LETTERS
https://doi.org/10.1038/s41588-018-0151-7
nature genetics

Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways

Mats Nagel[1,2,11], Philip R. Jansen[1,3,11], Sven Stringer[1], Kyoko Watanabe[1], Christiaan A. de Leeuw[1], Julien Bryois[4], Jeanne E. Savage[1], Anke R. Hammerschlag[1], Nathan G. Skene[5], Ana B. Muñoz-Manchado[5], 23andMe Research Team[6], Tonya White[3], Henning Tiemeier[3,7], Sten Linnarsson[5], Jens Hjerling-Leffler[5,8], Tinca J. C. Polderman[1], Patrick F. Sullivan[4,9,10], Sophie van der Sluis[1,2,12] and Danielle Posthuma[1,2,12]*

Neuroticism is an important risk factor for psychiatric traits, including depression[1], anxiety[2,3], and schizophrenia[4–6]. At the time of analysis, previous genome-wide association studies[7–12] (GWAS) reported 16 genomic loci associated to neuroticism[10–12]. Here we conducted a large GWAS meta-analysis ($n = 449{,}484$) of neuroticism and identified 136 independent genome-wide significant loci (124 new at the time of analysis), which implicate 599 genes. Functional follow-up analyses showed enrichment in several brain regions and involvement of specific cell types, including dopaminergic neuroblasts ($P = 3.49 \times 10^{-8}$), medium spiny neurons ($P = 4.23 \times 10^{-8}$), and serotonergic neurons ($P = 1.37 \times 10^{-7}$). Gene set analyses implicated three specific pathways: neurogenesis ($P = 4.43 \times 10^{-9}$), behavioral response to cocaine processes ($P = 1.84 \times 10^{-7}$), and axon part ($P = 5.26 \times 10^{-8}$). We show that neuroticism's genetic signal partly originates in two genetically distinguishable subclusters[13] ('depressed affect' and 'worry'), suggesting distinct causal mechanisms for subtypes of individuals. Mendelian randomization analysis showed unidirectional and bidirectional effects between neuroticism and multiple psychiatric traits. These results enhance neurobiological understanding of neuroticism and provide specific leads for functional follow-up experiments.

The meta-analysis of neuroticism comprised data from the UK Biobank study (UKB, full release[14]; $n = 372{,}903$; Methods and Supplementary Figs. 1 and 2), 23andMe, Inc.[15] ($n = 59{,}206$), and the Genetics of Personality Consortium (GPC1[9]; $n = 17{,}375$; Methods) ($n = 449{,}484$ in total). In all of the samples, neuroticism was measured through (digital) questionnaires (Methods and Supplementary Note). To achieve optimal power, SNP associations were subjected to meta-analysis using METAL[16], with weighting by sample size (Methods). We chose to perform meta-analysis on the available samples rather than use a two-stage discovery–replication strategy because Skol et al.[17] showed that this is almost always more powerful, even though less correction for multiple testing is required in the replication stage.

The quantile–quantile plot of the genome-wide meta-analysis on 449,484 subjects and 14,978,477 SNPs showed inflation (linkage disequilibrium score regression (LDSC)[18]: $\lambda_{GC} = 1.65$, mean χ^2 statistic = 1.91; Fig. 1a and Supplementary Table 1), yet the LDSC intercept (1.02; standard error (s.e.) = 0.01) and ratio (2.1%) both indicated that the inflation was largely due to true polygenicity and the large sample size[19]. The λ_{GC} value of 1.65 is consistent with values observed in recent large-sample GWAS ($n > 100{,}000$) for diverse and polygenic traits (Supplementary Note). The LDSC SNP-based heritability (h^2_{SNP}) of neuroticism was 0.100 (s.e. = 0.003). The GWAS meta-analysis identified 9,745 genome-wide significant SNPs ($P < 5 \times 10^{-8}$), of which 157 and 2,414 were located in known associated inversion regions on chromosomes 8 and 17[10–12], respectively (Fig. 1b and Supplementary Fig. 3; see Supplementary Table 2 for cohort-specific information). We used FUMA[20], a tool to functionally map and annotate results from GWAS (Methods), and extracted 170 independent lead SNPs (158 new; see the Methods for definition of lead SNPs) that mapped to 136 independent genomic loci (124 new at the time of analysis) (Methods, Supplementary Tables 3–8, and Supplementary Note). Of all the lead SNPs, 4 were in exonic regions, 88 were in intronic regions, and 52 were in intergenic regions. Of the 17,794 SNPs in high linkage disequilibrium (LD) with one of the independent significant SNPs (see the Methods for definition), most were intronic (9,147; 51.4%) or intergenic (5,460; 30.7%), and 3.8% were annotated as potentially having a functional impact, with 0.9% (155 SNPs) being exonic (Fig. 1c and Supplementary Table 9; see Supplementary Tables 10 and 11 for an overview of the chromatin state and regulatory functions of these SNPs). Of these 155 SNPs, 70 were exonic nonsynonymous (ExNS) (Table 1 and Supplementary Table 12). The ExNS SNP with

[1]Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. [2]Department of Clinical Genetics, Section of 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]A list of members and affiliations appears at the end of the paper. [7]Department of Social and Behavioral Science, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [8]UCL Institute of Neurology, Queen Square, London, UK. [9]Department of Genetics, University of North Carolina, Chapel Hill, NC, USA. [10]Department of Psychiatry, University of North Carolina, Chapel Hill, NC, USA. [11]These authors contributed equally: Mats Nagel, Philip R. Jansen. [12]These authors jointly supervised this work: Sophie van der Sluis, Danielle Posthuma. *e-mail: d.posthuma@vu.nl

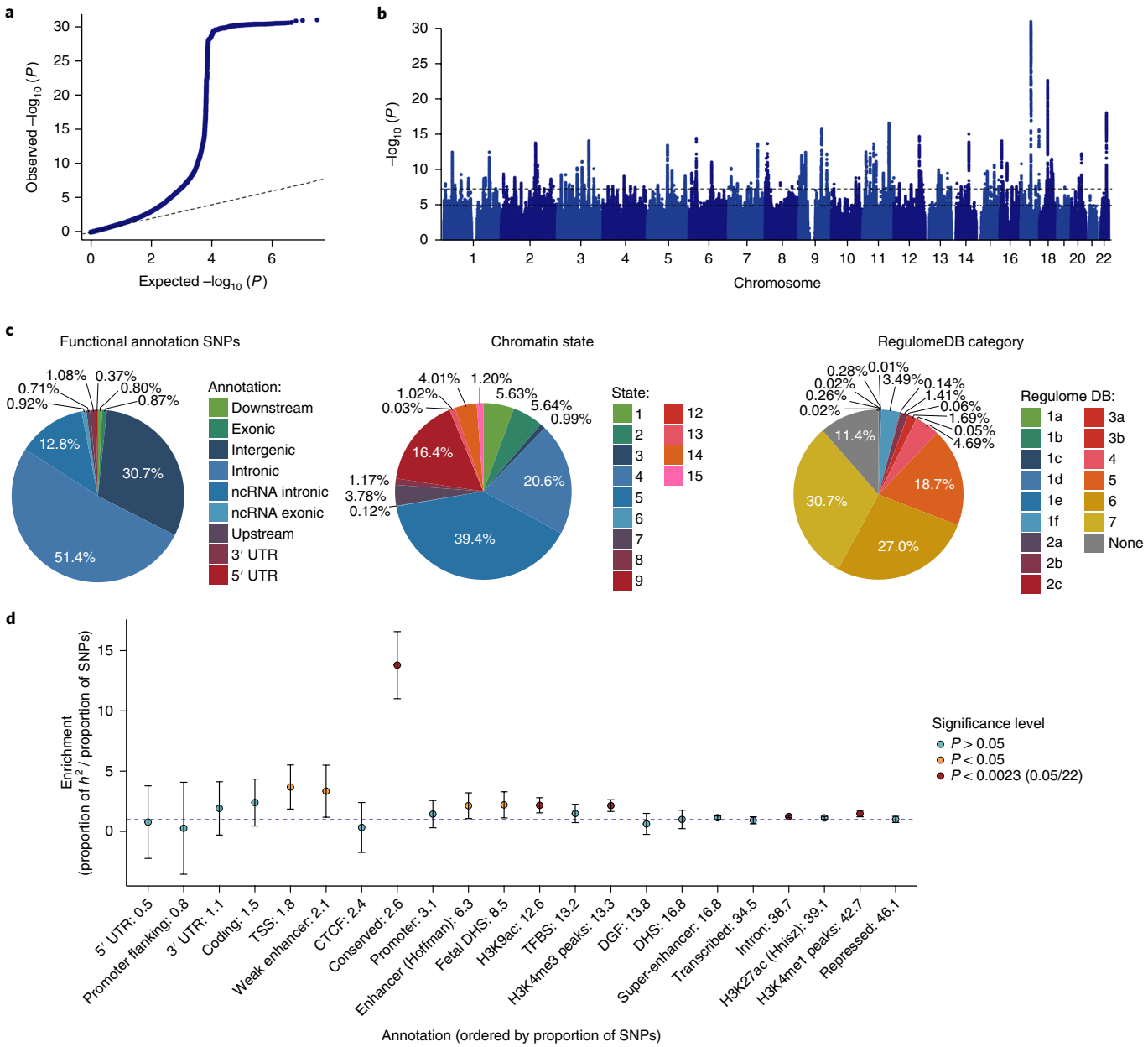

Fig. 1 | SNP-based associations with neuroticism in the GWAS meta-analysis. a, Quantile–quantile plot of the SNP-based associations with neuroticism ($n = 449{,}484$ individuals). SNP *P* values were computed in METAL using a two-sided, sample-size-weighted *z*-score method. **b**, Manhattan plot showing the $-\log_{10}$-transformed *P* value of each SNP on the *y* axis and base-pair positions along the chromosomes on the *x* axis ($n = 449{,}484$ individuals). SNP *P* values were computed in METAL using a two-sided, sample-size-weighted *z*-score method. The upper dashed line indicates genome-wide significance ($P < 5 \times 10^{-8}$), and the lower dashed line shows the threshold for suggestive associations ($P < 1 \times 10^{-5}$). **c**, Pie charts showing the distribution of functional consequences of SNPs in LD with genome-wide significant lead SNPs in the meta-analysis, the minimum chromatin state across 127 tissue and cell types, and the distribution of RegulomeDB score (a categorical score between 1a and 7, indicating biological evidence of a SNP being a regulatory element, with a low score denoting a higher likelihood of a SNP being regulatory). **d**, Heritability enrichment of 22 functional SNP annotations calculated with stratified LD Score regression (summary statistics of the meta-analysis of neuroticism were used as input for this analysis). The circles signify the estimated enrichment, whereas the dashed line indicates enrichment of 1. Error bars represent 95% confidence intervals. TSS, transcription start site; CTCF, CCCTC binding factor; DGF, digital genomic footprint; TFBS, transcription factor binding site; DHS, DNase I hypersensitivity site.

the highest combined annotation-dependent depletion[21] (CADD) score (which indicates the likelihood of the SNP being deleterious; Methods) was rs17651549 (CADD score of 34), located on chromosome 17 in exon 6 of *MAPT*, with a GWAS *P* value of 1.11×10^{-28}, in high LD with the lead SNP in that region ($r^2 = 0.97$). rs17651549 is a missense mutation (c.1108 C > T:p.Arg370Trp) that leads to an arginine-to-tryptophan change with allele frequencies matching the inversion in that region. The ancestral C allele is associated with a lower neuroticism score (see Table 1 and Supplementary Table 12 for a detailed overview of all functional variants in genomic risk loci).

Stratified LDSC[22] (Methods) showed significant enrichment for h^2 of SNPs located in conserved regions (enrichment = 13.79, $P = 5.14 \times 10^{-16}$), intronic regions (enrichment = 1.24,

Table 1 | Exonic nonsynonymous variants in the genomic loci associated with neuroticism and in LD ($r^2 > 0.6$) with one of the independent genome-wide significant SNPs

rsID	Exon	Gene	A1	MAF	GWAS *P*	*z* score	r^2	Independent significant SNP	Locus	CADD	RDB	Minimum chromatin state
rs41266050	14	*RABGAP1L*	T	0.25	5.65×10^{-6}	−4.54	0.84	rs7536102	6	2.85	7	5
rs34605051	10	*KDM3A*	T	0.16	1.64×10^{-7}	−5.24	0.99	rs11127043	14	13.85	4	4
rs2073498	3	*RASSF1*	A	0.11	2.71×10^{-8}	5.56	0.98	rs6776145	25	19.43	7	3
rs4434138	62	*STAB1*	A	0.46	3.10×10^{-7}	5.12	0.93	rs2015971	26	5.48	5	4
rs66782572	1	*NT5DC2*	A	0.46	2.65×10^{-6}	4.70	0.67	rs2015971	26	0.00	4	1
rs11177	3	***GNL3***	A	0.38	2.49×10^{-7}	−5.16	0.75	rs2015971	26	22.90	1d	1
rs2289247	11	***GNL3***	A	0.41	2.28×10^{-6}	−4.73	0.65	rs2015971	26	12.82	1f	3
rs6617	1	*SPCS1*	C	0.41	1.96×10^{-6}	4.76	0.65	rs2015971	26	0.00	1f	1
rs1029871	5	*NEK4*	C	0.38	2.29×10^{-7}	−5.17	0.75	rs2015971	26	24.10	1f	2
rs678	12	***ITIH1***	A	0.36	1.44×10^{-6}	4.82	0.65	rs2015971	26	25.90	1f	2
rs1042779	12	***ITIH1***	A	0.37	6.09×10^{-6}	4.52	0.63	rs2015971	26	0.15	1f	2
rs198844	1	*HIST1H1T*	C	0.47	3.07×10^{-8}	−5.54	0.98	rs198825	45	0.10	1f	5
rs200484	1	*HIST1H2BL*	A	0.13	1.53×10^{-7}	5.25	0.61	rs200965	46	2.68	1f	1
rs240780	39	*ASCC3*	C	0.43	3.01×10^{-8}	5.54	0.96	rs240769	49	19.95	7	4
rs3173615	6	*TMEM106B*	C	0.41	2.08×10^{-8}	−5.61	0.67	rs11509880	51	21.40	6	4
rs11765552	11	*LMTK2*	A	0.46	7.68×10^{-8}	5.38	0.98	rs34320230	55	12.24	6	4
rs10821128	3	***FAM120AOS***	T	0.33	2.73×10^{-9}	5.95	0.99	rs10821129	71	0.05	4	4
rs41274386	2	***FAM120AOS***	T	0.08	1.10×10^{-7}	5.31	0.66	rs78046549	71	2.36	4	1
rs1055710	1	***FAM120AOS***	A	0.33	1.11×10^{-9}	−6.09	0.99	rs10821129	71	0.05	4	1
rs3816614	33	*LRP4*	T	0.23	5.69×10^{-7}	5.00	0.90	rs7940441	84	22.70	NA	4
rs2030166	5	*NDUFS3*	T	0.35	2.02×10^{-10}	−6.36	0.93	rs11039389	84	3.13	6	4
rs1064608	13	*MTCH2*	C	0.35	1.15×10^{-10}	−6.45	0.93	rs11039389	84	25.40	6	4
rs12286721	13	*AGBL2*	A	0.45	7.81×10^{-8}	−5.37	0.78	rs7107356	84	14.22	1f	5
rs3816605	5	*NUP160*	T	0.46	5.68×10^{-8}	−5.43	0.75	rs7107356	84	6.61	6	4
rs4926	7	*SERPING1*	A	0.27	6.12×10^{-7}	4.99	0.86	rs73480560	85	23.50	5	4
rs11604671	6	***ANKK1***	A	0.49	2.57×10^{-10}	−6.32	0.64	rs2186800	88	1.39	5	4
rs2734849	8	***ANKK1***	A	0.49	8.43×10^{-10}	6.14	0.64	rs2186800	88	0.00	3a	4
rs1800497	8	***ANKK1***	A	0.20	8.45×10^{-6}	4.45	0.69	rs11214607	88	0.81	4	4
rs3825393	30	*MYO1H*	T	0.36	2.95×10^{-7}	−5.13	0.79	rs2111216	94	10.93	1f	4
rs2058804	2	*KCTD10*	A	0.48	1.82×10^{-10}	−6.38	0.76	rs2111216	94	2.11	6	4
rs7298565	12	*UBE3B*	A	0.48	2.24×10^{-10}	6.34	0.76	rs2111216	94	22.70	6	4
rs9593	9	*MMAB*	A	0.48	8.54×10^{-10}	−6.14	0.76	rs2111216	94	0.53	1f	4
rs8007859	7	*EXD2*	T	0.39	2.28×10^{-8}	5.59	0.80	rs1275411	108	3.95	5	4
rs2286913	4	***RPS6KL1***	A	0.37	1.46×10^{-7}	5.26	0.89	rs3213716	110	12.96	5	2
rs7156590	3	***RPS6KL1***	T	0.37	2.79×10^{-7}	5.14	0.86	rs3213716	110	19.46	5	4
rs35755513	-	*CSNK1G1*	T	0.07	2.87×10^{-8}	5.55	1.00	rs35755513	114	23.90	4	1
rs12443627	1	*ENSG00000268863*	C	0.37	1.28×10^{-10}	6.43	0.77	rs3751855	119	3.58	2b	1
rs9938550	7	*HSD3B7*	A	0.37	1.48×10^{-10}	−6.41	0.79	rs3751855	119	0.04	1d	3
rs35713203	1	***ZNF646***	C	0.38	3.67×10^{-11}	−6.62	0.98	rs3751855	119	0.05	2b	3
rs7196726	1	***ZNF646***	A	0.38	1.29×10^{-11}	−6.77	1.00	rs3751855	119	0.00	2b	3
rs7199949	8	*PRSS53*	C	0.38	1.32×10^{-11}	−6.77	1.00	rs3751855	119	0.00	2b	2
rs3803704	3	*CMTR2*	T	0.25	7.14×10^{-8}	−5.39	0.97	rs1424144	121	0.07	6	4
rs3748400	12	*ZCCHC14*	T	0.23	8.83×10^{-9}	−5.75	0.98	rs2042395	122	24.00	5	4
rs12949256	1	*ARHGAP27*	T	0.19	1.47×10^{-23}	10.00	0.73	rs77804065	126	11.97	4	1
rs16940674	6	***CRHR1***	T	0.23	5.24×10^{-29}	11.18	0.97	rs77804065	126	12.86	1f	5
rs16940681	13	***CRHR1***	C	0.23	2.18×10^{-30}	11.46	0.97	rs77804065	126	1.76	4	5

Continued

Table 1 | Exonic nonsynonymous variants in the genomic loci associated with neuroticism and in LD ($r^2 > 0.6$) with one of the independent genome-wide significant SNPs (Continued)

rsID	Exon	Gene	A1	MAF	GWAS *P*	*z* score	r^2	Independent significant SNP	Locus	CADD	RDB	Minimum chromatin state
rs62621252	1	***SPPL2C***	T	0.23	9.05×10^{-31}	−11.53	0.97	rs77804065	126	0.00	5	5
rs242944	1	***SPPL2C***	A	0.44	2.88×10^{-12}	−6.98	1.00	rs242947	126	0.00	5	5
rs62054815	1	***SPPL2C***	A	0.23	1.74×10^{-30}	11.48	0.97	rs77804065	126	0.00	5	5
rs12185233	1	***SPPL2C***	C	0.23	6.76×10^{-29}	11.16	0.96	rs77804065	126	25.60	1f	5
rs12185268	1	***SPPL2C***	A	0.23	4.08×10^{-30}	−11.40	0.97	rs77804065	126	0.00	1f	5
rs12373123	1	***SPPL2C***	T	0.23	7.80×10^{-29}	−11.14	0.97	rs77804065	126	22.70	1f	5
rs12373139	1	***SPPL2C***	A	0.23	2.19×10^{-30}	11.46	0.97	rs77804065	126	0.53	1f	5
rs12373142	1	***SPPL2C***	C	0.22	2.60×10^{-28}	−11.04	0.97	rs77804065	126	0.12	1f	5
rs754512	1	***MAPT***	A	0.23	1.17×10^{-28}	−11.11	0.97	rs77804065	126	2.39	1d	4
rs63750417	6	***MAPT***	T	0.23	4.89×10^{-30}	11.39	0.97	rs77804065	126	8.68	5	4
rs62063786	6	***MAPT***	A	0.23	1.05×10^{-29}	11.32	0.97	rs77804065	126	7.65	5	4
rs62063787	6	***MAPT***	T	0.23	4.57×10^{-30}	−11.39	0.97	rs77804065	126	0.00	5	4
rs17651549	6	***MAPT***	T	0.23	1.11×10^{-28}	11.11	0.97	rs77804065	126	34.00	1f	4
rs10445337	8	***MAPT***	T	0.23	1.41×10^{-28}	−11.09	0.96	rs77804065	126	9.93	1f	4
rs62063857	1	*STH*	A	0.23	3.71×10^{-30}	−11.41	0.97	rs77804065	126	0.00	7	4
rs34579536	15	***KANSL1***	A	0.23	1.92×10^{-30}	−11.47	0.96	rs77804065	126	8.02	3a	3
rs34043286	8	***KANSL1***	A	0.23	3.14×10^{-30}	−11.43	0.97	rs77804065	126	15.71	4	4
rs4969391	14	*BAIAP2*	A	0.16	1.69×10^{-14}	7.67	0.90	rs56084168	128	12.58	4	4
rs2282632	11	***ASXL3***	A	0.50	3.37×10^{-8}	−5.52	0.73	rs10460051	129	1.54	6	4
rs7232237	12	***ASXL3***	A	0.50	1.59×10^{-8}	−5.65	0.84	rs10460051	129	0.00	5	4
rs17522826	1	*TCF4*	A	0.18	2.17×10^{-10}	6.35	0.60	rs10503002	133	14.22	5	1
rs20551	15	*EP300*	A	0.29	3.44×10^{-18}	−8.70	0.98	rs9611519	138	3.23	5	4
rs139431	2	*L3MBTL2*	T	0.37	$9.45E\times10^{-7}$	−4.90	0.63	rs7289932	138	10.26	7	4
rs739134	2	*C22orf46*	T	0.19	4.55×10^{-7}	5.04	0.61	rs761366	138	22.60	1f	4

SNP *P* values and *z* scores were computed in METAL by a weighted *z*-score method (two-sided test). Per-SNP *n* values are reported in Supplementary Table 2 (for genome-wide significant SNPs) and in the publicly available summary statistics. rsID, rs number of the ExNS SNP; Exon, exon in which the SNP is located; Gene, nearest gene; A1, effect allele; MAF, minor allele frequency; GWAS *P*, SNP *P* value in the GWAS meta-analysis; *z* score, *z* score from the GWAS meta-analysis; r^2, maximum r^2 of the SNP with one of the independent significant SNPs; Locus, index of the genomic risk locus; CADD, CADD score; RDB, RegulomeDB score; Minimum chromatin state, minimum chromatin state of the SNP. Results are reported on hg19 coordinates (GRCh37); NA, not available in RegulomeDB (alleles do not match). Genes containing multiple ExNS SNPs are annotated in bold.

$P = 1.27\times10^{-6}$), and trimethylated Lys4 on histone H3 (H3K4me3; enrichment = 2.14, $P = 1.02\times10^{-5}$) and acetylated Lys9 on histone H3 (H3K9ac; enrichment = 2.17, $P = 3.06\times10^{-4}$) regions (Fig. 1d and Supplementary Table 13).

Polygenic scores (PGSs) calculated using PRSice[23] (clumping followed by *P*-value thresholding) and LDpred[24] in three randomly drawn hold-out samples (UKB only, $n = 3{,}000$ each; Methods) explained up to 4.2% ($P = 1.39\times10^{-30}$) of the variance in neuroticism (Supplementary Fig. 4, Supplementary Table 14, and Supplementary Note). Although the current sample size is considered to be large for GWAS and PGSs can be calculated with relatively low standard errors, the variance explained by all SNPs combined in the PGSs was still relatively small, although this was not unexpected given the h^2_{SNP} of 10%. Our current results thus have little predictive power in independent samples, mostly owing to the low average effect sizes of contributing SNPs, and indicate that the genetic architecture of neuroticism is extremely polygenic. We do note that our current meta-analysis did not include possible genetic interactions (as even with the current sample sizes power would be limited) but that adding these in the future may increase the predictive value of PGSs for neuroticism.

We used four strategies to link our SNP results to genes: positional, expression quantitative trait locus (eQTL), and chromatin interaction mapping (Methods) and genome-wide gene-based association study (GWGAS, using MAGMA[25]). GWGAS evaluates the joint association effect of all SNPs within a gene to yield a gene-based *P* value. Basesd on our meta-analysis results, 283 genes were implicated through positional mapping, 369 were implicated through eQTL mapping, and 119 were implicated through chromatin interaction mapping (Fig. 2a and Supplementary Table 15). GWGAS identified 336 genome-wide significant genes ($P < 2.75\times10^{-6}$; Fig. 2b,c, Supplementary Table 16, and Supplementary Note), of which 203 overlapped with genes implicated by FUMA, resulting in 599 unique neuroticism-related genes. Of these, 50 were implicated by all four methods, of which 49 had chromatin interaction and eQTL associations in the same tissue or cell type (Fig. 2a and Supplementary Table 15).

Nineteen of the 119 genes implicated through chromatin interaction mapping are especially notable, as they were implicated via interactions between two independent genome-wide significant genomic risk loci. There were several chromatin interactions in seven tissue types (aorta, hippocampus, left ventricle, right ventricle, liver, spleen, and pancreas) across two risk loci on chromosome 6 (Fig. 3a). Two genes are present in locus 45 and were mapped by chromatin interactions from risk locus 46 (*HFE* and *HIST1H4C*), and 16 genes encode histones in locus 46 and were mapped by interactions from locus 45 (Supplementary Table 15). One gene, *XKR6*, located on chromosome 8 in risk locus 61 is implicated by

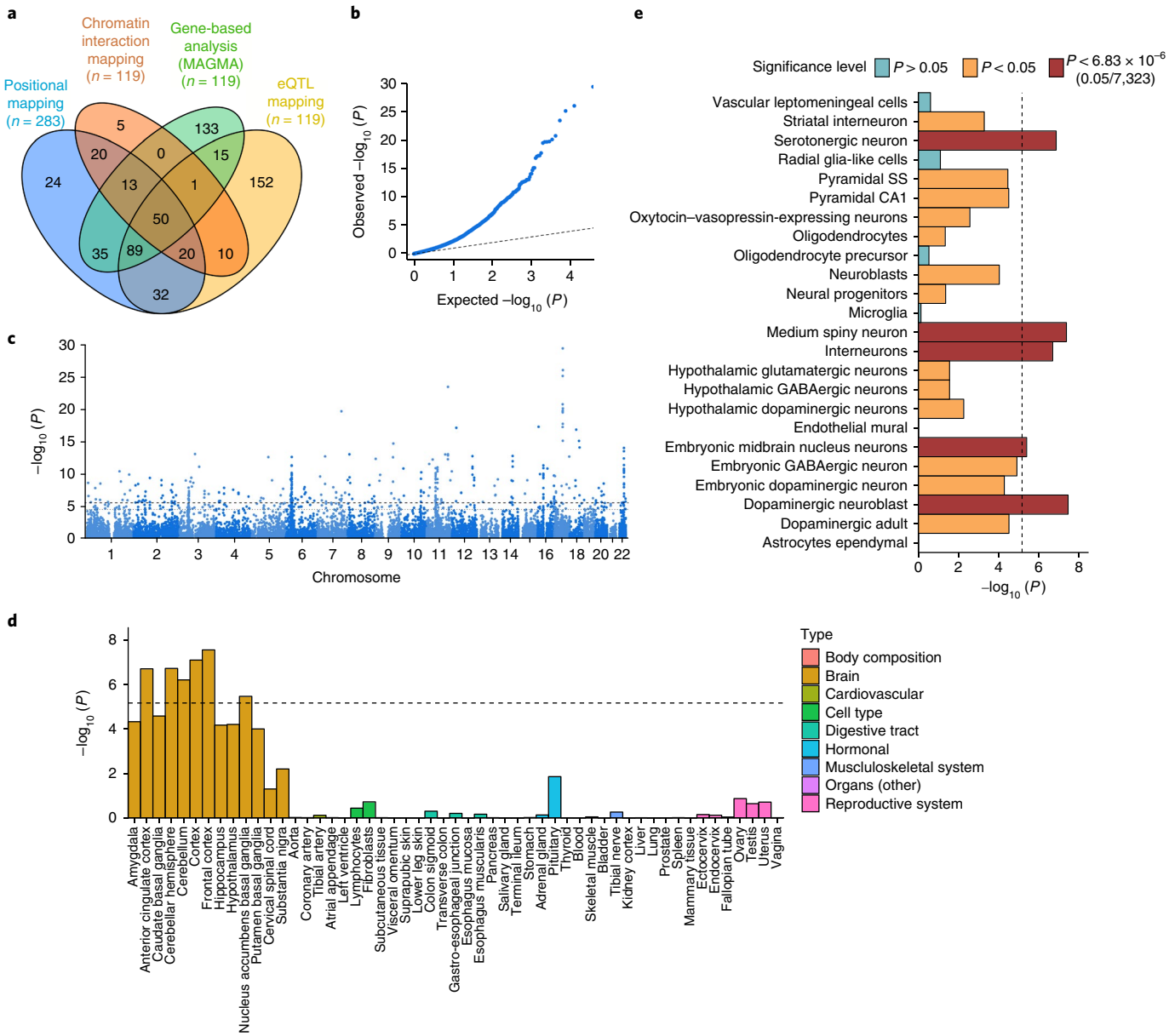

Fig. 2 | Mapping of genes and tissue expression and cell expression profiles. a, Venn diagram showing overlap of genes implicated by positional mapping, eQTL mapping, chromatin interaction mapping, and GWGAS. **b**, Quantile–quantile plot of GWGAS ($n = 449{,}484$ individuals). Gene *P* values were computed using MAGMA's gene-based test. **c**, Manhattan plot of the GWGAS on neuroticism ($n = 449{,}484$ individuals). Gene *P* values were computed using MAGMA's gene-based test. The *y* axis shows the $-\log_{10}$-transformed *P* value of each gene, and the *x* axis shows the chromosomal position (start position). The upper dashed line indicates the threshold for genome-wide significance of the gene-based test ($P < 2.76 \times 10^{-6}$; 0.05/18,128), and the lower dashed line indicates the suggestive threshold ($P < 2.76 \times 10^{-5}$; 0.5/18,128). **d**, Gene expression profiles of identified genes for 53 tissue types. Expression data were extracted from the GTEx database. Expression values (Reads Per Kilobase Million - RPKM) were $\log_2$ transformed with pseudocount 1 after Winsorization at 50 and averaged per tissue. Gene set tests for tissue expression were calculated with MAGMA (Methods). **e**, Enrichment of genetic signal for neuroticism in 24 cell types derived from mouse brain (Methods). The dashed line indicates the Bonferroni-corrected significance threshold ($P = 0.05/7{,}323 = 6.83 \times 10^{-6}$).

chromatin interactions in five tissue types (aorta, left ventricle, liver, pancreas, and spleen), including cross-locus interactions from locus 60 (Fig. 3b and Supplementary Table 15). This gene was also mapped by eQTLs in blood and transformed fibroblasts. Of the 19 genes mapped by two loci, 4 were located outside of the risk loci (*HIST1H2AI*, *HIST1H3H*, *HIST1H2AK*, and *HIST1H4L*) and 7 were also implicated by eQTLs in several tissue types (*HFE* in subcutaneous adipose, aorta, esophagus muscularis, lung, tibial nerve, sun-exposed skin, and thyroid; *HIST1H4J* in blood and adrenal gland; and *HIST1H4K*, *HIST1H2AK*, *HIST1H2BO*, and *XKR6* in blood).

We used the gene-based *P* values for gene set analysis in MAGMA[25] and tested 7,246 predefined gene sets derived from MSigDB[26], gene expression profiles in 53 tissue types obtained from the Genotype-Tissue Expression (GTEx) Project[27], and 24 cell-type-specific expression profiles using RNA-seq information[28] (Methods). Neuroticism was significantly associated with genes predominantly expressed in six brain tissue types (Fig. 2d and Supplementary Tables 17 and 18) and with seven Gene Ontology (GO) gene sets, with the strongest association for neurogenesis ($P = 4.43 \times 10^{-9}$) and neuron differentiation ($P = 3.12 \times 10^{-8}$) (Supplementary Table 17). Conditional

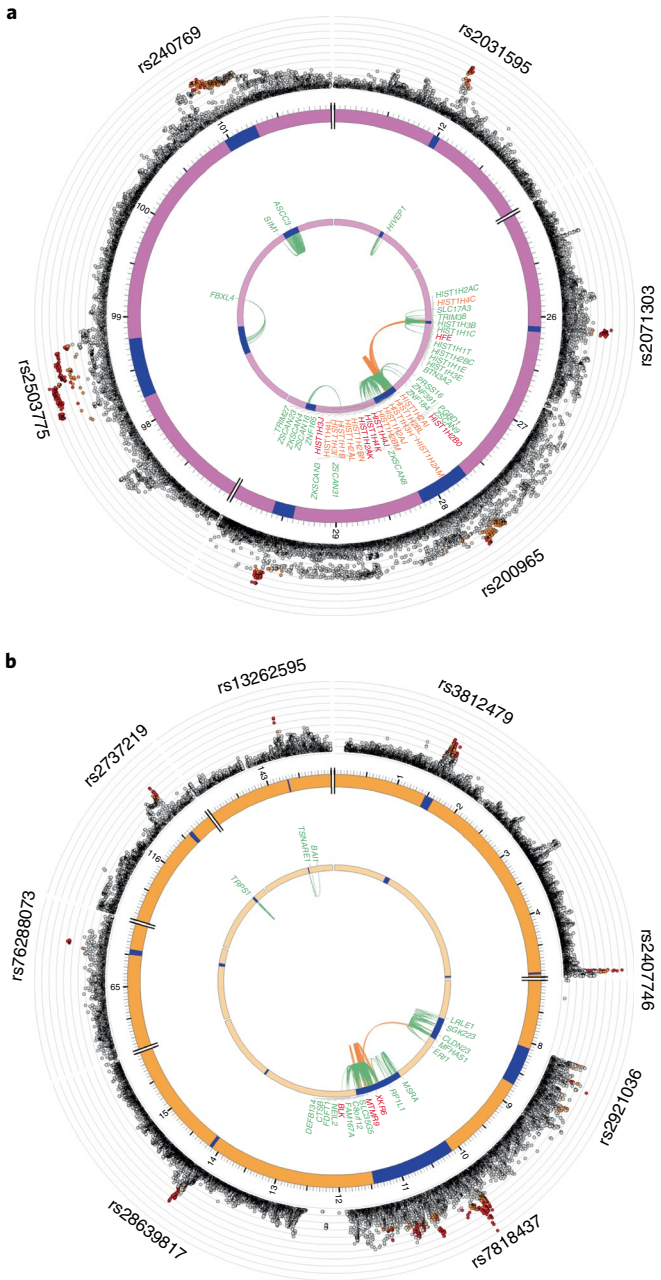

Fig. 3 | Genomic risk loci, eQTL associations, and chromatin interactions for chromosomes 6 and 8, containing cross-locus interactions. Circos plots showing genes on chromosome 6 (**a**) and chromosome 8 (**b**) that were implicated through the genomic risk loci (blue areas) by positional mapping, by chromatin interaction mapping (CTI; orange font), eQTL mapping (green font), or by both chromatin interaction and eQTL mapping (red font). The outer layer shows a Manhattan plot containing the $-\log_{10}$-transformed *P* value of each SNP in the GWAS meta-analysis of neuroticism ($n = 449{,}484$ individuals). Empty regions in the Manhattan plot layer indicate regions where no SNPs with $P < 0.05$ were situated.

gene set analyses (Methods) suggested that three of the seven gene sets (neurogenesis, $P = 4.43 \times 10^{-9}$; behavioral response to cocaine, $P = 1.84 \times 10^{-7}$; axon part, $P = 5.26 \times 10^{-8}$) had largely independent associations, implying a role in neuroticism (Supplementary Table 19). Conditional analyses of the tissue-specific expression ascertained general involvement of (frontal) cortex–expressed genes (Supplementary Fig. 5 and Supplementary Table 20).

Cell-type-specific gene set analysis showed significant association with genes expressed in multiple mouse-derived brain cell types (Fig. 2e, Methods, and Supplementary Table 21), with dopaminergic neuroblasts ($P = 3.49 \times 10^{-8}$), medium spiny neurons ($P = 4.23 \times 10^{-8}$), and serotonergic neurons ($P = 1.37 \times 10^{-7}$) showing the strongest associations. Conditional analysis indicated that these three cell types were also independently associated with neuroticism.

With the aim to further specify neuroticism's neurobiological interpretation, we compared the genetic signal of the full neuroticism trait to that of two genetically distinguishable neuroticism subclusters, 'depressed affect' and 'worry' (Methods), which we had previously established through hierarchical clustering of the genetic correlations between the 12 neuroticism items[13]. As a validation of the depressed affect dimension, we also compared the genetic signal of neuroticism and the two subclusters to that of depression. Genome-wide association analyses of the subclusters were conducted on the UKB data only (dictated by item-level data availability (Methods); depressed affect, $n = 357{,}957$; worry, $n = 348{,}219$). For depression, our meta-analysis comprised data from the UKB[14] ($n = 362{,}696$; Supplementary Fig. 6), 23andMe[15] ($n = 307{,}354$), and the Psychiatric Genetics Consortium (PGC[29]; $n = 18{,}759$) (total $n = 688{,}809$, which is the largest *n* for depression thus far; r_g between samples: 0.61–0.80; Methods and Supplementary Table 22; see the Supplementary Note for details on the depression GWAS results). Genetic correlations of neuroticism with all three phenotypes were considerable (depression, $r_g = 0.79$; depressed affect, $r_g = 0.88$; worry, $r_g = 0.87$; Supplementary Table 23). The positive genetic correlations between neuroticism and depression might in part be due to overlap in item content between the instruments used to gauge these phenotypes, reducing their operational distinctness[13].

The subclusters showed notable differences in genetic signal (for example, exclusive genome-wide significant associations on chromosomes 2 and 19 for depressed affect and on chromosomes 3 and 22 for worry; Supplementary Figs. 7–13 and Supplementary Tables 24–26). Of the 136 genetic loci associated with neuroticism, 32 were also genome-wide significant for depressed affect (7 shared with depression) but not for worry, and 26 were also genome-wide significant for worry (3 shared with depression) but not for depressed affect (Supplementary Fig. 13 and Supplementary Table 27). These results were mirrored by gene-based analyses (Supplementary Fig. 14, Supplementary Tables 28–30, and Supplementary Note), suggesting that part of neuroticism's genetic signal originated specifically in one of the two subclusters, possibly implicating different causal genetic mechanisms. To further verify the biological distinctness of the two clusters, cluster-specific functional annotation was conducted, which demonstrated that, with respect to SNPs that were highly likely to have functional consequences (ExNS), the clusters were (i) distinct and (ii) adding information to the results of neuroticism sum-score analysis (Supplementary Fig. 15, Supplementary Tables 31–34, and Supplementary Note).

To test whether the signal of the gene sets implicated in neuroticism originated from one of the specific subclusters, we conducted conditional analyses, correcting neuroticism for depressed affect and worry scores separately (Supplementary Fig. 16 and Supplementary Table 35). The association with axon part was markedly lower after correction for worry scores (uncorrected, $P = 5.26 \times 10^{-8}$; corrected for depressed affect, $P = 2.42 \times 10^{-6}$; corrected for worry, $P = 0.0013$), suggesting that the involvement of axon part in neuroticism originates predominantly from the worry component.

To examine the genetic correlational pattern of neuroticism and to compare it to the patterns observed for depression, depressed affect, and worry, we used LDSC[18,30] to calculate genetic correlations with 35 traits for which large-scale GWAS summary statistics were available (Methods and Supplementary Table 36). We observed 11

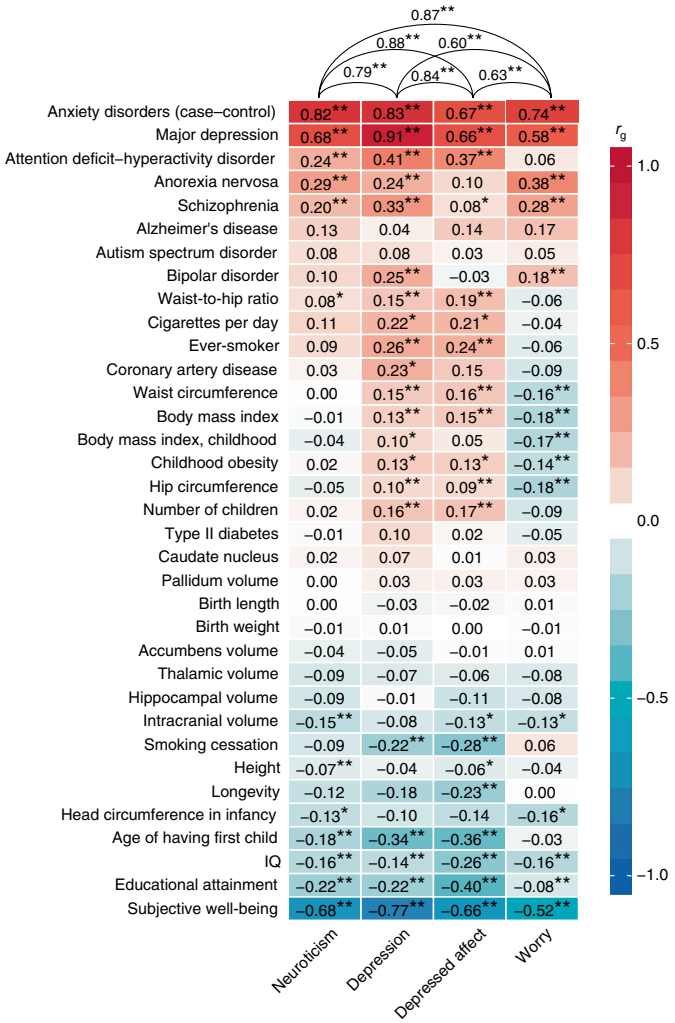

Fig. 4 | Genetic correlations between neuroticism and other traits. Genetic correlations of neuroticism, depression, depressed affect, and worry with various traits and diseases (computed using cross-trait LD Score regression). LD Score regression (Methods) tested genome-wide SNP associations for the neuroticism score against previously published results for 35 neuropsychiatric outcomes, anthropometric and health-related traits, and brain morphology (Supplementary Tables 36 and 37). Genetic correlations among neuroticism, depression, depressed affect, and worry are displayed in the top part of the figure. Red and blue indicate positive and negative genetic correlations, respectively, whereas the hue indicates the strength of the genetic correlations. Sample sizes for the traits in this figure are presented in Supplementary Table 36. *$P < 0.01$; **Bonferroni-corrected P value threshold ($P < 3.6 \times 10^{-4}$).

Bonferroni-corrected significant genetic correlations between neuroticism and other traits ($\alpha = 0.05/(4 \times 35)$; $P < 3.6 \times 10^{-4}$) (Fig. 4 and Supplementary Table 37), which covered previously reported psychiatric traits (r_g range: 0.20 to 0.82) and subjective well-being ($r_g = -0.68$). These correlations were supported by enrichment of genes associated with neuroticism in sets of genes that had previously been implicated in psychiatric traits (Supplementary Table 38). The r_g values of depression and depressed affect strongly mirrored each other (the correlation between their r_g values was $r = 0.98$; Supplementary Note), which validated the depressed affect cluster. The correlational patterns for depressed affect and worry were markedly different (for example, anorexia nervosa, schizophrenia, and ever-smoker) and sometimes in opposite directions (for example, body mass index (BMI)). The genetic correlations of the full neuroticism trait seemed to be a mix of the genetic signal of both clusters, with neuroticism's r_g values generally in between the cluster-specific r_g values.

To investigate whether these genetic correlations reflected directional effects, we performed Mendelian randomization (MR) analysis using the GSMR package[31] (Methods). Among other things, we observed unidirectional effects of BMI on depression and depressed affect ($b_{xy} = 0.061$, $P = 4.96 \times 10^{-12}$ and $b_{xy} = 0.049$, $P = 5.35 \times 10^{-6}$, respectively) and bidirectional associations between neuroticism and depression, as well as between all four main traits and subjective well-being, cognition, and several psychiatric disorders (Supplementary Table 39 and Supplementary Note).

We aimed to identify gene–drug interactions (using the Drug Gene Interaction database (DGIdb)[32,33]; Methods) of genes identified for each of the four traits, and we observed a large number of potential targets for pharmacotherapeutic intervention that were either shared between traits or distinct for each phenotype (Supplementary Fig. 17, Supplementary Tables 40 and 41, and Supplementary Note).

In conclusion, we identified 124 new genetic loci for neuroticism (73 taking into account a simultaneously conducted study by Luciano et al.[34]; Supplementary Table 42 and Supplementary Note). Extensive functional annotations highlighted several genes implicated through multiple routes. We demonstrated the involvement of specific neuronal cell types and three independently associated genetic pathways, and we established the genetic multidimensionality of the neuroticism phenotype and its link with depression. The current study provides new leads and testable functional hypotheses for unraveling the neurobiology of neuroticism, its subtypes, and its genetically associated traits.

URLs. UK Biobank, http://www.ukbiobank.ac.uk/; MAGMA, http://ctg.cncr.nl/software/magma; MSigDB, http://software.broadinstitute.org/gsea/msigdb/collections.jsp; METAL, http://genome.sph.umich.edu/wiki/METAL_Program; LDSC, https://github.com/bulik/ldsc; FUMA, http://fuma.ctglab.nl/; GSMR, http://cnsgenomics.com/software/gsmr/; DGIdb, http://dgidb.org/; Ethical and Independent Review Services, http://www.eandireview.com/; Genetics of Personality Consortium, http://www.tweelingenregister.org/GPC/; Psychiatric Genomics Consortium, http://www.med.unc.edu/pgc/results-and-downloads; GTEx Portal, https://www.gtexportal.org/home/; GWAS summary statistics https://ctg.cncr.nl/software/summary_statistics.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41588-018-0151-7.

Received: 1 September 2017; Accepted: 20 April 2018;
Published online: 25 June 2018

References

1. Kendler, K. S. & Myers, J. The genetic and environmental relationship between major depression and the five-factor model of personality. *Psychol. Med.* **40**, 801–806 (2010).
2. Middeldorp, C. M. et al. in *Biology of Personal and Individual Differences* (ed. Canli, T.) Ch. 12, 251–272 (Guilford Press, New York and London, 2006).
3. Hettema, J. M., Neale, M. C., Myers, J. M., Prescott, C. A. & Kendler, K. S. A population-based twin study of the relationship between neuroticism and internalizing disorders. *Am. J. Psychiatry* **163**, 857–864 (2006).
4. Hayes, J. F., Osborn, D. P. J., Lewis, G., Dalman, C. & Lundin, A. Association of late adolescent personality with risk for subsequent serious mental illness among men in a Swedish nationwide cohort study. *JAMA Psychiatry* **74**, 703–711 (2017).
5. Smeland, O. B. et al. Identification of genetic loci shared between schizophrenia and the Big Five personality traits. *Sci. Rep.* **7**, 2222 (2017).

6. Van Os, J. & Jones, P. B. Neuroticism as a risk factor for schizophrenia. *Psychol. Med* **31**, 1129–1134 (2001).
7. Genetics of Personality Consortium. Meta-analysis of genome-wide association studies for neuroticism, and the polygenic association with major depressive disorder. *JAMA Psychiatry* **72**, 642–650 (2015).
8. Terracciano, A. et al. Genome-wide association scan for five major dimensions of personality. *Mol. Psychiatry* **15**, 647–656 (2010).
9. de Moor, M. H. M. et al. Meta-analysis of genome-wide association studies for personality. *Mol. Psychiatry* **17**, 337–349 (2012).
10. Lo, M. T. et al. Genome-wide analyses for personality traits identify six genomic loci and show correlations with psychiatric disorders. *Nat. Genet.* **49**, 152–156 (2017).
11. Smith, D. J. et al. Genome-wide analysis of over 106,000 individuals identifies 9 neuroticism-associated loci. *Mol. Psychiatry* **21**, 1–9 (2016).
12. Okbay, A. et al. Genetic variants associated with subjective well-being, depressive symptoms and neuroticism identified through genome-wide analyses. *Nat. Genet.* **48**, 624–633 (2016).
13. Nagel, M., Watanabe, K., Stringer, S., Posthuma, D. & van der Sluis, S. Item-level analyses reveal genetic heterogeneity in neuroticism. *Nat. Commun.* **9**, 905 (2018).
14. Bycroft, C. et al. Genome-wide genetic data on ~500,000 UK Biobank participants. Preprint at bioRxiv https://doi.org/10.1101/166298 (2017).
15. Eriksson, N. et al. Web-based, participant-driven studies yield novel genetic associations for common traits. *PLoS Genet.* **6**, e1000993 (2010).
16. Willer, C. J., Li, Y., Abecasis, G. R. & Overall, P. METAL: fast and efficient meta-analysis of genome-wide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
17. Skol, A. D., Scott, L. J., Abecasis, G. R. & Boehnke, M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat. Genet.* **38**, 209–213 (2006).
18. Bulik-Sullivan, B. K. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
19. Yang, J. et al. Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* **19**, 807–812 (2011).
20. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
21. Kircher, M. et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
22. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
23. Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: polygenic risk score software. *Bioinformatics* **31**, 1466–1468 (2015).
24. Vilhjálmsson, B. J. et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
25. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
26. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
27. GTEx Consortium. The genotype–tissue expression (GTEx) pilot analysis: multi-tissue gene regulation in humans. *Science* **348**, 648–660 (2015).
28. Skene, N. G. et al. Genetic identification of brain cell types underlying schizophrenia. *Nat. Genet.* **50**, 825–833 (2018).
29. Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol. Psychiatry* **18**, 497–511 (2013).
30. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
31. Zhu, Z. et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat. Commun.* **9**, 224 (2018).
32. Griffith, M. et al. DGIdb: mining the druggable genome. *Nat. Methods* **10**, 1209–1210 (2013).
33. Cotto, K. C. et al. DGIdb 3.0: a redesign and expansion of the drug–gene interaction database. *Nucleic Acids Res.* **46**, D1068–D1073 (2017).
34. Luciano, M. et al. Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. *Nat. Genet.* **50**, 6–11 (2018).

Acknowledgements

We would like to thank the participants, including the 23andMe customers who consented to participate in research, and the researchers who collected and contributed to the data. This work was funded by the Netherlands Organization for Scientific Research through the following grants: NWO Brain and Cognition 433-09-228 (D.P.), NWO MagW VIDI 452-12-014 (S.v.d.S.), NWO VICI 435-13-005 (D.P.) and 645-000-003) (D.P.). P.R.J. was funded by the Sophia Foundation for Scientific Research (SSWO, grant no. S14-27). J.H.-L. was funded by the Swedish Research Council (Vetenskapsrådet, award 2014-3863), StratNeuro, the Wellcome Trust (108726/Z/15/Z), and the Swedish Brain Foundation (Hjärnfonden). N.G.S. was supported by the Wellcome Trust (108726/Z/15/Z). J.B. was funded by the Swiss National Science Foundation. The work of H.T. was supported by a NWO–VICI grant (NWO-ZonMW 016.VICI.170.200). Analyses were carried out on the Genetic Cluster computer, which is financed by the Netherlands Scientific Organization (NWO award 480-05-003 to D.P.), VU University (Amsterdam, The Netherlands), and the Dutch Brain Foundation and is hosted by the Dutch National Computing and Networking Services, SurfSARA. This research has been conducted using the UK Biobank resource (application 16406).

Author contributions

S.v.d.S. and D.P. conceived the study; M.N. and P.R.J. performed the analyses; S.S. performed the quality control on the UKB data and wrote a pipeline to facilitate data processing; K.W. constructed the tool for biological annotation and ran the analyses; H.T. and T.W. read and commented on the pre-final version of the manuscript; A.R.H., C.A.d.L., J.E.S., and T.J.C.P. wrote part of the analysis pipeline and assisted in interpreting results; N.G.S., A.B.M.-M., S.L., and J.H.-L. provided single-cell RNA-seq data for mouse brain cell types; J.B. and P.F.S. performed the single-cell gene expression analysis; and M.N., P.R.J., S.v.d.S., and D.P. wrote the paper. All authors discussed the results and commented on the paper.

Competing interests

J.H.-L. is a scientific advisor at Cartana and has received a grant from Roche. P.F. has received a grant from Lundbeck and is currently a member of the advisory committee. Over the last 3 years, P.F. has been on the scientific advisory board at Pfizer, received a consultation fee from Element Genomics, and received speaker reimbursement fees from Roche.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0151-7.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to D.P.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

23andMe Research Team

Michelle Agee[13], Babak Alipanahi[13], Adam Auton[13], Robert K. Bell[13], Katarzyna Bryc[13], Sarah L. Elson[13], Pierre Fontanillas[13], Nicholas A. Furlotte[13], David A. Hinds[13], Bethann S. Hromatka[13], Karen E. Huber[13], Aaron Kleinman[13], Nadia K. Litterman[13], Matthew H. McIntyre[13], Joanna L. Mountain[13], Elizabeth S. Noblin[13], Carrie A. M. Northover[13], Steven J. Pitts[13], J. Fah Sathirapongsasuti[13], Olga V. Sazonova[13], Janie F. Shelton[13], Suyash Shringarpure[13], Chao Tian[13], Joyce Y. Tung[13], Vladimir Vacic[13] and Catherine H. Wilson[13]

[13]23andMe, Inc, Mountain View, CA, USA

Methods

Samples. *UK Biobank.* The UKB study is a major data resource that contains genetic, as well as a wide range of phenotypic, data of ~500,000 participants aged 39–73 years at recruitment[14]. We used data released in July 2017, and selection (discussed below) resulted in final sample sizes of $n = 372{,}903$ and $n = 362{,}696$ individuals for neuroticism and depression, respectively (Supplementary Note). The UKB received ethical approval from the National Research Ethics Service Committee North West–Haydock (reference 11/NW/0382), and all study procedures were performed in accordance with the World Medical Association for medical research. The current study was conducted under UKB application 16406.

23andMe. 23andMe, Inc., is a large personal genomics company that provides genotype and health-related information to customers. For the neuroticism and depression meta-analyses, we used neuroticism and depression GWAS summary statistics, respectively, from a subset of 23andMe research participants (neuroticism, $n = 59{,}206$; depression, $n = 307{,}354$), which is described in more detail elsewhere[10,35]. All included participants provided informed consent and were of European ancestry, and related individuals were excluded. Online data collection procedures were approved by the Ethical and Independent Review Services (E&I Review), an AAHRPP-accredited private institutional review board (see URLs).

Genetics of Personality Consortium. The GPC is a large collaboration concerning GWAS on personality. We used summary statistics of neuroticism from the first GPC personality meta-analysis (GPC1; see URLs)[9] on ten discovery cohorts (SardiNIA, NTR/NESDA, ERF, SAGE, HBCS, NAG, IRPG, QIMR, LBC1936, BLSA, and EGPUT), which included in total $n = 17{,}375$ participants of European descent. All included studies were approved by local ethics committees, and informed consent was obtained from all of the participants.

Psychiatric Genetics Consortium. The PGC unites investigators worldwide to conduct genetic meta- and mega-analyses for psychiatric disorders. We used summary statistics from the latest published PGC meta-analysis on depression (see URLs)[29], which included data from eight cohorts (Bonn-Mannheim, GAIN, GenRED, GSK, MDD2000, MPIP, RADIANT, and STAR*D), which covered $n = 18{,}759$ participants of European descent. All included studies were approved by local ethics committees, and informed consent was obtained from all of the participants.

Phenotype assessment: neuroticism. *UK Biobank.* Neuroticism was measured with 12 dichotomous (yes or no) items of the Eysenck Personality Questionnaire Revised Short Form (EPQ-RS)[36], using a touchscreen-based questionnaire at the UKB assessment centers (Supplementary Note). Participants with valid responses to < 10 items were excluded from analyses. A weighted neuroticism sum-score was calculated by adding up individual valid item responses and dividing that sum by the total number of valid responses. In addition, we constructed two scores based on subsets of genetically homogeneous neuroticism items, as established previously[13] through hierarchical clustering analysis of the genetic correlations between the 12 neuroticism items (Supplementary Note). Specifically, the sum of scores on four EPQ-RS items (i.e., "Do you often feel lonely?", "Do you ever feel 'just miserable' for no reason?", "Does your mood often go up and down?", and "Do you often feel 'fed up'?") was used to obtain scores for the cluster depressed affect. Similarly, the sum of scores on four other EPQ-RS items (i.e., "Are you a worried?", "Do you suffer from nerves?", "Would you call yourself a nervous person?", and "Would you call yourself tense or highly strung") was used to obtain scores for the cluster worry. In the item-cluster analyses, only participants with complete scores on all four items were included, which resulted in $n = 357{,}957$ and $n = 348{,}219$ for depressed affect and worry, respectively.

23andMe. Neuroticism was operationalized as the sum of eight neuroticism items (five-point Likert scale: from 'disagree strongly' to 'agree strongly') from the Big Five Inventory (BFI)[37,38], as obtained in an online survey. Only participants with valid responses to all items were included in the analyses (Supplementary Note).

Genetic Personality Consortium. All ten cohorts included in the first meta-analysis of the GPC used sums of the scores on 12 items (five-point Likert scale: from 'strongly disagree' to 'strongly agree') of the NEO-FFI[39] to measure neuroticism. If < 4 item scores were missing, data on invalid items were imputed by taking an individual's average score on valid items. Participants were excluded from analyses if they had invalid scores on > 3 items[9] (Supplementary Note).

Phenotype assessment: depression. *UK Biobank.* Depression was operationalized by adding up the scores on two continuous items ("Over the past two weeks, how often have you felt down, depressed or hopeless?" and "Over the past two weeks, how often have you had little interest or pleasure in doing things?"; both were evaluated on a four-point Likert scale: from 'not at all' to 'nearly every day'), resulting in a continuous depression score (as used previously[12]). Only participants with scores on both items were included in the analyses, which resulted in $n = 362{,}696$ (Supplementary Note).

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

Psychiatric Genetics Consortium. This concerns a case–control sample. Cases had a DSM-IV lifetime (sometimes (early-onset) recurrent) major depressive disorder (MDD) diagnosis, which was established through either structured diagnostic interviews or clinician-administered DSM-IV checklists. Most cases were ascertained from clinical sources, whereas controls were randomly selected from population resources and screened for a lifetime history of MDD[29] (Supplementary Note).

Genotyping and imputation. *UK Biobank: neuroticism.* We used genotype data released by the UKB in July 2017. The genotype data collection and processing are described in detail by the responsible UKB group[14]. In short, 489,212 individuals were genotyped on two customized SNP arrays (the UK BiLEVE Axiom array ($n = 50{,}520$) and the UK Biobank Axiom array ($n = 438{,}692$)), which covered 812,428 unique genetic markers (95% overlap in SNP content). After quality-control procedures[14], 488,377 individuals and 805,426 genotypes remained. Genotypes were phased and imputed by the coordinating team to approximately 96 million genotypes by using a combined reference panel, including the Haplotype Reference Consortium and the UK10K haplotype panel. Imputed and quality-controlled genotype data were available for 487,422 individuals and 92,693,895 genetic variants. As recommended by the UKB team, variants imputed from the UK10K reference panel were removed from the analyses due to technical errors in the imputation process.

In our analyses, only individuals of European descent (based on genetic principal components) were included. Therefore, principal components from the 1000 Genomes reference populations[40] were projected onto the called genotypes available in UKB. Subjects were identified as European if their projected principal-component score was closest (based on Mahalanobis distance) to the average score of the European 1000 Genomes sample[41]. European subjects with a Mahalanobis distance > 6 s.d. were excluded. In addition, participants were excluded based on withdrawn consent, UKB-provided relatedness (subjects with the most inferred relatives, third degree or closer, were removed until no related subjects were present), discordant sex, and sex aneuploidy. After selecting individuals based on available neuroticism sum-score and active consent for participation, 372,903 individuals remained for the analyses.

To correct for population stratification, 30 principal components were calculated on the subset of quality-controlled unrelated European subjects based on 145,432 independent ($r^2 < 0.1$) SNPs with MAF > 0.01 and INFO = 1 using FlashPC2[42]. Subsequently, imputed variants were converted to a hard call by using a certainty threshold of 0.9. Multiallelic SNPs, indels, and SNPs without unique rs identifiers were excluded, as well as SNPs with a low imputation scores (INFO score < 0.9), low MAF (< 0.0001), and high missingness (> 0.05). This resulted in a total of 10,847,151 SNPs that were used for downstream analysis.

UK Biobank: depression. A genotyping, imputation, and filtering procedure similar to the one described above for the UKB neuroticism GWAS was used for the UKB depression GWAS, which resulted in $n = 362{,}696$.

Genome-wide association analyses. *UK Biobank: neuroticism.* Genome-wide association analyses were performed in PLINK[43,44], using a linear regression model of additive allelic effects with age, sex, Townsend deprivation index, genotype array, and ten genetic European-based principal components as covariates (Supplementary Note).

UK Biobank: depression, depressed affect, and worry. The settings, covariates, and exclusion criteria for the UKB depression, UKB depressed affect, and UKB worry GWAS were the same as those described above for the UKB neuroticism GWAS, with 10,847,151 SNPs remaining after all exclusion steps (Supplementary Note).

Other samples. Summary statistics were used for 23andMe, GPC, and PGC. Details on the genome-wide association analyses of these samples can be found elsewhere (23andMe neuroticism[10]; 23andMe depression[35]; GPC neuroticism[9]; PGC depression[29]).

Meta-analysis. To maximize the statistical power to detect associated genetic variants of small effect, we conducted meta-analyses for both neuroticism and depression[17] (Supplementary Note). All meta-analyses were performed in METAL[16].

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

Depression. Because the UKB GWAS concerned a continuous operationalization of the depression phenotype, whereas 23andMe and PGC used case–control phenotypes, the odds ratios from the 23andMe and PGC summary statistics were converted to log odds, which reflected the direction of the effect. The meta-analysis was then performed on the *P* value of each SNP by using a two-sided sample-size-weighted fixed-effects analysis. Bonferroni correction was applied to correct for multiple testing. Genetic correlations between the three samples were moderate to strong (r_g range: 0.61 to 0.80; Supplementary Table 22).

Genomic risk loci and functional annotation. Functional annotation was performed with FUMA[17] (see URLs), an online platform for functional mapping of genetic variants. We first defined independent significant SNPs, which had a genome-wide significant *P* value (5×10^{-8}) and were independent at $r^2<0.6$. A subset of these independent significant SNPs, which were independent from each other at $r^2<0.1$, was marked as lead SNPs (based on LD information from UKB genotypes; see the Supplementary Note for a more detailed explanation). Subsequently, genomic risk loci were defined by merging lead SNPs that physically overlapped or for which LD blocks were less than 250 kb apart. Note that when analyzing multiple phenotypes, as in the current study, a locus may be discovered for different phenotypes while different lead SNPs are identified.

All SNPs in the meta-analysis results that were in LD ($r^2>0.6$) with one of the independent significant SNPs and that had $P<1.0\times10^{-5}$ and MAF > 0.0001 were selected for annotation. The rationale behind this inclusive approach was that the most significant SNP in the locus was not necessarily the causal SNP but that it might be in LD with the causal SNP. We thus annotated all SNPs in LD with the most significant SNP to get insight into the possible biological reasons for observing a statistical association. We note that liberalizing the r^2 and *P*-value thresholds can dilute the functional annotation results, whereas more stringent thresholds may result in exclusion of possibly interesting functional variants. Functional consequences for these SNPs were obtained by performing ANNOVAR[45] gene-based annotation using Ensembl genes. In addition, CADD scores (indicating the deleteriousness of a SNP, with scores > 12.37 seen as likely deleterious[21]) and RegulomeDB scores[46] (for which a higher probability of having a regulatory function is indicated by a lower score) were annotated to SNPs by matching chromosome, position, reference, and alternative alleles. CADD scores integrate a number of diverse annotations into a single measure that correlates with pathogenicity, disease severity, and experimentally measured regulatory effects and complex trait associations[21].

Gene mapping. SNPs in genomic risk loci that were genome-wide significant or were in LD ($r^2>0.6$) with one of the independent significant SNPs were mapped to genes in FUMA[20] using one of three strategies.

First, positional mapping uses the physical distances (i.e., within 10-kb windows) from known protein-coding genes in the human reference assembly (GRCh37 or hg19) to map SNPs to genes. The second strategy, eQTL mapping, uses information from three data repositories (GTEx, Blood eQTL browser, and BIOS QTL browser) and maps SNPs to genes based on a significant eQTL association (i.e., where the expression of the gene is associated with allelic variation at the SNP). eQTL mapping is based on cis-eQTLs, which can map SNPs to genes up to 1 Mb away. FUMA applied a false discovery rate (FDR) of 0.05 to define significant eQTL associations. Third, chromatin interaction mapping mapped SNPs to genes based on a significant chromatin interaction between a genomic region in a risk locus and promoter regions of genes (250 bp upstream and 500 bp downstream of a TSS). This type of mapping does not have a distance boundary (as in eQTL mapping) and may therefore involve long-range interactions. Currently, FUMA contains Hi-C data for 14 tissue types from the study of Schmitt et al.[47]. Notably, as chromatin interactions are usually defined in a certain resolution (in the current study, 40 kb), an interacting region may span several genes. Hence, this method would map all SNPs within these regions to genes in the corresponding interaction region. By integrating predicted enhancers and promoters in 111 tissue and cell types from the Roadmap Epigenomics Project[48], we aimed to prioritize candidate genes from chromatin interaction mapping. Using this information, FUMA selected chromatin interactions for which one region involved in the interaction overlapped with predicted enhancers and the other overlapped with predicted promoters 250 bp upstream and 500 bp downstream of the TSS of a gene. Similar to eQTL mapping, we used an FDR of 1×10^{-5} to define significant interactions.

Gene-based analysis. GWGAS can identify genes in which multiple SNPs show moderate association to the phenotype of interest without reaching the stringent genome-wide significance level. At the same time, because GWGAS takes all SNPs within a gene into account, a gene harboring a genome-wide significant SNP may not be implicated by GWGAS analyses when multiple other SNPs within that gene show only very weak association signal. The *P* values from the SNP-based GWAS meta-analyses for neuroticism and depression, and the GWAS for depressed affect and worry, were used as input for the GWGAS in MAGMA (see URLs)[25], and all 19,427 protein-coding genes from the NCBI 37.3 gene definitions were used. We annotated all of the SNPs in our genome-wide association (meta-)analyses to these genes, resulting in 18,187, 18,187, 18,182, and 18,182 genes that were represented by at least one SNP in the neuroticism meta-analysis, the depression meta-analysis, the depressed affect GWAS, and the worry GWAS, respectively. We included a window around each gene of 2 kb before the TSS and 1 kb after the transcription stop site. Gene association tests were performed, taking into account the LD between SNPs, and a stringent Bonferroni correction was applied to correct for multiple testing (0.05/number of genes tested: $P<2.75\times10^{-6}$).

Gene set analysis. We used MAGMA[25] to test for association of predefined gene sets with neuroticism, depression, depressed affect, and worry. A total of 7,246 gene sets were derived from several resources, including BioCarta, KEGG, Reactome[49], and GO. All gene sets were obtained from MSigDB version 6.0 (see URLs). Additionally, we performed gene set analysis on 53 tissue expression profiles obtained from the GTEx portal (see URLs) and on 24 cell-type-specific expression profiles.

For all gene sets, we computed competitive *P* values, which result from testing whether the combined effect of genes in a gene set is significantly larger than the combined effect of the same number of randomly selected genes (in contrast to testing against the null hypothesis of no effect; self-contained test). Here we only report Bonferroni-corrected ($\alpha=0.05/7{,}323=6.83\times10^{-6}$) competitive *P* values, which were more conservative than the self-contained *P* values.

Cell-type-specific expression analysis. Definition and calculation of gene sets for cell-type-specific expression is described in detail elsewhere[28,50]. Briefly, brain-cell-type expression data were drawn from single-cell RNA-seq (scRNA-seq) data from mouse brain[28]. For each gene, the value for each cell type was calculated by dividing the mean unique molecular identifier (UMI) counts for the given cell type by the summed mean UMI counts across all cell types[28]. MAGMA[25] was used to calculate associations between gene-wise *P* values from the meta-analysis and cell-type-specific gene expression. Genes were grouped into 40 equally sized bins by specificity of expression, and bin membership was subsequently regressed on gene-wise association with neuroticism in the meta-analysis. Results were deemed significant if the association *P* values were smaller than the relevant Bonferroni threshold.

Conditional gene set and tissue expression analyses. Conditional gene set analyses were performed using MAGMA[25] to determine which tissue expression levels and MSigDB gene sets represented independent associations. In these regression-based analyses, the effect of a gene set (or tissue expression) of interest was conditioned on the effects of another gene set (or tissue expression) to correct the association of the tested gene set for any effect it shared with the conditioned-on gene set.

For the MSigDB gene sets, we conducted two series of conditional analyses. First, we performed forward selection on the initially significant gene sets, in each step selecting the most strongly associated gene set after conditioning on all already-selected gene sets (Supplementary Table 19). Second, to test whether the association of gene sets to neuroticism was primarily driven by the association signal of one specific subcluster, we also re-ran the GO gene set analyses, conditioning on the gene *z* scores of depressed affect or worry (Supplementary Table 35). If the gene set association decreased after conditioning on one cluster but did not decrease or did so to a lesser extent when conditioned on the other, then this suggested that neuroticism's association to that gene set was primarily driven by the genetic effects of the first, and not the second, item cluster.

Genetic correlations. Genetic correlation (r_g) values were computed using LD Score regression[18,30] (see URLs). The significance of the genetic correlations of neuroticism, depression, depressed affect, and worry with 35 behavioral, social, and (mental) health phenotypes for which summary statistics were available was determined by correcting for multiple testing through a stringent Bonferroni-corrected threshold of $P<0.05/(4\times35)=3.6\times10^{-4}$.

Mendelian randomization. We performed MR analysis to test whether genetic correlations could be explained by directional effects between traits. Generalized summary-data-based MR (GSMR)[31], a summary-statistics-based MR method that uses independent genome-wide significant variants as instrumental variables, was used for MR analysis. Causal associations were tested between the 4 traits and the 21 traits that showed significant genetic correlations (r_g values) in LD Score regression analysis with at least one of the 4 traits. To test for unidirectional and bidirectional effects, we performed both forward and reverse GSMR analyses (i.e., using the four GWAS traits either as predictor or as outcome). Associations were Bonferroni corrected for multiple testing with $P<0.05/(21\times4\times2)=2.98\times10^{-4}$.

Partitioned heritability. To investigate the relative contribution to the overall SNP-based heritability annotated to 22 specific genomic categories, we partitioned SNP heritability by binary annotations using stratified LD Score regression[22]. Information about binary SNP annotations was obtained from the LD Score website (see URLs). Enrichment results reflected the X-fold increase in h^2 proportional to the number of SNPs (for example, enrichment = 13.79 for SNPs in conserved regions implies that a 13.79-fold increase in h^2 is carried by SNPs in these region, corrected for the proportion of SNPs in these regions compared to all tested SNPs).

Gene drug targets. We aimed to identify potential druggable targets by searching for the implicated genes (by one of the gene-mapping strategies) in DGIdb[32,33] (version 3.0; see URLs). The DGIdb contains mined data from several resources and provides a comprehensive overview of the druggability of gene targets. First, we searched 20 drug–gene databases for interactions with existing medicines based on 48 known interaction types with genes that were implicated in each of the four phenotypes. Filtering was performed based on known interaction types and interactions with US Food and Drug Administration (FDA)-approved pharmaceutical compounds. Second, to identify genes that may form targets for novel therapies in addition to existing medicines, we searched for the potential gene druggability of gene targets and performed an additional search in ten DGIdb databases containing information about gene targetability.

Polygenic risk scoring. To test the predictive accuracy (ΔR^2) of our meta-analysis results for neuroticism, we calculated a PGS based on the SNP effect sizes of the current analysis. For independent samples, we used three hold-out samples; we removed 3,000 individuals from the discovery sample (UKB only, as we had access only to individual-level data from this sample) and re-ran the genome-wide analyses. We repeated this three times, to create three randomly drawn, independent hold-out samples. Next, we calculated a PGS on the individuals in each of the three hold-out samples. PGSs were calculated using LDpred[24] and PRSice[23] (clumping followed by *P*-value thresholding).

For LDpred, PGSs were calculated based on different LDpred priors (P_{LDpred} = 0.01, 0.05, 0.1, 0.5, 1, and infinitesimal). The explained variance (R^2) was derived from the linear model, using the neuroticism summary score as the outcome, while correcting for age, sex, array, Townsend deprivation index, and genetic principal components.

Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability. Our policy is to make genome-wide summary statistics (sumstats) publicly available. Sumstats from our neuroticism meta-analysis, our depression meta-analysis, and the genome-wide association analyses for depressed affect and worry are available for download at the website of the Department of Complex Trait Genetics, CNCR (see URLs).

Note that our freely available meta-analysis sumstats concern results excluding the 23andMe sample. This is a non-negotiable clause in the 23andMe data transfer agreement, which is intended to protect the privacy of the 23andMe research participants. To fully recreate our meta-analysis results for neuroticism (i) obtain the sumstats from Lo et al. (2017) for 23andMe (see below) and (ii) conduct a meta-analysis of our sumstats with those from Lo et al. To fully recreate our meta-analysis results for depression (i) obtain the sumstats from Hyde et al.[35] for 23andMe (see below) and (ii) conduct a meta-analysis of our sumstats with those for Hyde et al.[35].

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

References

35. Hyde, C. L. et al. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat. Genet.* **48**, 1031–1036 (2016).
36. Eysenck, B. G., Eysenck, H. J. & Barrett, P. A revised version of the psychoticism scale. *Pers. Individ. Dif.* **6**, 21–29 (1985).
37. John, O. P. & Srivastava, S. The Big Five trait taxonomy: history, measurement and theoretical perspectives. *Handb. Personal. Theory Res.* **2**, 102–138 (1999).
38. Soto, C. J. & John, O. P. Ten facet scales for the Big Five Inventory: convergence with NEO PI-R facets, self-peer agreement and discriminant validity. *J. Res. Pers.* **43**, 84–90 (2009).
39. Costa, P. & McCrae, R. R. *Professional Manual: Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor-Inventory (NEO-FFI).* (Psychological Assessment Resources, Odessa, FL, USA, 1992).
40. Auton, A. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
41. Webb, B. T. et al. Molecular genetic influences on normative and problematic alcohol use in a population-based sample of college students. *Front. Genet.* **8**, 30 (2017).
42. Abraham, G. & Inouye, M. Fast principal component analysis of large-scale genome-wide data. *PLoS One* **9**, e93766 (2014).
43. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
44. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
45. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
46. Boyle, A. P. et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* **22**, 1790–1797 (2012).
47. Schmitt, A. D. et al. A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep.* **17**, 2042–2059 (2016).
48. Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
49. Croft, D. et al. The Reactome pathway knowledgebase. *Nucleic Acids Res.* **42**, D472–D477 (2014).
50. Coleman, J. R. I. et al. Functional consequences of genetic loci associated with intelligence in a meta-analysis of 87,740 individuals. *Mol. Psychiatry* https://doi.org/10.1038/s41380-018-0040-6 (2018).

nature research

Corresponding author(s): Danielle Posthuma

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

We made use of data collected by external sources (UK biobank, 23andMe, Genetics of Personality Consortium and the Psychiatric genetics Consortium). For all samples the sample size consists of all individuals that remain after quality control of the data and exclusion of withdrawn subjects. Detailed information on the samples used, as well as the exclusion/inclusion criteria, are provided in the Online Methods (sections: Samples, Genotyping and imputation, Genome-wide association analyses).

2. Data exclusions

Describe any data exclusions.

See Online Methods.
For UKB data: we excluded participants from further analyses if they had excessive missing phenotypic data (section: Phenotype assessment), did not pass standard quality control or withdrew their consent to participate in the UK biobank study (section: Genotyping and imputation).

3. Replication

Describe whether the experimental findings were reliably reproduced.

We used a meta-analytic approach, which inherently evaluates the combined evidence for significant association across samples. In addition, we used polygenic risk score profiling to determine whether our current results were predictive of the same outcome in three independent samples.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

NA

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

NA

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

Standard statistical genetics software packages were used for the analyses described in the current manuscript (all are described in more detail in the Online Methods). Below we list the software used:
Plink - Open-source software (Purcell et al., 2007; Chang et al., 2015) used to conduct genome-wide association analyses.
MAGMA - In-house developed software (de Leeuw et al., 2015) for performing gene-based analyses.
FUMA - In-house developed online platform for functional annotation of GWAS results (Watanabe et al., 2017)
LD Score Regression - Used to compute SNP-based heritability and genetic correlations (Bulik-Sullivan et al., 2015)
PRSice - Polygenic Risk Score analysis (Euesden et al., 2015)
LDpred - Polygenic Risk Score analysis (Vilhjálmsson et al., 2015)
GSMR - Generalized Summary-based Mendelian Randomization (Zhu, Z. et al. 2018)

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	Summary statistics from our neuroticism meta-analysis, our depression meta-analysis, and the GWA analyses for depressed affect and worry will be made available for download at https://ctg.cncr.nl/. Note that our freely available meta-analytic sumstats concern results excluding the 23andMe sample. This is a non-negotiable clause in the 23andMe data transfer agreement, intended to protect the privacy of the 23andMe research participants. 23andMe participant data are shared according to community standards that have been developed to protect against breaches of privacy. Currently, these standards allow for the sharing of summary statistics for at most 10,000 SNPs. The full set of summary statistics can be made available to qualified investigators who enter into an agreement with 23andMe that protects participant confidentiality. Interested investigators should contact David Hinds (dhinds@23andme.com) for more information. Neuroticism summary statistics from the Genetics of Personality Consortium (GPC1) are freely available for download at: http://www.tweelingenregister.org/GPC/ Depression summary statistics from the Psychiatric Genetics Consortium (PGC) are freely available for download at: https://www.nimhgenetics.org/available_data/data_biosamples/pgc_public.php For further information, see Online Methods, section Data Availability.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	No antibodies were used.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	No eukaryotic cell lines were used.
b. Describe the method of cell line authentication used.	No eukaryotic cell lines were used.
c. Report whether the cell lines were tested for mycoplasma contamination.	No eukaryotic cell lines were used.
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	No commonly misidentified cell lines were used.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	No animals were used in this study.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	We utilized data collected previously by external sources. All individuals included in the study provided informed consent, and all original studies were approved by the concerned ethical committee (see Online Methods; section 'Samples')