eriksson, N. *et al.* Web-based, participant-driven studies yield novel genetic
185 associations for common traits. *PLoS Genet.* **6**, 1–20 (2010).
- 186 16. Willer, C. J., Li, Y., Abecasis, G. R. & Overall, P. METAL: fast and efficient meta-
187 analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
- 188 17. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from
189 polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
- 190 18. Bulik-Sullivan, B. K. *et al.* An atlas of genetic correlations across human diseases and
191 traits. *Nat. Genet.* **47**, 1–9 (2015).
- 192 19. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. FUMA: Functional
193 mapping and annotation of genetic associations. *bioRxiv* (2017). at
194 <<http://biorxiv.org/content/early/2017/02/20/110023.abstract>>
- 195 20. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-
196 wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
- 197 21. Euesden, J., Lewis, C. M. & O’Reilly, P. F. PRSice: Polygenic Risk Score software.
198 *Bioinformatics* **31**, 1466–1468 (2015).
- 199 22. Vilhjálmsson, B. J. *et al.* Modeling Linkage Disequilibrium Increases Accuracy of

- 200 Polygenic Risk Scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
- 201 23. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-
202 Set Analysis of GWAS Data. *PLoS Comput. Biol.* **11**, 1–19 (2015).
- 203 24. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for
204 interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102**, 15545–15550
205 (2005).
- 206 25. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue
207 gene regulation in humans. *Science (80-.)*. **348**, 648–660 (2015).
- 208 26. Skene, N. G. *et al.* Genetic Identification Of Brain Cell Types Underlying
209 Schizophrenia. *bioRxiv* (2017). at
210 <<http://biorxiv.org/content/early/2017/06/02/145466.abstract>>
- 211 27. Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium. A
212 mega-analysis of genome-wide association studies for major depressive disorder. *Mol.*
213 *Psychiatry* **18**, 497–511 (2013).
- 214 28. Hyde, C. L. *et al.* Identification of 15 genetic loci associated with risk of major
215 depression in individuals of European descent. *Nat. Genet.* **48**, 1031–1036 (2016).
- 216 29. Eysenck, B. G., Eysenck, H. J. & Barrett, P. A Revised Version of the Psychoticism
217 Scale. *Pers. Individ. Dif.* **6**, 21–29 (1985).
- 218 30. John, O. P. & Srivastava, S. The Big Five trait taxonomy: History, measurement, and
219 theoretical perspectives. *Handb. Personal. Theory Res.* **2**, 102–138 (1999).
- 220 31. Soto, C. J. & John, O. P. Ten facet scales for the Big Five Inventory: Convergence with
221 NEO PI-R facets, self-peer agreement, and discriminant validity. *J. Res. Pers.* **43**, 84–
222 90 (2009).
- 223 32. Costa, P. & McCrae. *Professional Manual: Revised NEO Personality Inventory (NEO-*
224 *PI-R) and NEO Five-Factor-Inventory (NEO-FFI)*. (Psychological Assessment

- 225 Resources: Odessa, FL, 1992).
- 226 33. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74
227 (2015).
- 228 34. Webb, B. T. *et al.* Molecular Genetic Influences on Normative and Problematic Alcohol
229 Use in a Population-Based Sample of College Students. *Front. Genet.* **8**, 1–11 (2017).
- 230 35. Abraham, G. & Inouye, M. Fast Principal Component Analysis of Large-Scale Genome-
231 Wide Data. *PLoS One* **9**, 1–5 (2014).
- 232 36. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based
233 linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
- 234 37. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer
235 datasets. *Gigascience* **4**, 1 (2015).
- 236 38. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of
237 human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
- 238 39. Schmitt, A. D. *et al.* A Compendium of Chromatin Contact Maps Reveals Spatially
239 Active Regions in the Human Genome. *Cell Rep.* **17**, 2042–2059 (2016).
- 240 40. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human
241 epigenomes. *Nature* **518**, (2015).
- 242 41. Croft, D. *et al.* The Reactome pathway knowledgebase. *Nucleic Acids Res.* **42**, D472–
243 D477 (2014).
- 244 42. Coleman, J. *et al.* Functional consequences of genetic loci associated with intelligence
245 in a meta-analysis of 87,740 individuals. *bioRxiv* (2017). at
246 <<http://biorxiv.org/content/early/2017/07/31/170712.abstract>>
- 247 43. Gazal, S. *et al.* Linkage disequilibrium dependent architecture of human complex traits
248 reveals action of negative selection. *bioRxiv* (2016). at
249 <<http://biorxiv.org/content/early/2016/10/19/082024.abstract>>

250

251 **URLs:**

252 <http://ukbiobank.ac.uk>

253 <http://ctg.cncr.nl/software/magma>

254 <http://software.broadinstitute.org/gsea/msigdb/collections.jsp>

255 http://genome.sph.umich.edu/wiki/METAL_Program

256 <https://github.com/bulik/ldsc>

257 <http://fuma.ctglab.nl>

258

259 **Acknowledgements:** This work was funded by The Netherlands Organization for Scientific
260 Research (NWO Brain & Cognition 433-09-228, NWO MagW VIDI 452-12-014, NWO VICI
261 435-14-005 and 453-07-001, 645-000-003), Sophia Foundation for Scientific Research
262 (SSWO, grant nr: S-1427). JB was funded by the Swiss National Science Foundation. Analyses
263 were carried out on the Genetic Cluster Computer, which is financed by the Netherlands
264 Scientific Organization (NWO: 480-05-003), by the VU University, Amsterdam, the
265 Netherlands, and by the Dutch Brain Foundation, and is hosted by the Dutch National
266 Computing and Networking Services SurfSARA. This research has been conducted using the
267 UK Biobank Resource (application number 16406). We would like to thank the participants,
268 including the 23andMe customers who consented to participate in research, and researchers
269 who collected and contributed to the data.

270

271 **Author Contributions:** S.vd.S and D.P. conceived the study. M.N. and P.R.J. performed the
272 analyses. S. St. performed the quality control on the UK Biobank data and wrote a pipeline to
273 facilitate data processing. K.W. constructed the tool for biological annotation and ran the
274 analyses. J.B. and P.F.S. performed the single-cell gene-expression analysis. M.N., P.R.J.,
275 S.vd.S and D.P. wrote the paper. All authors discussed the results and commented on the paper.

276

277 **23andMe Research Team contributors:** Michelle Agee, Babak Alipanahi, Adam Auton,
278 Robert K. Bell, Katarzyna Bryc, Sarah L. Elson, Pierre Fontanillas, Nicholas A. Furlotte, David
279 A. Hinds, Bethann S. Hromatka, Karen E. Huber, Aaron Kleinman, Nadia K. Litterman,
280 Matthew H. McIntyre, Joanna L. Mountain, Elizabeth S. Noblin, Carrie A.M. Northover,
281 Steven J. Pitts, J. Fah Sathirapongsasuti, Olga V. Sazonova, Janie F. Shelton, Suyash
282 Shringarpure, Chao Tian, Joyce Y. Tung, Vladimir Vacic, and Catherine H. Wilson.

283

284 **Author Information:** Reprints and permissions information is available at
285 www.nature.com/reprints. The authors declare no competing financial interest.
286 Correspondence and requests for materials should be addressed to d.posthuma@vu.nl.

287

288 **Table 1. Exonic non-synonymous (ExNS) variants in the genomic loci associated with neuroticism and in LD ($r^2>0.6$) with one of the**
 289 **independent GWS SNPs.**

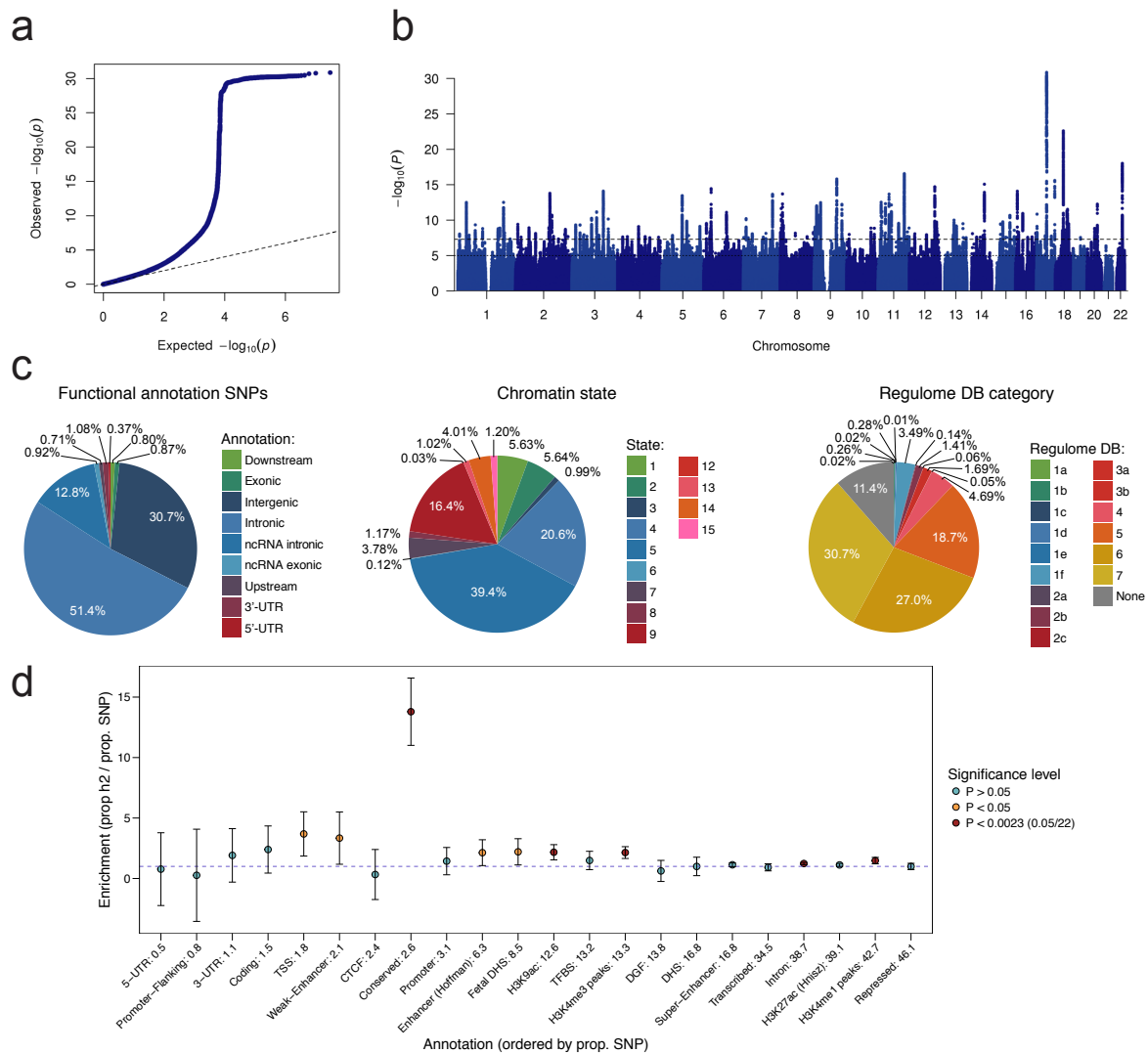
290 CADD: CADD score; rdb: regulome DB score; MAF: minor allele frequency; Z-score: z-score from the GWAS meta-analysis in METAL. Results
 291 are reported on hg19 coordinates (NCBI b37). Genes containing multiple ExNS are annotated in red.
 292

rsID	Exon	Gene	A1	MAF	gwas P	beta	r^2	Independent Sign. SNP	Locus	CADD	RDB	Min. Chromatin state
rs41266050	14	RABGAP1L	T	0.25	5.65E-06	-4.54	0.84	rs7536102	6	2.85	7	5
rs2073498	2	RASSF1	A	0.11	2.71E-08	5.56	0.98	rs6776145	25	19.43	7	3
rs111177	3	GNL3	A	0.38	2.49E-07	-5.16	0.75	rs2015971	26	22.90	NA	1
rs2289247	11	GNL3	A	0.41	2.28E-06	-4.73	0.65	rs2015971	26	12.82	NA	3
rs1029871	4	NEK4	C	0.38	2.29E-07	-5.17	0.75	rs2015971	26	24.10	1f	2
rs240780	39	ASCC3	C	0.43	3.01E-08	5.54	0.96	rs240769	49	19.95	7	4
rs11765552	11	LMTK2	A	0.46	7.68E-08	5.38	0.98	rs34320230	55	12.24	6	4
rs41274386	2	FAM120AOS	T	0.08	1.10E-07	5.31	0.66	rs78046549	71	2.36	4	1
rs1055710	1	FAM120AOS	A	0.33	1.11E-09	-6.09	0.99	rs10821129	71	0.05	NA	1
rs3816614	33	LRP4	T	0.23	5.69E-07	5.00	0.90	rs7940441	84	22.70	NA	4
rs2030166	5	NDUFS3	T	0.35	2.02E-10	-6.36	0.93	rs11039389	84	3.13	6	4
rs1064608	13	MTCH2	C	0.35	1.15E-10	-6.45	0.93	rs11039389	84	25.40	6	4
rs12286721	13	AGBL2	A	0.45	7.81E-08	-5.37	0.78	rs7107356	84	14.22	1f	5
rs4926	8	SERPING1	A	0.27	6.12E-07	4.99	0.86	rs73480560	85	23.50	5	4
rs11604671	6	ANKK1	A	0.49	2.57E-10	-6.32	0.64	rs2186800	88	1.39	5	4
rs1800497	8	ANKK1	A	0.20	8.45E-06	4.45	0.69	rs11214607	88	0.81	NA	4
rs7298565	12	UBE3B	A	0.48	2.24E-10	6.34	0.76	rs2111216	94	22.70	6	4
rs8007859	10	EXD2	T	0.39	2.28E-08	5.59	0.80	rs1275411	108	3.95	5	4
rs2286913	4	RPS6KL1	A	0.37	1.46E-07	5.26	0.89	rs3213716	110	12.96	5	2
rs7156590	3	RPS6KL1	T	0.37	2.79E-07	5.14	0.86	rs3213716	110	19.46	5	4
rs12443627	1	ENSG00000268863	C	0.37	1.28E-10	6.43	0.77	rs3751855	119	3.58	2b	1
rs35713203	2	ZNF646	C	0.38	3.67E-11	-6.62	0.98	rs3751855	119	0.05	2b	3
rs7196726	2	ZNF646	A	0.38	1.29E-11	-6.77	1.00	rs3751855	119	0.00	2b	3
rs7199949	8	PRSS53	C	0.38	1.32E-11	-6.77	1.00	rs3751855	119	0.00	2b	2
rs3748400	12	ZCCHC14	T	0.23	8.83E-09	-5.75	0.98	rs2042395	122	24.00	5	4
rs12949256	1	ARHGAP27	T	0.19	1.47E-23	10.00	0.73	rs77804065	126	11.97	4	1
rs16940674	6	CRHR1	T	0.23	5.24E-29	11.18	0.97	rs77804065	126	12.86	1f	5
rs16940681	13	CRHR1	C	0.23	2.18E-30	11.46	0.97	rs77804065	126	1.76	4	5
rs242944	1	SPPL2C	A	0.44	2.88E-12	-6.98	1.00	rs242947	126	0.00	NA	5
rs62054815	1	SPPL2C	A	0.23	1.74E-30	11.48	0.97	rs77804065	126	0.00	5	5
rs12185233	1	SPPL2C	C	0.23	6.76E-29	11.16	0.96	rs77804065	126	25.60	1f	5
rs12373139	1	SPPL2C	A	0.23	2.19E-30	11.46	0.97	rs77804065	126	0.53	1f	5
rs63750417	6	MAPT	T	0.23	4.89E-30	11.39	0.97	rs77804065	126	8.68	5	4

rsID	Exon	Gene	A1	MAF	gwas P	beta	r ²	Independent Sign. SNP	Locus	CADD	RDB	Min. Chromatin state
rs62063786	6	MAPT	A	0.23	1.05E-29	11.32	0.97	rs77804065	126	7.65	5	4
rs17651549	6	MAPT	T	0.23	1.11E-28	11.11	0.97	rs77804065	126	34.00	1f	4
rs17522826	1	TCF4	A	0.18	2.17E-10	6.35	0.60	rs10503002	133	14.22	NA	1
rs139431	2	L3MBTL2	T	0.37	9.45E-07	-4.90	0.63	rs7289932	138	10.26	7	4

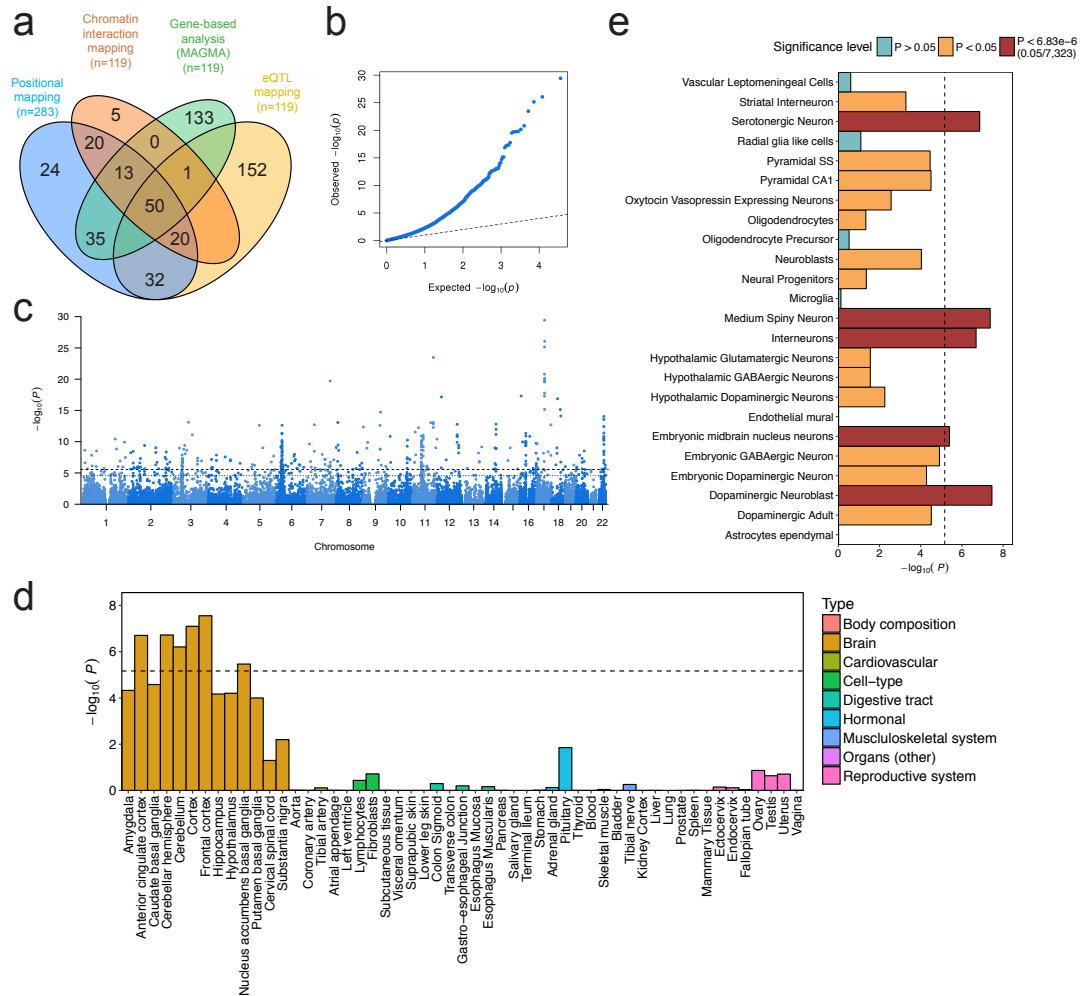
293

294



295

296 **Fig. 1. SNP-based associations with neuroticism in the GWAS meta-analysis.**
 297 **(a)** Quantile-quantile plot of the SNP-based associations with neuroticism. **(b)** Manhattan plot
 298 showing the $-\log_{10}$ transformed P -value of each SNP on the y-axis and base pair positions
 299 along the chromosomes on the x-axis. The dashed line indicates genome-wide significance
 300 ($P < 5 \times 10^{-8}$), the dotted line the threshold for suggestive associations ($P < 1 \times 10^{-5}$). **(c)** Pie charts
 301 showing the distribution of functional consequences of SNPs in linkage disequilibrium (LD)
 302 with genome-wide significant lead SNPs in the meta-analysis, the minimum chromatin state
 303 across 127 tissue and cell types and the distribution of regulome DB score, a categorical score
 304 between 1a and 7, indicating biological evidence of a SNP being a regulatory element, with a
 305 low score denoting a higher likelihood of being regulatory. **(d)** Heritability enrichment of 22
 306 functional SNP annotations calculated with stratified LD score regression.
 307

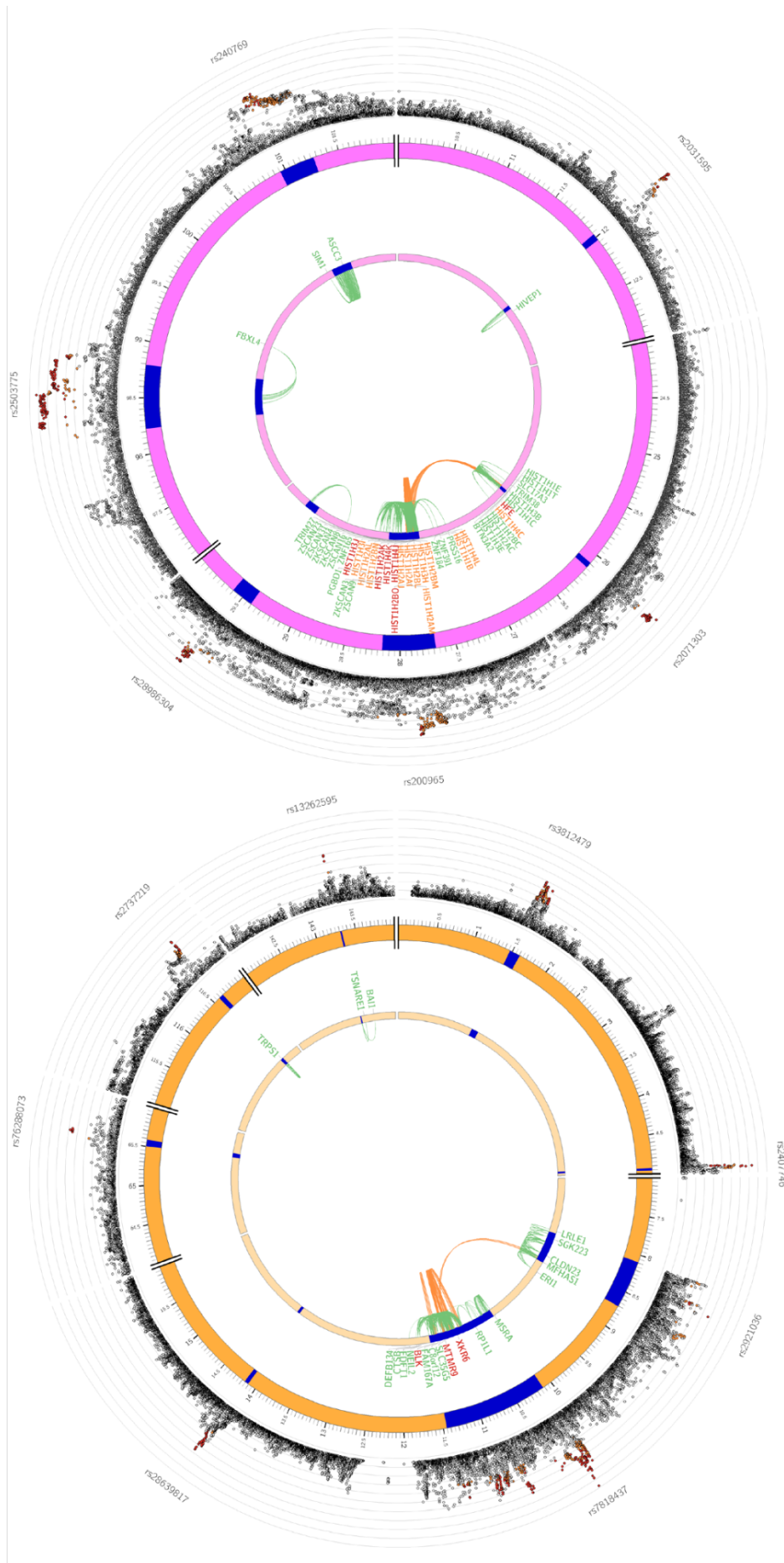


308

309 **Fig. 2. Mapping of genes and tissue- and cell expression profiles.**

310 (a) Venn diagram showing overlap of genes implicated by positional mapping, eQTL mapping,
 311 chromatin interaction mapping, and gene-based genome-wide association (GWGAS). (b)
 312 Quantile-quantile plot of the GWGAS. (c) Manhattan plot of the genome-wide gene-based
 313 association analysis (GWGAS) on neuroticism. The y-axis shows the $-\log_{10}$ transformed P -
 314 value of each gene, and the chromosomal position (start position) on the x-axis. The dashed
 315 line indicates the threshold for genome-wide significance of the gene-based test ($P < 2.76 \times 10^{-6}$;
 316 0.05/18,128), and the dotted line indicates the suggestive threshold ($P < 2.76 \times 10^{-5}$;
 317 0.5/18,128). (d) Gene expression profiles of identified genes for 53 tissue types. Expression
 318 data were extracted from the Genotype-Tissue Expression (GTEx) database. Expression values
 319 (RPKM) were \log_2 transformed with pseudocount 1 after winsorization at 50 and averaged per
 320 tissue. Gene-set tests for tissue expressions were calculated using MAGMA (**Online**
 321 **Methods**). (e) Enrichment of genetic signal for neuroticism in 24 brain cell types. The dashed
 322 line indicates the Bonferroni-corrected significance threshold ($P = 0.05/7,323 = 6.83 \times 10^{-6}$).

323



325 **Fig 3. Genomic risk loci, eQTL associations and chromatin interaction for chromosome**
326 **6 and 8, containing cross-locus interactions.**
327 Circos plot showing genes on **(a)** chromosome 6 and **(b)** chromosome 8 that were implicated
328 by the genomic risk (blue areas) loci by chromatin interaction (CTI; orange), eQTL (green) or
329 implicated by both eQTL and CTI mapping (red). The outer layer shows a Manhattan plot
330 containing the $-\log_{10}$ transformed P -value of each SNP in the GWAS meta-analysis. Empty
331 regions in the Manhattan plot layer indicate regions where no SNPs with $P < 0.05$ are situated.
332



333

334 **Fig. 4. Genetic correlations between neuroticism and other traits.**

335 Genetic correlations of neuroticism, depression, *depressed* affect and worry with various traits
 336 and diseases. LD score regression (**Online Methods**) tested genome-wide SNP associations
 337 for the neuroticism score against previously published results for 35 neuropsychiatric
 338 outcomes, antropometric and health-related traits, and brain morphology (**Supplementary**
 339 **Table 32-33**).

340

341 **Online methods**

342

343 ***Samples***

344 ***UK Biobank:*** The UK Biobank (UKB) Study is a major data resource, containing genetic and
345 a wide range of phenotypic data of ~500,000 participants aged 40-69 at recruitment¹⁴. We used
346 data released in July 2017, and selection (discussed below) resulted in final sample sizes of
347 N=372,903 and N=362,696 individuals for neuroticism and depression, respectively
348 (**Supplementary information**). The UKB received ethical approval from the National
349 Research Ethics Service Committee North West–Haydock (reference 11/NW/0382), and all
350 study procedures were performed in accordance with the World Medical Association for
351 medical research. The current study was conducted under UKB application number 16406.

352 ***23andMe:*** 23andMe, Inc. is a large personal genomics company that provides genotype and
353 health-related information to customers. For the neuroticism meta-analysis, we used
354 neuroticism GWAS summary statistics from a subset of 23andMe research participants
355 (N=59,206), described in more detail elsewhere¹⁰. For our depression meta-analysis, we used
356 depression GWAS summary statistics from a subset of 23andMe research participants
357 (N=307,354), described in detail elsewhere²⁸. All included participants provided informed
358 consent and were of European ancestry, and related individuals were excluded. Online data
359 collection procedures were approved by the Ethical & Independent Review Services (E&I
360 Review), an AAHRPP-accredited private institutional review board
361 (<http://www.eandireview.com>).

362 ***Genetics of Personality Consortium:*** The Genetics of Personality Consortium (GCP) is a large
363 body of cooperation concerning GWAS on personality. We used summary statistics of
364 neuroticism from the first GCP personality meta-analysis (GPC1,
365 <http://www.tweelingenregister.org/GPC/>)⁹, on 10 discovery cohorts (SardiNIA, NTR/NESDA,

366 ERF, SAGE, HBCS, NAG, IRPG, QIMR, LBC1936, BLSA, EGPOT), including in total
367 N=17,375 participants of European descent. All included studies were approved by local ethic
368 committees, and informed consent was obtained from all participants.

369 ***Psychiatric Genetics Consortium:*** The Psychiatric Genetics Consortium (PGC) unites
370 investigators worldwide to conduct genetic meta- and mega-analyses for psychiatric disorders.
371 We used summary statistics from the latest published PGC meta-analysis on depression
372 (<http://www.med.unc.edu/pgc/results-and-downloads>)²⁷, which included data from 8 cohorts
373 (Bonn/Mannheim, GAIN, GenRED, GSK, MDD2000, MPIP, RADIANT, STAR*D), covering
374 N=18,759 participants of European descent. All included studies were approved by local ethic
375 committees, and informed consent was obtained from all participants.

376

377 ***Phenotype assessment – Neuroticism***

378 ***UK Biobank:*** Neuroticism was measured with 12 dichotomous (yes/no) items of the Eysenck
379 Personality Questionnaire Revised Short Form (EPQ-RS)²⁹, using a touchscreen-questionnaire
380 at the UKB assessment centers (**Supplementary Information 1.1**). Participants with valid
381 responses to <10 items were excluded from analyses. A weighted neuroticism sum score was
382 calculated by adding up individual valid item responses, and dividing that sum by the total
383 number of valid responses. Scores on 4 EPQ-RS items (i.e., “Do you often feel lonely?”, “Do
384 you ever feel ‘just miserable’ for no reason?”, “Does your mood often go up and down?”, and
385 “Do you often feel ‘fed-up’?”) were summed to obtain scores for the cluster *depressed affect*.
386 Similarly, scores on 4 other EPQ-RS items (i.e., “Are you a worried?”, “Do you suffer from
387 nerves?”, “Would you call yourself a nervous person?”, and “Would you call yourself tense or
388 highly strung”) were summed to obtain scores for the cluster *worry*. In the item-cluster
389 analyses, only participants with complete scores on all 4 items were included, resulting in
390 N=357,957 and N=348,219 for *depressed affect* and *worry*, respectively.

391 **23andMe**: Neuroticism was operationalized as of the sum of 8 neuroticism items (5-point
392 Likert scale; ‘Disagree strongly’ to ‘Agree strongly’) from the Big Five Inventory (BFI^{30,31}),
393 as obtained in an online survey. Only participants with valid responses to all items were
394 included in the analyses (**Supplementary Information 1.2**).

395 **Genetic Personality Consortium**: All 10 cohorts included in the first meta-analysis of the GPC
396 used sums of the scores on 12 items (5-point Likert scale; ‘Strongly disagree’ to ‘Strongly
397 agree’) of the NEO-FFI³² to measure neuroticism. If <4 item scores were missing, data on
398 invalid items were imputed by taking an individual’s average score on valid items. Participants
399 were excluded from analyses if they had invalid scores on >3 items⁹ (**Supplementary**
400 **Information 1.3**).

401

402 ***Phenotype assessment - Depression***

403 **UK Biobank**: Depression was operationalized by adding up the scores on two continuous items
404 (“Over the past two weeks, how often have you felt down, depressed or hopeless?”, “Over the
405 past two weeks, how often have you had little interest or pleasure in doing things?”; both
406 evaluated on a 4-point Likert scale; ‘Not at all’ to ‘Nearly every day’), resulting in a continuous
407 depression score (as used previously¹²). Only participants with scores on both items were
408 included in the analyses, resulting in N=362,696 (**Supplementary Information 1.4**).

409

410 **23andMe**: This concerns a case-control sample. Four self-report survey items were used to
411 determine case-control status. Cases were defined as replying affirmatively to at least one of
412 these questions, and not replying negatively to previous ones. Controls replied negatively to at
413 least one of the questions, and did not report being diagnosed with depression on previous ones
414 (**Supplementary Information 1.5**).

415 ***Psychiatric Genetics Consortium:*** This concerns a case-control sample. Cases had a DSM-IV
416 lifetime (sometimes (early onset) recurrent) major depressive disorder (MDD) diagnosis, either
417 established through structured diagnostic interviews or clinician-administered DSM-IV
418 checklists. Most cases were ascertained from clinical sources, while controls were randomly
419 selected from population resources and screened for lifetime history of MDD²⁷
420 (**Supplementary Information 1.6**).

421

422 ***Genotyping and imputation***

423 ***UK Biobank - Neuroticism:*** We used genotype data released by the UKB in July 2017. The
424 genotype data collection and processing are described in detail by the responsible UKB
425 group¹⁴. In short, 489,212 individuals were genotyped on two customized SNP arrays (the UK
426 BiLEVE Axiom array (n=50,520) and UK Biobank Axiom array (n=438,692)), covering
427 812,428 unique genetic markers (95% overlap in SNP content). After quality control
428 procedures¹⁴, 488,377 individuals and 805,426 genotypes remained. Genotypes were phased
429 and imputed by the coordinating team to approximately 96 million genotypes using a combined
430 reference panel including the Haplotype Reference Consortium and the UK10K haplotype panel.
431 Imputed and quality controlled genotype data was available for 487,422 individuals and
432 92,693,895 genetic variants. As recommended by the UKB team, variants imputed from the
433 UK10K reference panel were removed from the analyses due to technical errors in the
434 imputation process.

435 In our analyses, only individuals from European descent (based on genetic principal
436 components) were included. Therefore principal components from the 1000 Genomes
437 reference populations³³ were projected onto the called genotypes available in UK Biobank.
438 Subjects were identified as European if their projected principal component score was closest
439 (based on the Mahalanobis distance) to the average score of the European 1000 Genomes

440 sample³⁴. European subjects with a Mahalanobis distance > 6 S.D. were excluded. In addition,
441 participants were excluded based on withdrawn consent, UKB provided relatedness (subjects
442 with most inferred relatives, 3rd degree or closer, were removed until no related subjects were
443 present), discordant sex, sex aneuploidy. After selecting individuals based on available
444 neuroticism sum-score and active consent for participation, 372,903 individuals remained for
445 the analyses.

446 To correct for population-stratification, 30 principal components were calculated on the subset
447 of QC-ed unrelated European subjects based on 145,432 independent ($r^2 < 0.1$) SNPs with
448 MAF > 0.01 and INFO = 1 using FlashPC³⁵. Subsequently, imputed variants were converted to
449 hard call using a certainty threshold of 0.9. Multi-allelic SNPs, indels, and SNPs without unique
450 rs id were excluded, as well as SNPs with a low imputation score (INFO score < 0.9), low minor
451 allele frequency (MAF < 0.0001) and high missingness (> 0.05). This resulted in a total of
452 10,847,151 SNPs used for downstream analysis.

453

454 **UK Biobank – Depression:** Similar genotyping/imputation/filtering procedures as described
455 above for the UKB neuroticism GWAS were followed for the UKB depression GWAS,
456 resulting in N=362,696.

457 **Other samples:** Summary statistics were used for 23andMe and PGC. Genotyping and
458 imputation of these samples are described in detail elsewhere (23andMe depression²⁸; PGC
459 depression²⁷).

460

461 **Genome-wide association analyses**

462 **UK Biobank - Neuroticism:** Genome-wide association analyses were performed in PLINK^{36,37},
463 using a linear regression model of additive allelic effects with age, sex, townsend deprivation
464 index, genotype array, and 10 genetic European-based principal components as covariates.

465 **UK Biobank – Depression, depressed affect, worry:** The settings, covariates, and exclusion
466 criteria for the UKB depression, UKB *depressed affect*, and UKB *worry* GWAS were the same
467 as described above for UKB neuroticism GWAS, with 10,847,151 SNPs remaining after all
468 exclusion steps.

469 **Other samples:** Summary statistics were used for 23andMe, GPC and PGC. Details on the
470 genome-wide association analyses of these samples can be found elsewhere (23andMe
471 neuroticism¹⁰; 23andMe depression²⁸; GPC neuroticism⁹; PGC depression²⁷).

472

473 **Meta-analysis**

474 **Neuroticism:** Meta-analysis of the neuroticism GWAS in UKB, 23andMe, and GPC was
475 carried out in METAL¹⁶. The meta-analysis was performed on the *P*-value of each SNP using
476 a sample size-weighted fixed-effects analysis. Bonferroni correction was applied to correct for
477 multiple testing. The genetic signal correlated strongly between the three samples (r_g range:
478 0.83 – 1.07; **Supplementary Table 1**), supporting the decision to meta-analyze.

479

480 **Depression:** Meta-analysis of the depression GWAS in UKB, 23andMe and PGC was carried
481 out in METAL¹⁶. As the UKB GWAS concerned a continuous operationalization of the
482 depression phenotype, while 23andMe and PGC used case-control phenotypes, the odds ratio
483 from the 23andMe and PGC summary statistics were converted to log odds, reflecting the
484 direction of the effect. The meta-analysis was then performed on the *P*-value of each SNP using
485 a sample size-weighted fixed-effects analysis. Bonferroni correction was applied to correct for
486 multiple testing. Genetic correlations between the three samples were moderate to strong (r_g
487 range: 0.61 – 0.80; **Supplementary Table 22**).

488

489

490 *Functional Annotation*

491 Functional annotation was performed using FUMA¹⁷ (<http://fuma.ctglab.nl/>), an online
492 platform for functional mapping of genetic variants. We first defined independent significant
493 SNPs which have a genome-wide significant P -value (5×10^{-8}) and are independent at $r^2 < 0.6$.
494 Lead SNPs were defined by retaining those independent significant SNPs that were
495 independent from each other at $r^2 < 0.1$ (based on LD information from UK Biobank genotypes;
496 see **Supplementary Information** for a more detailed explanation). Subsequently, risk loci
497 were defined by merging lead SNPs that physically overlapped or whose LD blocks were closer
498 than 250kb apart. As a result, when analyzing multiple phenotypes, as in the current study, the
499 same locus may be discovered for different phenotypes, whilst different lead SNPs are
500 identified.

501 We selected all SNPs with $r^2 > 0.6$ with one of the independent significant SNPs, a P -value
502 lower than 0.05 and minor allele frequency (MAF) higher than 0.0001 for annotations.
503 Functional consequences for all independent significant SNPs and SNPs in LD with them were
504 obtained by performing ANNOVAR gene-based annotation using Ensembl genes. In addition,
505 CADD scores (indicating the deleteriousness of SNP, with scores > 12.37 seen as likely
506 deleterious)³⁸ and RegulomeDB scores (with lower scores indicating a higher probability of
507 having a regulatory function) were annotated to SNPs by matching chromosome, position,
508 reference and alternative alleles.

509

510 *Gene-mapping*

511 SNPs in genomic risk loci that were GWS or were in LD (> 0.6) with one of the independent
512 GWS SNPs were mapped to genes in FUMA¹⁹ using three strategies:

- 513 1. Positional mapping maps SNPs to genes based on the physical distances (i.e., within
514 10kb window) from known protein coding genes in the human reference assembly
515 (GRCh37/hg19).
- 516 2. eQTL mapping maps SNPs to genes with which they show a significant eQTL
517 association (i.e. the expression of that gene is associated with allelic variation at the
518 SNP). eQTL mapping uses information from 3 data repositories (GTEx, Blood eQTL
519 browser BIOS QTL browser, and is based on cis-eQTLs which can map SNPs to genes
520 up to 1Mb apart. A false discovery rate (FDR) of 0.05 was applied to define significant
521 eQTL associations.
- 522 3. Chromatin interaction mapping was performed to map SNPs to genes based on a
523 significant chromatin interaction between a genomic region in a risk locus and promoter
524 regions of genes (250bp up and 500bp downstream of transcription start site (TSS)).
525 Chromatin interaction mapping can involve long-range interactions as it does not have
526 a distance boundary as in eQTL mapping. FUMA currently contains Hi-C data of 14
527 tissue types from the study of³⁹. Since chromatin interactions are often defined in a
528 certain resolution, such as 40kb, an interacting region may span multiple genes. All
529 SNPs within these regions would be mapped by this method to genes in the
530 corresponding interaction region. To further prioritize candidate genes from chromatin
531 interaction mapping, we integrated predicted enhancers and promoters in 111 tissue/cell
532 types from the Roadmap Epigenomics Project⁴⁰; chromatin interactions are selected in
533 which one region involved in the interaction overlaps with predicted enhancers and the
534 other region overlaps with predicted promoters in 250bp up- and 500bp downstream of
535 TSS site of a gene. We used a FDR of 1×10^{-5} to define significant interactions.
- 536
- 537

538 ***Gene-based analysis***

539 A genome-wide gene association analysis (GWGAS) can identify genes in which multiple
540 SNPs show moderate association to the phenotype of interest without reaching the stringent
541 genome-wide significance level. At the same time, as a GWGAS takes all SNPs within a gene
542 into account, a gene harbouring a genome-wide significant SNP may not be implicated by a
543 GWGAS analyses when multiple other SNPs within that gene show only very weak association
544 signal. The *P*-values from the SNP-based GWAS meta-analyses for neuroticism and
545 depression, and the GWAS for *depressed affect* and *worry*, were used as input for the genome-
546 wide gene association analysis (GWGAS) in MAGMA (<http://ctg.cncr.nl/software/magma>)²³,
547 and all 19,427 protein-coding genes from the NCBI 37.3 gene definitions were used. We
548 annotated all SNPs in our GWA (meta-) analyses to these genes, resulting in 18,187, 18,187,
549 18,182, and 18,182 genes that were represented by at least one SNP in the neuroticism meta-
550 analysis, the depression meta-analysis, the *depressed affect* GWAS, and the *worry* GWAS,
551 respectively. We included a window around each gene of 2 kb before the transcription start site
552 and 1 kb after the transcription stop site. Gene association tests were performed taking into
553 account the LD between SNPs, and a stringent Bonferroni correction was applied to correct for
554 multiple testing (0.05/number of genes tested: $P < 2.75 \times 10^{-6}$).

555

556 ***Gene-set analysis***

557 We used MAGMA²³ to test for association of predefined gene-sets with neuroticism,
558 depression, *depressed affect*, and *worry*. A total of 7,246 gene-sets were derived from several
559 resources, including BioCarta, KEGG, Reactome⁴¹ and GO. All gene-sets were obtained from
560 the MsigDB version 5.2 (<http://software.broadinstitute.org/gsea/msigdb/collections.jsp>). In
561 addition, we performed gene-set analysis on 53 tissue expression profiles obtained from the
562 GTEx portal (<https://www.gtexportal.org/home/>), and 24 cell-type specific expression profiles.

563 Definition and calculation of gene-sets for cell-type specific expression is described in detail
564 elsewhere^{26,42}. Briefly, brain cell-type expression data was drawn from scRNAseq data from
565 mouse brain²⁶. For each gene, the value for each cell-type was calculated by dividing the mean
566 Unique Molecular Identifier (UMI) counts for the given cell type by the summed mean UMI
567 counts across all cell types²⁶. Associations between gene-wise P -values from the meta-analysis
568 and cell-type specific gene expression were calculated using MAGMA²³, by grouping genes
569 into 40 equal bins by specificity of expression, and regressing bin-membership on gene-wise
570 association with neuroticism in the meta-analysis. Results were considered significant if the
571 association P -values were smaller than the relevant Bonferroni threshold.
572 For all gene-sets we computed competitive P -values, which result from the test whether the
573 combined effect of genes in a gene-set is significantly larger than the combined effect of a same
574 number of randomly selected genes (in contrast, self-contained P -values result from testing
575 against the null hypothesis of no effect). We only report competitive P -values, which are more
576 conservative compared to self-contained P -values. Competitive P -values were Bonferroni
577 corrected ($\alpha=0.05/7,323=6.83\times 10^{-6}$).
578 Conditional gene-set analyses were performed with MAGMA as a secondary analysis to test
579 whether each observed enriched cell-type was independent of all others. Full details of the
580 method implemented are provided in ²⁶.

581

582 ***Genetic correlations***

583 Genetic correlations (r_g) were computed using LDscore regression^{17,18}
584 (<https://github.com/bulik/ldsc>). The significance of the genetic correlations of neuroticism,
585 depression, *depressed affect* and *worry* with 35 behavioral, social and (mental) health
586 phenotypes for which summary statistics were available was determined while correcting for

587 multiple testing through a stringent Bonferroni corrected threshold of $P < 0.05 / (4 \times 35)$ (3.6×10^{-4}).

589

590 ***Partitioned heritability***

591 To investigate the relative contribution to the overall heritability of SNPs annotated to 22
592 specific genomic categories, we partitioned SNP heritability by binary annotations using
593 stratified LD score regression^{20,43}. Information about binary SNP annotations were obtained
594 from the LD score website (<https://github.com/bulik/ldsc>). Enrichment results reflect the X-
595 fold increase in h^2 proportional to the number of SNPs (e.g., enrichment=13.79 for SNPs in
596 conserved regions implies that a 13,79-fold increase in h^2 is carried by SNPs in these region,
597 corrected for the proportion of SNPs in these regions compared to all tested SNPs).

598

599 ***Polygenic risk scoring***

600 To test the predictive accuracy (ΔR^2) of our meta-analytic results, we calculated a polygenic
601 risk score (PGS) based on the SNP effect sizes of the current analysis. As independent samples
602 we used holdout samples; we removed 3,000 individuals from the discovery sample (UKB
603 only, as we only had access to raw data from this sample) and reran the genome-wide analyses.
604 We repeated this three times, to create 3 randomly drawn, independent hold-out samples. Next,
605 we calculated a PGS on the individuals in each of the 3 holdout samples. PGS were calculated
606 using LDpred²² and PRSice²¹ (clumping followed by P -value thresholding).

607 For LDpred, PGS were calculated based on different LDpred priors ($P_{LDpred} = 0.01, 0.05, 0.1,$
608 $0.5, 1$ and infinitesimal). The explained variance (R^2) was derived from the linear model, using
609 the neuroticism summary score as the outcome, while correcting for age, gender, array, batch
610 and genetic principal components.

611

612 ***Data availability***

613 Our policy is to make genome-wide summary statistics (sumstats) publically available.

614 Sumstats from our neuroticism meta-analysis, our depression meta-analysis, and the GWA
615 analyses for *depressed affect* and *worry* are available for download at <https://ctg.cncr.nl/>.

616 Note that our freely available meta-analytic sumstats concern results excluding the 23andMe
617 sample. This is a non-negotiable clause in the 23andMe data transfer agreement, intended to
618 protect the privacy of the 23andMe research participants. To fully recreate our meta-analytic
619 results for neuroticism: (a) obtain Lo et al. (2016) sumstats from 23andMe (see below); (b)
620 conduct a meta-analysis of our sumstats with the Lo et al. sumstats. To fully recreate our meta-
621 analytic results for depression: (a) obtain Hyde et al. (2016) sumstats from 23andMe (see
622 below); (b) conduct a meta-analysis of our sumstats with the Hyde et al. sumstats.

623 23andMe participant data are shared according to community standards that have been
624 developed to protect against breaches of privacy. Currently, these standards allow for the
625 sharing of summary statistics for at most 10,000 SNPs. The full set of summary statistics can
626 be made available to qualified investigators who enter into an agreement with 23andMe that
627 protects participant confidentiality. Interested investigators should contact David Hinds
628 (dhinds@23andme.com) for more information.

629

630 **References**

- 631 1. Kendler, K. S. & Myers, J. The genetic and environmental relationship between major
632 depression and the five-factor model of personality. *Psychol. Med.* **40**, 801–806 (2010).
- 633 2. Middeldorp, C. M. *et al.* The Association of Personality with Anxious and Depressive
634 Psychopathology. *Biol. Personal. Individ. Differ.* 251–272 (2006).
- 635 3. Hettema, J. M., Neale, M. C., Myers, J. M., Prescott, C. A. & Kendler, K. S. A
636 Population-Based Twin Study of the Relationship Between Neuroticism and
637 Internalizing Disorders. *Am. J. Psychiatry* **163**, 857–864 (2006).
- 638 4. Hayes, J. F. *et al.* Association of Late Adolescent Personality With Risk for Subsequent
639 Serious Mental Illness Among Men in a Swedish Nationwide Cohort Study. *JAMA*
640 *Psychiatry* **54**, 948–963 (2017).
- 641 5. Smeland, O. B. *et al.* Identification of genetic loci shared between schizophrenia and the
642 Big Five personality traits. *Sci. Rep.* **7**, 1–9 (2017).
- 643 6. Van Os, J. & Jones, P. B. Neuroticism as a risk factor for schizophrenia. *Psychol. Med.*
644 **31**, 1129–1134 (2001).
- 645 7. Genetics of Personality Consortium. Meta-analysis of Genome-wide Association
646 Studies for Neuroticism, and the Polygenic Association With Major Depressive
647 Disorder. *JAMA Psychiatry* **72**, 642–650 (2015).
- 648 8. Terracciano, A. *et al.* Genome-wide association scan for five major dimensions of
649 personality. *Mol. Psychiatry* **15**, 647–656 (2010).
- 650 9. Moor, M. H. M. De *et al.* Meta-analysis of genome-wide association studies for
651 personality. *Mol. Psychiatry* **17**, 337–349 (2012).
- 652 10. Lo, M. *et al.* Genome-wide analyses for personality traits identify six genomic loci and
653 show correlations with psychiatric disorders. *Nat. Genet.* (2016). doi:10.1038/ng.3736
- 654 11. Smith, D. J. *et al.* Genome-wide analysis of over 106 000 individuals identifies 9

- 655 neuroticism-associated loci. *Mol. Psychiatry* **21**, 1–9 (2016).
- 656 12. Okbay, A. *et al.* Genetic variants associated with subjective well-being, depressive
657 symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* **48**,
658 624–636 (2016).
- 659 13. Nagel, M., Watanabe, K., Stringer, S., Posthuma, D. & Van der Sluis, S. Item-level
660 Analyses Reveal Genetic Heterogeneity in Neuroticism. Manuscript submitted for
661 publication (2017).
- 662 14. Bycroft, C. *et al.* Genome-wide genetic data on ~500,000 UK Biobank participants.
663 *bioRxiv* (2017). at <<http://biorxiv.org/content/early/2017/07/20/166298.abstract>>
- 664 15. Eriksson, N. *et al.* Web-based, participant-driven studies yield novel genetic
665 associations for common traits. *PLoS Genet.* **6**, 1–20 (2010).
- 666 16. Willer, C. J., Li, Y., Abecasis, G. R. & Overall, P. METAL: fast and efficient meta-
667 analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
- 668 17. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from
669 polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
- 670 18. Bulik-Sullivan, B. K. *et al.* An atlas of genetic correlations across human diseases and
671 traits. *Nat. Genet.* **47**, 1–9 (2015).
- 672 19. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. FUMA: Functional
673 mapping and annotation of genetic associations. *bioRxiv* (2017). at
674 <<http://biorxiv.org/content/early/2017/02/20/110023.abstract>>
- 675 20. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-
676 wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
- 677 21. Euesden, J., Lewis, C. M. & O’Reilly, P. F. PRSice: Polygenic Risk Score software.
678 *Bioinformatics* **31**, 1466–1468 (2015).
- 679 22. Vilhjálmsón, B. J. *et al.* Modeling Linkage Disequilibrium Increases Accuracy of

- 680 Polygenic Risk Scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
- 681 23. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-
682 Set Analysis of GWAS Data. *PLoS Comput. Biol.* **11**, 1–19 (2015).
- 683 24. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for
684 interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102**, 15545–15550
685 (2005).
- 686 25. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue
687 gene regulation in humans. *Science (80-.)*. **348**, 648–660 (2015).
- 688 26. Skene, N. G. *et al.* Genetic Identification Of Brain Cell Types Underlying
689 Schizophrenia. *bioRxiv* (2017). at
690 <<http://biorxiv.org/content/early/2017/06/02/145466.abstract>>
- 691 27. Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium. A
692 mega-analysis of genome-wide association studies for major depressive disorder. *Mol.*
693 *Psychiatry* **18**, 497–511 (2013).
- 694 28. Hyde, C. L. *et al.* Identification of 15 genetic loci associated with risk of major
695 depression in individuals of European descent. *Nat. Genet.* **48**, 1031–1036 (2016).
- 696 29. Eysenck, B. G., Eysenck, H. J. & Barrett, P. A Revised Version of the Psychoticism
697 Scale. *Pers. Individ. Dif.* **6**, 21–29 (1985).
- 698 30. John, O. P. & Srivastava, S. The Big Five trait taxonomy: History, measurement, and
699 theoretical perspectives. *Handb. Personal. Theory Res.* **2**, 102–138 (1999).
- 700 31. Soto, C. J. & John, O. P. Ten facet scales for the Big Five Inventory: Convergence with
701 NEO PI-R facets, self-peer agreement, and discriminant validity. *J. Res. Pers.* **43**, 84–
702 90 (2009).
- 703 32. Costa, P. & McCrae. *Professional Manual: Revised NEO Personality Inventory (NEO-*
704 *PI-R) and NEO Five-Factor-Inventory (NEO-FFI)*. (Psychological Assessment

- 705 Resources: Odessa, FL, 1992).
- 706 33. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74
707 (2015).
- 708 34. Webb, B. T. *et al.* Molecular Genetic Influences on Normative and Problematic Alcohol
709 Use in a Population-Based Sample of College Students. *Front. Genet.* **8**, 1–11 (2017).
- 710 35. Abraham, G. & Inouye, M. Fast Principal Component Analysis of Large-Scale Genome-
711 Wide Data. *PLoS One* **9**, 1–5 (2014).
- 712 36. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based
713 linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
- 714 37. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer
715 datasets. *Gigascience* **4**, 1 (2015).
- 716 38. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of
717 human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
- 718 39. Schmitt, A. D. *et al.* A Compendium of Chromatin Contact Maps Reveals Spatially
719 Active Regions in the Human Genome. *Cell Rep.* **17**, 2042–2059 (2016).
- 720 40. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human
721 epigenomes. *Nature* **518**, (2015).
- 722 41. Croft, D. *et al.* The Reactome pathway knowledgebase. *Nucleic Acids Res.* **42**, D472–
723 D477 (2014).
- 724 42. Coleman, J. *et al.* Functional consequences of genetic loci associated with intelligence
725 in a meta-analysis of 87,740 individuals. *bioRxiv* (2017). at
726 <<http://biorxiv.org/content/early/2017/07/31/170712.abstract>>
- 727 43. Gazal, S. *et al.* Linkage disequilibrium dependent architecture of human complex traits
728 reveals action of negative selection. *bioRxiv* (2016). at
729 <<http://biorxiv.org/content/early/2016/10/19/082024.abstract>>

730