Title

Personality gene discovery and correlation with psychiatric disorders

Authors

Min-Tzu Lo[1], David A. Hinds[2], Joyce Y. Tung[2], Carol Franz[3], Chun-Chieh Fan[1,4], Yunpeng Wang[5,6], Olav B. Smeland[6], Andrew Schork[1,4], Dominic Holland[5], Karolina Kauppi[1,7], Nilotpal Sanyal[1], Valentina Escott-Price[8], Daniel J. Smith[9], Michael O'Donovan[8], Hreinn Stefansson[10], Gyda Bjornsdottir[10], Thorgeir E. Thorgeirsson[10], Kari Stefansson[10], Linda K. McEvoy[1], Anders M. Dale[1,3,5], Ole A. Andreassen[6], Chi-Hua Chen[1]*

Affiliations

[1]Department of Radiology, University of California, San Diego, La Jolla, California, USA; [2]23andMe, Inc., Mountain View, California, USA; [3]Department of Psychiatry, University of California, San Diego, La Jolla, California, USA; [4]Department of Cognitive Science, University of California, San Diego, La Jolla, California, USA; [5]Department of Neurosciences, University of California, San Diego, La Jolla, California, USA; [6]NORMENT, KG Jebsen Centre for Psychosis Research, Institute of Clinical Medicine, University of Oslo and Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway; [7]Department of Radiation Sciences, Umea University, Sweden; [8]MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, UK; [9]Institute of Health and Wellbeing, University of Glasgow, Glasgow, UK; [10]deCODE Genetics/Amgen, Reykjavik, Iceland

24 Correspondence should be addressed to C.-H.C. (chc101@ucsd.edu)

25

Summary (146 words):

Personality is influenced by genetic and environmental factors[1], and associated with mental health. However, the underlying genetic determinants are largely unknown. We identified six genetic loci, including five novel loci[2,3], significantly associated with personality traits in a meta-analysis of genome-wide association studies (N=123,132-260,861). Of these genome-wide significant loci, extraversion was associated with variants in *WSCD2* and near *PCDH15*, and neuroticism with variants on chromosome 8p23.1 and in *L3MBTL2*. We performed a principal component analysis to extract major dimensions underlying genetic variations among five personality traits and six psychiatric disorders (N=5,422-18,759). The first genetic dimension separated personality traits and psychiatric disorders, except that neuroticism and openness to experience were clustered with the disorders. High genetic correlations were found between extraversion and attention-deficit/hyperactivity disorder (ADHD), and between openness and schizophrenia/bipolar disorder. The second genetic dimension was closely aligned with extraversion-introversion and grouped neuroticism with internalizing psychopathology (e.g., depression/anxiety).

Main Text (2012 words)

The Five Factor Model (FFM) of personality, also known as "the Big Five" is commonly used to measure individual differences in personality. It models personality according to five broad domains[4]. Extraversion (versus introversion) reflects talkativeness, assertiveness and activity level. Neuroticism (versus emotional stability) denotes negative affect like anxiety and depression. Agreeableness (versus antagonism) measures cooperativeness and compassion. Conscientiousness (versus undependability) depicts order and discipline. Openness to experience (versus closedness) captures intellectual curiosity and creativity[4,5]. Personality phenotypes, measured by various questionnaires, are represented by continuous quantitative scores on each of the five traits[4].

A meta-analysis of twin and family studies found that approximately 40% of the variance in personality could be attributed to genetic factors[1]. Genome-wide association studies (GWAS) have discovered several variants associated with FFM traits[6-8]. Neuroticism was reported to be associated with an intronic variant in *MAGI1* (P=9.26×10^{-9}, N=63,661)[7], conscientiousness with an intronic variant in *KATNAL2* (P=4.9×10^{-8}, N=17,375)[6], and openness with variants near *RASA1* (P=2.8×10^{-8}, N=17,375)[6] and near *PTPRD* (P=1.67×10^{-8}, N=1,089)[8]. Recent UK Biobank studies (N=106,716-170,908) yielded several single nucleotide polymorphisms (SNPs) associated with neuroticism[2,3].

Another large study, 23andMe, contains well-phenotyped data on personality and offers opportunity to identify additional genetic variants, since all five personality traits were measured in all individuals using the same personality inventory (Online Methods).

We performed a meta-analysis based on GWAS summary statistics to identify genetic variants associated with FFM traits. We included participants with European ancestry from 23andMe (N=59,225) and two samples from the Genetics of Personality Consortium (GPC)[6,7]. GPC-1 (N=17,375)[6] contains data on agreeableness, conscientiousness, and openness, whereas GPC-2 (N=63,661)[7] contains information on extraversion and neuroticism.

Summary statistics of GWAS from 23andMe (available in Supplementary Data Sets 1-5 for the top 10K SNPs) were combined with the two GPC samples separately, yielding a total of 76,600 and 122,886 subjects as the discovery/stage 1 sample. Eight linkage-disequilibrium (LD)-independent SNPs (LD r^2<0.05) were discovered exceeding GWAS significance ($P<5\times10^{-8}$) in the combined meta-analysis (Table 1 and Fig. 1).

To evaluate the consistency of association signals between 23andMe and GPC samples, we conducted genome-wide polygenic analyses using LD Score regression to examine genetic correlations (r_g)[9] of personality traits between the two samples. The estimated r_g were highly significant (r_g=0.86-0.96), suggesting that genetic effects are consistent and replicable between the samples at a polygenic level (Supplementary Fig. 1), and that a considerable number of SNPs below the GWAS significance threshold contain trait-associated genetic effects.

To assess replicability of the eight significant SNPs identified in the discovery/stage1 sample, we obtained their summary statistics from three independent samples, including an independent 23andMe replication sample, UK Biobank cohort (only neuroticism) and an Icelandic sample from deCODE Genetics (Online Methods and Table 1). In the final combined meta-analysis, six SNPs remained GWAS significant. The

other two fell just below GWAS significance but had consistent direction of effects in all samples, suggesting that these may be significant in larger samples. Overall, the directions of effects were consistent for all eight SNPs between the discovery and replication tests, except two SNPs in the smaller (N=7,137) deCODE sample.

The strongest associations were detected for neuroticism within a subregion of 8p23.1, which spans ~4 Mb (Chr8: 8,091,701-11,835,712) with highly correlated SNPs in one big LD block (Fig. 2a). The 8p23.1 region comprises genes related to innate immunity and the nervous system, and is considered as a potential hub for cancer and developmental neuropsychiatric disorders[10]. Our conditional analysis indicated the existence of multiple associations (conditional $P \sim 10^{-7}$) independent of the top SNP within the 8p23.1 locus but these were not GWAS significant.

The UK Biobank studies also identified multiple associations with neuroticism in 8p23.1[2,3], which were attributed to an inversion polymorphism[2]. Our association signals reside in the same inversion region, with an LD of r^2=0.35 (LDlink) between the lead SNP found here and in the UK Biobank study[3]. Additionally, we identified an intronic variant of *MTMR9* within 8p23.1 that was associated with extraversion, with opposite direction of association with neuroticism (Fig.2b). Together, these findings provide converging evidence for the association of 8p23.1 with personality.

For extraversion, we found a significant locus on 12q23.3 within *WSCD2*. This locus has been implicated in a GWAS of temperament in bipolar disorder[11], and linkage analysis[12], suggesting that 12q likely harbors important alleles for temperament and personality. Another SNP significantly associated with extraversion is near *PCDH15*, a member of the cadherin superfamily important for calcium-dependent cell-cell adhesion.

All six SNPs discovered here reside in loci for which genome-wide significant associations with other phenotypes have been reported (NHGRI GWAS catalog). For example, we found a variant associated with neuroticism in *L3MBTL2*, a gene reported to be associated with schizophrenia[13]. Etiologically, neuroticism has been associated with schizophrenia risk[14]. Further, one gene in which we found a variant associated with extraversion, *MTMR9* has been related to response to antipsychotic medications[15]. The SNP associated with conscientiousness in the discovery sample, though not significant in the final meta-analysis, was located in a locus linked to educational attainment[16], and high conscientiousness was found to correlate positively with academic performance[17].

These six SNPs have been found to be significantly associated with gene expression and all are listed as expression quantitative trait loci (eQTL) for brain tissues, to varying degrees (Supplementary Table 1). We performed a Bayesian test[18] to examine whether GWAS signals co-localize with eQTL. The COLOC-estimated posterior probabilities[18] (see Online Methods) indicated that one SNP-associated locus (rs57590327) and its corresponding eQTL (Supplementary Table 1) were probably attributable to a common causal variant (posterior probability=0.76). Another SNP (rs216273) showed evidence of independence with eQTL (posterior probability =0.75). For the rest of the SNPs, the posterior probability ranged between 0 and 0.45, failing to support any of the specified hypotheses. Our analyses did not show consistent evidence for these SNPs influencing personality traits through gene expression levels in the brain, but caution is warranted owing to the small eQTL sample (N=134).

Beyond identifying single genetic variants that each account for very little phenotypic variance, we estimated SNP-based heritability of the traits. All heritability

estimates were significant in our 23andMe discovery sample, with the largest estimate for extraversion (0.18) (Supplementary Table 2). These findings extend those from a previous heritability analysis of FFM traits (N=5,011), in which SNP-based heritability estimates were significant for neuroticism and openness[19]. As expected, SNP-based heritability estimates were lower than those reported in family studies[1].

Relationships among personality traits are also of interest. Although the FFM traits were derived through factor analysis and thus orthogonal in the original findings, most studies observe some degree of phenotypic correlation between traits[19]. Using 23andMe data, we found that neuroticism was inversely related to the other personality traits, whereas agreeableness, conscientiousness, extroversion, and openness were positively correlated. Almost all phenotypic correlations were highly significant, except for openness vs. conscientiousness (Supplementary Table 3). Genetic correlation patterns were congruent with phenotypic correlations but the association was more apparent in genetic structure, reflecting clear shared genetic factors contributing to the correlations (Fig. 3a).

A notable feature of personality is its link with a wide range of social, mental and physical health outcomes[5]. High levels of neuroticism, extraversion and openness have been associated with bipolar disorder[20], and high neuroticism with major depression and anxiety[21]. Low agreeableness has been associated with narcissism, Machiavellianism and psychopathy[22]. In addition to phenotypic relationships, twin and GWAS studies have demonstrated genetic correlations between personality traits and psychiatric disorders[3,21,23], though most focus on only neuroticism (Supplementary Note for details).

We thus sought to quantify the genetic correlations between the five personality

traits and six psychiatric disorders from the Psychiatric Genomics Consortium: schizophrenia (N=17,115), bipolar disorder (N=16,731), major depressive disorder (N=18,759), ADHD (N=5,422) and autism spectrum disorder (N=10,263), and from Genetic Consortium for Anorexia Nervosa (N=17,767) (see Online Methods and Supplementary Table 2). A pair-wise genetic correlation matrix (11×11) was constructed, which revealed several significant correlations (Fig. 3a, Supplementary Table 4). For example, neuroticism was highly correlated with depression, and extraversion with ADHD. To complement genetic correlation estimation via LD Score regression[9], we compared the pattern of GWAS results by assessing whether signs of genetic effects were concordant between the top associations among these traits and disorders. The results of the sign tests of directional effects closely matched the genetic correlations (Supplementary Fig. 2).

Given the moderate and high genetic correlations, we subsequently conducted a principal component analysis (PCA) to extract principal components of genetic variation (Fig. 3b). We projected all phenotypes onto a two dimensional space spanned by the top two principal components (PC1 and PC2) of genetic variation. This loading plot summarizes the genetic relationships between personality traits and psychiatric disorders. The analysis integrates genomic information with traditionally defined phenotypes to better understand basic dimensions of the full range of human behavior, from typical to pathological, in line with the research strategy of the Research Domain Criteria (RDoC)[24].

Our results indicate that openness, bipolar disorder, and schizophrenia cluster in the first quadrant (Fig. 3b). Interestingly, all three share phenotypic commonality in that they

have been linked to heightened creativity and dopamine activity[25,26]. Most personality traits (conscientiousness, agreeableness and extraversion) cluster in the second quadrant. Neuroticism and depression are in the fourth quadrant. Autism and anorexia nervosa are captured by factors in higher dimensions and have relatively low loadings on the first two components, as indicated by short arrows on these two dimensions. Notably, ADHD has a high genetic correlation with extraversion and low correlations with other psychiatric disorders (except bipolar disorder), as also shown in hierarchical clustering analysis in which ADHD clustered with personality traits rather than psychiatric disorders (Supplementary Fig. 3). This may indicate that ADHD, or some ADHD subtypes, represent a variant of extraversion personality trait. Of note, our ADHD data consists of cases ranging in age from 5-19 years old. Phenotypically, positive emotionality has been linked with a subgroup of children with ADHD[27]. Future genetic studies considering ADHD heterogeneity (e.g., subtypes, child/adult ADHD) may help characterize its diverse etiologies and relationships with personality traits.

Overall, we observed a systematic pattern with all psychiatric disorders showing positive loadings on PC1, and agreeableness and conscientiousness with negative loadings. A combination of low agreeableness and low conscientiousness is thought to reflect Eysenk's psychoticism personality[4]. PC2 is closely aligned with extraversion-introversion which has been associated with externalizing/internalizing traits and activation/inhibition[28,29]. Internalizing traits (e.g., neuroticism, depression, anxiety and withdrawal)[21] have negative loadings on PC2. Externalizing traits are predicted by high extraversion, low agreeableness and low conscientiousness[29].

These findings provide additional support for shared genetic influences between

personality traits and psychiatric disorders[3,21,23] and for the notion that personality traits and psychiatric disorders exist on a continuum in phenotypic and genomic space[5,11]. Maladaptive or extreme variants of personality may contribute to the persistence of, or vulnerability to, psychiatric disorders and comorbidity[5,11,21,23]. Further genomic research in which categorical disease entities are viewed as variants of quantitative dimensions in a polygenic framework may help elucidate this issue[30].

Caveats of this study include that the sample size, while large, may still be underpowered to detect the majority of associated SNPs, given the conservative GWAS significance threshold. Because we used only summary statistics of GWAS, we cannot estimate non-additive genetic variance such as dominance and epistasis, and genetic contribution from structural (e.g., inversions) and rare variants. Additionally, genetic correlations indicate the degree of shared genetic influences across traits at the genome-wide level, but other studies using different methods are needed to identify specific pleiotropic variants underlying the observed correlations.

In summary, by studying all FFM traits we found six replicable genetic variants associated with personality, five of which are novel and one replicates a recently published finding[2,3]. We also observed that personality traits are correlated at the genetic level, with neuroticism showing an inverse association with the other traits. Other novel aspects of this study include description of the genetic correlations among five personality traits and six psychiatric disorders, and depiction of their relationships through principal component analysis. Personality traits are likely influenced by many gene variants and by gene-environment interactions. We are only in the beginning of understanding the genetics of personality and their relation to psychiatric disorders. The

overall effort promises to have great relevance to public health.

URLs.

LDlink, http://analysistools.nci.nih.gov/LDlink/?tab=ldpair;GPC-1 and GPC-2 summary statistics, http://www.tweelingenregister.org/GPC/; LocusZoom, http://locuszoom.sph.umich.edu/locuszoom/; The Bain eQTL Almanac (Braineac), http://www.braineac.org/; Psychiatric Genomics Consortium (PGC) summary statistics (schizophrenia, bipolar disorder, major depressive disorder, ADHD, autism spectrum disorder and anorexia nervosa), https://www.med.unc.edu/pgc/results-and-downloads; LD score regression, https://github.com/bulik/ldsc; GCTA-COJO (Genome-wide Complex Trait Analysis - Conditional and Joint Genome-wide Association Analysis), http://cnsgenomics.com/software/gcta/cojo.html; METAL (Meta-analysis of Genome-wide Association Scans), http://csg.sph.umich.edu//abecasis/metal/; PLINK 1.9, https://www.cog-genomics.org/plink2.

Data Availability Statement

The top 10K SNPs for five personality traits from the 23andMe discovery data set are available in Supplementary Data Sets 1-5. The full GWAS summary statistics for the 23andMe discovery data set will be made available through 23andMe to qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants. Please contact David Hinds (dhinds@23andme.com) for more information and to apply for data access.

Acknowledgments

We would like to thank the customers, research participants, and employees of 23andMe for making this work possible. This project was funded by National Institute of Mental Health R01MH100351 (M.-T. Lo, N. Sanyal, C.-H. Chen), NARSAD Young Investigator award (C.-H. Chen), South-East Norway Regional Health Authority (2016-064) (O.B. Smeland), and Research Council of Norway through a FRIPRO Mobility Grant, contract no. 251134 (Y. Wang). The FRIPRO Mobility grant scheme (FRICON) is co-funded by the European Union's Seventh Framework Programme for research, technological development and demonstration under Marie Curie grant agreement no. 608695. D.J. Smith is a Lister Institute Prize Fellow. The research leading to deCODE results was supported in part by NIH (NIDA) (R01-DA017932 and R01-DA034076) and the Innovative Medicines Initiative Joint Undertaking under grant agreement no. 115008 of which resources are composed of EFPIA in-kind contribution and financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EU funded FP7-People-2011-IAPP grant PsychDPC (GA 28613) (H. Stefansson, G. Bjornsdottir, T. E. Thorgeirsson and K. Stefansson).

Author Contributions

C.-H.C., M.-T.L. and O.A.A. designed the study. M.-T.L. and C.-H.C. analysed data and wrote the manuscript. D.A.H. and J.Y.T. analysed the 23andMe data. V.E.-P., D.J.S. and M.O'D. analysed the UK Biobank data. H.S., G.B., T.E.T and K.S. analysed the deCODE data. C.F., C.-C.F., Y.W., O.B.S., A.S. D.H., K.K., N.S., L.K.M., A.M.D. and O.A.A. contributed to manuscript preparation. All authors commented on and approved the manuscript.

Competing Financial Interests Statement

H.S., T.E.T., G.B. and K.S. are employees of deCODE Genetics/Amgen. D.A.H. and J.Y.T. are employees of 23andMe, Inc. The remaining authors declare no competing financial interests.

References

1. Vukasovic, T. & Bratko, D. Heritability of personality: A meta-analysis of behavior genetic studies. *Psychol Bull* **141**, 769-85 (2015).
2. Okbay, A. *et al.* Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat Genet* **48**, 624-33 (2016).
3. Smith, D.J. *et al.* Genome-wide analysis of over 106 000 individuals identifies 9 neuroticism-associated loci. *Mol Psychiatry* **21**, 749-57 (2016).
4. Goldberg, L.R. The structure of phenotypic personality traits. *Am Psychol* **48**, 26-34 (1993).
5. Trull, T.J. & Widiger, T.A. Dimensional models of personality: the five-factor model and the DSM-5. *Dialogues Clin Neurosci* **15**, 135-46 (2013).
6. de Moor, M.H. *et al.* Meta-analysis of genome-wide association studies for personality. *Mol Psychiatry* **17**, 337-49 (2012).
7. Genetics of Personality Consortium. *et al.* Meta-analysis of Genome-wide Association Studies for Neuroticism, and the Polygenic Association With Major Depressive Disorder. *JAMA Psychiatry* **72**, 642-50 (2015).
8. Kim, H.N. *et al.* Genome-wide association study of the five-factor model of personality in young Korean women. *Journal of Human Genetics* **58**, 667-674 (2013).
9. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-5 (2015).
10. Tabares-Seisdedos, R. & Rubenstein, J.L.R. Chromosome 8p as a potential hub for developmental neuropsychiatric disorders: implications for schizophrenia, autism and cancer. *Molecular Psychiatry* **14**, 563-589 (2009).
11. Greenwood, T.A., Akiskal, H.S., Akiskal, K.K., Study, B.G. & Kelsoe, J.R. Genome-Wide Association Study of Temperament in Bipolar Disorder Reveals Significant Associations with Three Novel Loci. *Biological Psychiatry* **72**, 303-310 (2012).
12. Green, E. *et al.* Localization of bipolar susceptibility locus by molecular genetic analysis of the chromosome 12q23-q24 region in two pedigrees with bipolar disorder and Darier's disease. *Am J Psychiatry* **162**, 35-42 (2005).
13. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-7 (2014).
14. Van Os, J. & Jones, P.B. Neuroticism as a risk factor for schizophrenia. *Psychological Medicine* **31**, 1129-1134 (2001).
15. Aberg, K. *et al.* Genomewide association study of movement-related adverse antipsychotic effects. *Biol Psychiatry* **67**, 279-82 (2010).
16. Rietveld, C.A. *et al.* Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 13790-13794 (2014).
17. Poropat, A.E. A meta-analysis of adult-rated child personality and academic performance in primary education. *British Journal of Educational Psychology* **84**,

239-252 (2014).
18. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* **10**, e1004383 (2014).
19. Power, R.A. & Pluess, M. Heritability estimates of the Big Five personality traits based on common genetic variants. *Transl Psychiatry* **5**, e604 (2015).
20. Barnett, J.H. *et al.* Personality and bipolar disorder: dissecting state and trait associations between mood and personality. *Psychological Medicine* **41**, 1593-1604 (2011).
21. 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. *American Journal of Psychiatry* **163**, 857-864 (2006).
22. Jakobwitz, S. & Egan, V. The dark triad and normal personality traits. *Personality and Individual Differences* **40**, 331-339 (2006).
23. Kendler, K.S. & Myers, J. The genetic and environmental relationship between major depression and the five-factor model of personality. *Psychological Medicine* **40**, 801-806 (2010).
24. Insel, T. *et al.* Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders. *American Journal of Psychiatry* **167**, 748-751 (2010).
25. Deyoung, C.G. The neuromodulator of exploration: A unifying theory of the role of dopamine in personality. *Front Hum Neurosci* **7**, 762 (2013).
26. Power, R.A. *et al.* Polygenic risk scores for schizophrenia and bipolar disorder predict creativity. *Nat Neurosci* **18**, 953-5 (2015).
27. Karalunas, S.L. *et al.* Subtyping attention-deficit/hyperactivity disorder using temperament dimensions: toward biologically based nosologic criteria. *JAMA Psychiatry* **71**, 1015-24 (2014).
28. Gray, J.A. The psychophysiological basis of introversion-extraversion. *Behav Res Ther* **8**, 249-66 (1970).
29. Mezquita, L. *et al.* Five-factor model and internalizing and externalizing syndromes: A 5-year prospective study. *Personality and Individual Differences* **79**, 98-103 (2015).
30. Plomin, R., Haworth, C.M. & Davis, O.S. Common disorders are quantitative traits. *Nat Rev Genet* **10**, 872-8 (2009).

Figure Legends

Figure 1. Manhattan plots for personality traits in the combined sample of 23andMe and GPC data (discovery/stage1 sample).

Sample size: Agreeableness: N=76,551; conscientiousness: N=76,551; extraversion: N=122,886; neuroticism: N=122,867; openness: N=76,581. Number of SNPs: Agreeableness: N=2,165,398; conscientiousness: N=2,166,809; extraversion: N=6,343,667; neuroticism: N=6,337,541; openness: N=2,167,320.

Figure 2. Regional association plot. The figure shows the distribution of $-\log_{10}$(p-value) of SNPs on chromosome 8p of the significant SNPs for neuroticism (a) and extraversion (b) in the combined discovery analysis. These two SNPs (LD r^2=0.5 in LDlink) have opposite signs of β's in GWAS results of neuroticism and extraversion. The opposite signals might be attributable to negative phenotypic association between neuroticism and extraversion. Regional plots with detailed annotation information for significant SNPs are also shown in Supplementary Fig. 4.

Figure 3. Genetic correlations between personality traits (23andMe sample) and psychiatric disorders. (a) The heat map illustrates genetic correlations between phenotypes. The values in the color squares correspond to genetic correlations. Asterisks denote genetic correlations significantly different from zero: * P<0.05; ** P<0.00091 (Bonferroni correction threshold). (b) The loading plot shows loadings of the personality traits and psychiatric disorders on the first two

402 principal components derived from the genetic correlation matrix on the
403 left. A small angle between arrows indicates a high correlation between
404 variables and arrows pointing to opposite directions indicate a negative
405 correlation in the space of the two principal components.
406
407

Table 1. 18 LD-independent genetic variants significantly associated with personality traits

					Discovery/Stage 1								Replication/Stage 2									Final combined analysis of stage 1 and stage 2		
					23andMe (N = ～59,200)			GPC (N = 17,375/ 63,661)†			Combined analysis		23andMe replication (N = ～39,500)			deCODE (N = ～7,100)			UK Biobank (N = 91,370)					
SNP	Chr	Closest gene (region)	A1/ A2	Frq	β	SE	P-value	β	SE	P-value	P-value	N	β	SE	P-value	β	SE	P-value	β	SE	P-value	P-value	N	R^2 (%)
Conscientiousness																								
rs3814424	5q	*LINC00461**	T/C	0.17	-0.289	0.050	9.75×10^{-9}	-0.138	0.131	0.294	2.98×10^{-8}	76,551	-0.051	0.051	0.313	-0.005	0.027	0.855				6.19×10^{-7}	123,132	0.0202
Extraversion																								
rs57590327	3p	*GBE1* (intergenic)	T/G	0.26	0.236	0.054	1.37×10^{-5}	0.026	0.006	2.03×10^{-5}	1.61×10^{-9}	122,886	0.088	0.052	0.091	0.007	0.019	0.713				1.26×10^{-9}	169,507	0.0217
rs2164273	8p	*MTMR9* (intron)	G/A	0.39	0.179	0.047	1.14×10^{-4}	0.024	0.006	4.08×10^{-5}	1.79×10^{-8}	122,845	0.093	0.045	0.037	0.021	0.018	0.255				1.61×10^{-9}	169,466	0.0215
rs6481128	10q	*PCDH15* (intergenic)	G/A	0.45	0.205	0.046	7.10×10^{-6}	0.018	0.005	0.0010	4.15×10^{-8}	122,886	0.154	0.045	5.58×10^{-4}	-0.011	0.017	0.528				5.44×10^{-10}	169,507	0.0227
rs1426371	12q	*WSCD2* (intron)	A/G	0.28	-0.308	0.053	4.65×10^{-9}	-0.023	0.006	2.56×10^{-4}	2.09×10^{-11}	122,886	-0.177	0.051	5.09×10^{-4}	-0.037	0.021	0.077				9.54×10^{-15}	169,507	0.0354
rs7498702	16p	*RBFOX1* (intron)	C/T	0.29	-0.166	0.050	8.94×10^{-4}	-0.026	0.006	1.17×10^{-5}	4.73×10^{-8}	122,886	-0.006	0.048	0.907	-0.005	0.018	0.777				1.89×10^{-6}	169,507	0.0134
Neuroticism																								
rs6981523	8p	*XKR6* (intergenic)	T/C	0.50	0.250	0.042	2.68×10^{-9}	0.022	0.006	1.01×10^{-4}	4.25×10^{-12}	122,867	0.138	0.042	1.05×10^{-3}	0.032	0.018	0.070	0.098	0.015	1.04×10^{-10}	3.17×10^{-24}	260,861	0.0395
rs9611519	22q	*L3MBTL2* (exon) *CHADL* (intron)	T/C	0.31	0.235	0.046	4.05×10^{-7}	0.020	0.007	0.003	1.87×10^{-8}	122,867	0.002	0.047	0.966	-0.002	0.023	0.931	0.053‡	0.017‡	0.0015‡	9.16×10^{-9}	260,861	0.0127

Chr: chromosome; A1: effect allele; A2: non-effect allele; Frq: allele frequency of A1; β: linear regression association coefficient; SE: standard error; N: sample size. β and SE may have varying scales in different cohorts, thus sample-based meta-analyses were used.

*SNP in non-protein coding region.

†The sample sizes of GPC1 and GPC2 are 17,375 and 63,661, respectively.

‡Due to absence of rs9611519 in the UK Biobank data, a proxy SNP (rs2273085, LD $r^2 = 0.99$) was used.

Online Methods

23andMe sample

The GWAS summary statistics were obtained from a subset of 23andMe participants. 23andMe uses a survey design to collect a number of phenotypes including the personality traits reported here, and the sample has been described previously for other phenotypes[31,32]. We included only those participants (N=59,225) who showed >97% European ancestry as determined by analyzing local ancestry and comparing to three HapMap 2 populations[33]. Relatedness between participants was examined by a segmental identity-by-descent (IBD) method[34] to ensure that only unrelated individuals (sharing less than 700 cM IBD) were included in the sample. All participants included in the analyses provided informed consent and answered surveys online according to a human subject research protocol, which was reviewed and approved by Ethical & Independent Review Services, an AAHRPP-accredited private institutional review board (http://www.eandireview.com).

Additionally, we obtained independent replication results of GWAS from 23andMe replication sample. This sample included ~39,500 participants (N=39,452 for conscientiousness, 39,484 for extraversion and 39,488 for neuroticism) who met the same inclusion criteria as described above.

Genetics of Personality Consortium (GPC) sample

Genetics of Personality Consortium (GPC) is a large collaboration of GWAS for personality. Summary statistics of the PGC data used in the current study included the first meta-analysis of GWAS (GPC-1)[6] for three traits (agreeableness, conscientiousness and openness) and the second meta-analysis of GWAS (GPC-2) for neuroticism[7,35,36] and

444 extraversion. The results of 10 discovery cohorts for GPC-1 and of 29 discovery cohorts
445 for GPC-2 are available in the public domain, which respectively consist of 17,375 and
446 63,661 participants with European ancestry across Europe, Australia and United States.
447 These studies were performed with oversight from local ethic committees, and all
448 participants provided informed consent[6,7,35,36].

449 **UK Biobank sample**

450 UK (United Kingdom) Biobank is a large prospective cohort of more than 502,000
451 participants (aged 40-69 years)[3] with genetic data and a wide range of phenotypic data
452 including social, cognitive, personality (neuroticism trait), lifestyle, and physical health
453 measures collected at baseline. We used a subsample of this cohort for neuroticism
454 replication. Exclusion criteria included UK Biobank genomic analysis exclusions,
455 relatedness, gender mismatch, non-white UK ancestry and failure of quality control of
456 UK BiLEVE genotyping[3], resulting in a sample of 91,370 individuals. Association
457 analysis was conducted using linear regression under a model of additive allelic effects
458 with sex, age, array and the first eight PCs as covariates[3]. Informed consent was obtained
459 from all participants and the study was approved by National Health Service National
460 Research Ethics Service[3].

461 **deCODE sample**

462 Icelandic participants (N=7,137 for extraversion, 7,136 for neuroticism and 7,129
463 for conscientiousness) were enrolled in various ongoing deCODE studies administering
464 the NEO-FFI measure of the Big Five personality traits[37,38]. All deCODE studies were
465 approved by the appropriate bioethics and data protection authorities and all participating
466 subjects donating blood signed informed consent forms. The personal identities of

participants from whom phenotype information and biological samples were obtained were encrypted by a third-party system overseen by the Icelandic Data Protection Authority[39]. A generalized form of linear regression that accounts for relatedness between individuals was used to test the correlation between normalized NEO-FFI trait scores and genotypes.

Personality assessment

In the 23andMe sample, individuals completed a web-based implementation of the Big Five Inventory (BFI)[40,41], with 44 questions. Scores for agreeableness, conscientiousness, extraversion, neuroticism, and openness were computed using 8 to10 items per factor[40].

In GPC-1, scores of personality traits were based on the 60 item NEO Five-Factor Inventory (NEO-FFI) with 12 items per factor[6,37]. In GPC-2, harmonization of measures for neuroticism and extraversion across 9 inventories and 29 cohorts were performed by applying Item Response Theory (IRT) to avoid personality scores being influenced by the number of items and the specific inventory. Because the personality measures were not assessed similarly across GPC-2 cohorts, the harmonized/calibrated scores of personality are more comparable, thereby increasing power for meta-analysis of GWAS using fixed-effect models[7,35,36]. As described in the main text, high genetic correlations between 23andMe and GPC samples were found, suggesting a highly consistent pattern of associations despite the discrepancy in questionnaires (Supplementary Fig. 1).

In the UK Biobank sample, neuroticism was scored between 0 to 12 using the 12 items of the Eysenck Personality Questionnaire-Revised Short Form (EPQ-R-S)[42] with high reliability and concurrent validity[42].

In the deCODE sample, NEO-FFI personality trait scores[37,38] were adjusted for sex and age at measurement and were then normalized to a standard normal distribution using quantile normalization.

Distributions and correlations for personality scores in the 23andMe sample

Quantile-quantile (QQ) plots of covariate-adjusted personality scores to examine normality are shown in Supplementary Fig. 5. The distributions at the top tail deviates from normality due to the limited range of the scores and those at the bottom tail deviate due to the limited range (for neuroticism and extraversion) and/or extreme values. This violation of the normality assumption can be influential for genetic variants with very low minor allele frequencies (e.g., rare variants)[43]. However, this did not affect our results because our GWAS and LD Score regression[9] only include common variants.

Pearson correlations, unadjusted and after adjusting for the covariates (age, sex, top five principal components for population structure correction[44]), were used to assess phenotypic correlations among the five traits (Supplementary Table 3).

Genotyping and imputation

In the 23andMe sample, DNA extraction and genotyping were performed on saliva samples by National Genetics Institute (NGI), a CLIA licensed clinical laboratory and a subsidiary of Laboratory Corporation of America. Samples have been genotyped on one of four genotyping platforms. The V1 and V2 platforms were variants of the Illumina HumanHap550+ BeadChip, including about 25,000 custom SNPs selected by 23andMe, with a total of about 560,000 SNPs. The V3 platform was based on the Illumina OmniExpress+ BeadChip, with custom content to improve the overlap with 23andMe's V2 array, with a total of about 950,000 SNPs. The 23andMe's V4 platform in current use

is a fully custom array, including a lower redundancy subset of V2 and V3 SNPs with additional coverage of lower-frequency coding variation, and about 570,000 SNPs. Samples that failed to reach 98.5% call rate were re-analyzed. As part of 23andMe standard practice, individuals whose analyses failed repeatedly were re-contacted and asked to provide a new sample.

23andMe participant genotype data were imputed using the 1000 Genomes Project phase 1 version 3 reference panel[45]. The phasing and imputation for each genotyping platform were separated. First, chromosomal segments of no more than 10,000 genotyped SNPs, with overlaps of 200 SNPs, were phased using Beagle (version 3.3.1)[46]. Then, each phased segment was imputed against all-ethnicity 1000 Genomes Project haplotypes (excluding monomorphic and singleton sites) using a high-performance version of Minimac[47] for 5 rounds and 200 states to estimate parameters. SNPs were filtered by procedures including Hardy-Weinberg equilibrium $P<10^{-20}$ (stringent threshold for large sample size), call rate<95% and allele frequencies apparently different from European 1000 Genomes Project reference data. A total of 13,341,935 SNPs was retained after filtering and excluding chromosome X, Y and mitochondria. We focus on autosomal SNPs, which are available for 23andMe, GPC and UK Biobank samples.

Genotyping in cohorts of GPC-1[6] and GPC-2[7,35] was conducted on Illumina or Affymetrix platforms. Quality control of genotype data was examined in each cohort independently, including checks for European ancestry, sex inconsistencies, Mendelian errors, high genome-wide homozygosity, relatedness, minor allele frequencies (MAF), SNP call rate, sample call rate and Hardy-Weinberg equilibrium[6,7,35,36]. Genotype data of GPC-1 were then imputed using HapMap phase II CEU as a reference panel including

~2.5M SNPs[6] and, alternatively, a reference panel from 1000 Genomes Project phase 1 version 3 was used to impute the genotype data of GPC-2[7,35,36]. Poorly imputed SNPs (r^2<0.3 or proper_info<0.3[6] or 0.4[7,35]) and low MAF (<0.01[6] or $\sqrt{5/N}$ [7,35]) were excluded in the meta-analyses, resulting in a total number of 1.1-6.6 million SNPs[7,35] across cohorts.

In the UK Biobank first release genetic data of 152,729 participants (June 2015), about two thirds of the sample was genotyped using Affymetrix UK Biobank Axiom array (Santa Clara, CA, USA) and the remaining were genotyped using the Affymetrix UK BiLEVE Axiom array[3]. Outlier, multi-allelic and low-MAF (<1%) SNPs were excluded from phasing and imputation procedures. The reference panel of imputation was based on the 1000 Genomes Phase 3 and UK10K haplotype panels[3]. Further quality control procedures were applied after imputation, yielding a total of 8,268,322 SNPs for further analyses[3].

Genotyping, imputation methods and the association analysis method used in the deCODE sample are previously described[48]. A total of 676,913 autosomal SNPs were typed using Illumina SNP chips[48]. SNPs with low MAF (<0.1%) and low imputation information (<0.8) were excluded and 99.5% of SNPs remained after imputation.

Genome-wide association analysis

Association tests were performed by regressing personality traits on imputed dosages of SNPs in the 23andMe sample. Age, sex, and the top five principal components (PCs)[44] for population structure correction were included as covariates and p-values were computed using likelihood ratio tests. For all five personality traits, the correlation structure of SNPs was determined by an LD matrix of 9,270,523 autosomal SNPs

559 generated from European reference sample in 1000 Genomes Project phase 1 v3 within
560 1,000,000 base pairs (1 Mb)[49,50] using Plink 1.07[51]. The original 13,341,935 SNPs were
561 reduced into 9,270,523 SNPs in our subsequent analyses (e.g., LD correlation structure is
562 used to determine LD-independent SNPs). All SNPs' positions were mapped to Genome
563 Reference Consortium Human Build 37 (GRCh37) and UCSC Genome Browser on
564 Human hg19 assembly. We made QQ plots with GWAS summary statistics of the
565 23andMe sample. The QQ plots lie along the expected null line for large p-values ($P>10^{-3}$),
566
567 indicating that the GWAS results are not inflated by population stratification or cryptic
568 relatedness. This pattern is consistent with the genomic inflation factors (λ)[52] close to 1,
569 as shown in Supplementary Fig. 6.

570 In each cohort of GPC-1[6] and GPC-2[7,35], linear regressions with covariates of sex,
571 age and PCs were conducted for association tests using dosage data. The meta-analyses
572 of GWAS results of cohorts for GPC-1 and GPC-2 were performed by the inverse-
573 variance method using METAL[53] released on the GPC website (see URLs). Given
574 improved power for detection of genetic effects with larger sample sizes in GWAS, we
575 performed a combined meta-analysis of 23andMe and GPC samples using METAL[53]
576 based on the sample-size based method. SNPs available in one cohort only were
577 excluded. The totals of 2,305,461, 2,305,682 and 2,305,640 SNPs were available for
578 traits of agreeableness, conscientiousness and openness (respectively) in GPC-1, as well
579 as 6,941,603 SNPs for extraversion and 6,949,614 SNPs for neuroticism in GPC-2.
580 Genomic inflation factors (λ) are 1.01, 1.01, 1.03, 1.02 and 1.02 for agreeableness,
581 conscientiousness, extraversion, neuroticism and openness, respectively.

Meta-analysis of 23andMe and GPC samples

582 Given improved power for detection of genetic effects with larger sample sizes in
583 GWAS, we performed a combined meta-analysis of 23andMe and GPC samples using
584 METAL[53] based on the sample-size based method. To assess the quality of meta-analysis,
585 SNPs with heterogeneity p-values<0.05 were excluded. Eight significant LD-independent
586 SNPs were identified after removing correlated SNPs at LD r^2>0.05 that are within 1 Mb
587 of the top SNP. In Table 1, the percentage of variance explained by each SNP is
588 calculated using equation: $(z^2/(n\text{-}k\text{-}1+z^2))\times 100$, where z is the z value for each SNP
589 controlling for covariates, n is the sample size for each SNP and k is the number of
590 covariates in the regression model (k=7 for age, sex, and top five PCs)[54,55].

591 **Conditional analysis within 1 Mb region of significant SNPs**

592 We performed a conditional analysis[56] within the 1 Mb genomic region of each of
593 the six LD-independent SNPs. In our study, we used 1000 Genomes Project reference
594 panel of European ancestry to estimate LD correlations (r^2) and excluded SNPs correlated
595 at LD r^2>0.9 with the top associated SNP within 1 Mb window. We did not detect
596 additional significant SNPs conditional on the top SNPs under the stringent GWAS
597 threshold. However, for the significant loci in 8p, several SNPs still showed substantial
598 association signals ($P\sim10^{-7}$) conditioning on the top SNPs, rs6981523 or rs2164273.

599 **Regional association and annotation plot**

600 The regional plot of chromosome 8p (Fig. 2) was constructed by a web-interface
601 tool, LocusZoom[57]. In Fig. 2a and 2b, the most significant SNPs (rs6981523 and
602 rs2164273) are shown in purple, otherwise the colors of the circles denote their
603 correlations (LD r^2) with the top SNP. The bottom panel displays gene symbol and
604 location within the region derived from UCSC Genome Browser on Human hg19

assembly. The regional and annotation plots for other significant SNPs are also shown in Supplementary Fig. 4.

Genetic correlation analysis

We used the LD Score regression method to examine the pattern of genetic correlations (r_g)[9,58] across personality traits within/between 23andMe and GPC samples (Fig. 3a, Supplementary Fig. 1 and Supplementary Table 4) based on GWAS summary statistics. The LD Score for each SNP measures the amount of pair-wise LD r^2 with other SNPs within 1-cM windows from 1000 Genomes Project reference panel of European ancestry. All SNPs were filtered by LD Score regression built-in procedures, including INFO > 0.9 and MAF > 0.1, and merged to SNPs in HapMap 3 reference panel. Approximately 0.8-1.1 million SNPs (Supplementary Table 2) were retained to estimate genetic correlations.

We also examined genetic correlations among the five traits, which have been estimated previously using a twin design[59,60], and unrelated individuals' SNP data from a relatively smaller sample, in which many estimates did not converge[19]. Our LD Score regression analysis based on a large sample provided additional contribution to this effort.

We further quantified genetic correlations between personality traits and psychiatric disorders, including schizophrenia[61], bipolar disorder[62], major depressive disorder[63], ADHD[61], autism spectrum disorder[61] and anorexia nervosa[64].

Query for eQTL Database

We queried eQTL evidence for our significant SNPs from Braineac[65,66] (the Brain eQTL Almanac). The results are listed in Supplementary Table 1. We display the brain region with the lowest p-value among all 10 regions. To check the rank of eQTL p-values

of six LD-independent SNPs in the Braineac database, we randomly selected 50,000 SNPs and queried the database to extract the lowest p-value for each SNP, resulting in a total of 36,190 SNPs with eQTL results. In order to match allele frequency and distances to transcription start site (TSS) with the significant SNPs, the randomly selected SNPs were stratified into four groups: (1) within transcript, (2) downstream 0-200 kilobase pairs (kb), (3) upstream 0-200 kb and (4) upstream 200-400 kb. SNPs that fell outside these ranges were removed. The SNPs in the 'within transcript' group were further stratified into three subgroups according to allele frequency. This procedure resulted in six distributions of eQTL p-values that matched the significant SNPs in terms of allele frequency and TSS, and these were used to determine the ranking of eQTL associations (see Supplementary Table 1 & 5). Two SNPs are ranked high for their significance as eQTL compared to randomly sampled eQTL markers with matched allele frequencies and distance to TTS from the Braineac database (top 10-20% ranking: rs6981523 and top 20-30% ranking: rs9611519; see Supplementary Table 5).

Colocalisation analysis between GWAS and eQTL

To investigate whether GWAS significant SNPs and their eQTL are colocalised with a shared candidate causal variant, we performed a colocalisation analysis, COLOC, that use Bayesian posterior probability to assess colocalisation[18]. The SNP-associated locus was defined as within a 1 Mb window[18] for each of the six SNPs (Table 1). The prior probabilities that the locus is associated with only trait 1 (i.e., personality traits), only trait 2 (i.e., eQTL) and both are respectively 10^{-5}, 10^{-4} and 10^{-6}. The posterior probabilities (PP0, PP1, PP2, PP3 and PP4) for five hypotheses (H_0: no association with either trait; H_1: association with trait 1, not with trait 2; H_2: association with trait 2, not with trait 1;

H_3: independent association with two traits, two independent SNPs; H_4: association with both traits, one shared SNP)[18] were calculated to determine which hypothesis is supported by the data. A limitation of this analysis is the potentially low power in the small eQTL sample (N=134).

SNP concordant test for the top GWAS signals

To investigate concordance of SNP effects between personality traits and psychiatric disorders, we followed a similar procedure described previously[67,68] by counting the number of same direction effect sizes for the LD-independent top SNPs ($P<10^{-4}$) in the pairwise phenotypes data and calculated the proportion of the same direction effects in the total number of LD-independent top SNPs. The one-sided p-value for the proportion of each pairwise phenotype was computed using a binomial test to examine the deviation from 0.5 for the proportion. In Supplementary Fig. 2, a heat map of the proportions of the same direction effect for pairwise phenotypes shows a similar pattern with a heat map of genetic correlations in Fig 3a.

Hierarchical clustering analysis

We performed hierarchical clustering analysis using dissimilarity measures (1-genetic correlation) implemented in hclust function of R to investigate and display relationships between personality traits and psychiatric disorders. Based on genetic correlations, the more highly correlated phenotypes were grouped in the same clusters and displayed by a dendrogram (Supplementary Fig. 3), showing an agreement with classifications of the loading plot (Fig. 3b).

674

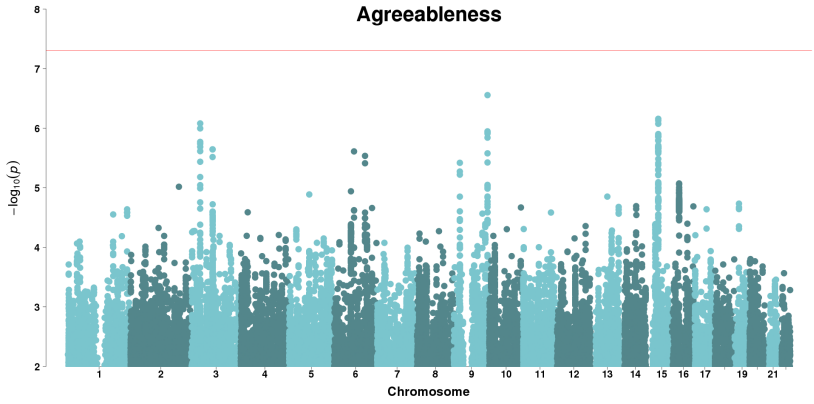
675

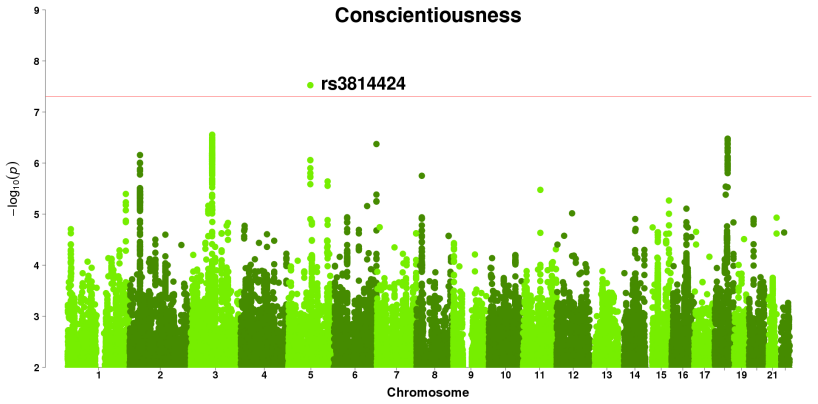
Methods-only References

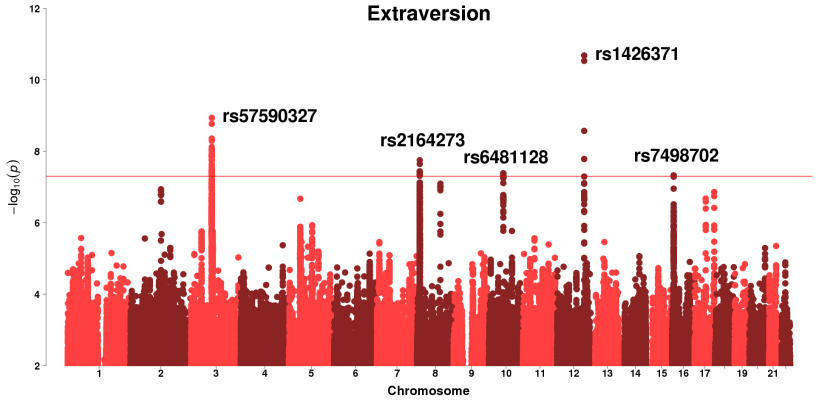
31. Hu, Y. *et al.* GWAS of 89,283 individuals identifies genetic variants associated with self-reporting of being a morning person. *Nat Commun* **7**, 10448 (2016).
32. Pickrell, J.K. *et al.* Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet* **48**, 709-17 (2016).
33. Falush, D., Stephens, M. & Pritchard, J.K. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* **164**, 1567-1587 (2003).
34. Henn, B.M. *et al.* Cryptic distant relatives are common in both isolated and cosmopolitan genetic samples. *PLoS One* **7**, e34267 (2012).
35. van den Berg, S.M. *et al.* Meta-analysis of Genome-Wide Association Studies for Extraversion: Findings from the Genetics of Personality Consortium. *Behav Genet* **46**, 170-82 (2016).
36. van den Berg, S.M. *et al.* Harmonization of Neuroticism and Extraversion phenotypes across inventories and cohorts in the Genetics of Personality Consortium: an application of Item Response Theory. *Behav Genet* **44**, 295-313 (2014).
37. Costa, P.T. & McCrae, R.R. *Revised NEO personality inventory (NEO PI-RTM) and NEO five-factor inventory (NEO-FFI) : professional manual*, vi, 101 p. (Psychological Assessment Resources, Florida, 1992).
38. Bjornsdottir, G. *et al.* Psychometric properties of the Icelandic NEO-FFI in a general population sample compared to a sample recruited for a study on the genetics of addiction. *Pers Individ Dif* **58**(2014).
39. Gulcher, J.R., Kristjansson, K., Gudbjartsson, H. & Stefansson, K. Protection of privacy by third-party encryption in genetic research in Iceland. *Eur J Hum Genet* **8**, 739-42 (2000).
40. John, O.P.D., E. M.; Kentle, R. L. The Big Five Inventory--Versions 4a and 54. *Berkeley, CA: University of California,Berkeley, Institute of Personality and Social Research.* (1991).
41. 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. *Journal of Research in Personality* **43**, 84-90 (2009).
42. Eysenck, S.B.G., Eysenck, H.J. & Barrett, P. A Revised Version of the Psychoticism Scale. *Personality and Individual Differences* **6**, 21-29 (1985).
43. Buzkova, P. Linear regression in genetic association studies. *PLoS One* **8**, e56976 (2013).
44. Lehoucq, R.B., Sorensen, D.C., Yang, C. & Society for Industrial and Applied Mathematics. ARPACK users' guide solution of large-scale eigenvalue problems with implicitly restarted Arnoldi methods. in *Software, environments, tools 6* 1 electronic text (xv, 142 p (Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104),, Philadelphia, Pa., 1998).
45. 1000 Genomes Project Consortium *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061-73 (2010).
46. Browning, B.L. & Browning, S.R. Rapid and accurate haplotype phasing and missing data inference for whole genome association studies using localized

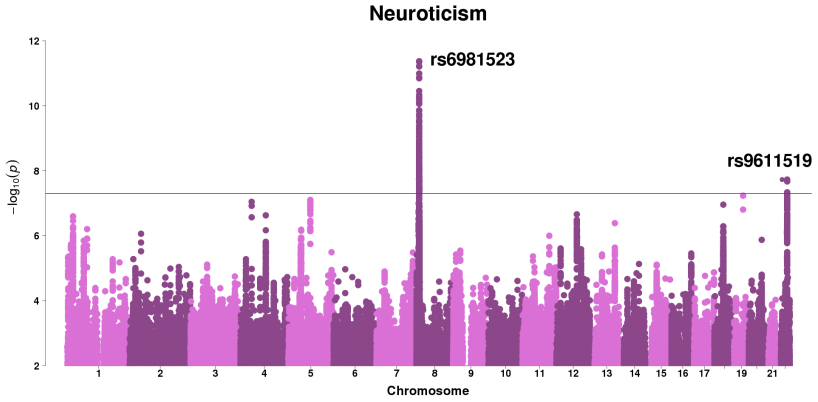
haplotype clustering. *Genetic Epidemiology* **31**, 606-606 (2007).
47. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G.R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet* **44**, 955-9 (2012).
48. Gudbjartsson, D.F. *et al.* Large-scale whole-genome sequencing of the Icelandic population. *Nat Genet* **47**, 435-44 (2015).
49. Thompson, W.K. *et al.* An Empirical Bayes Mixture Model for Effect Size Distributions in Genome-Wide Association Studies. *PLoS Genet* **11**, e1005717 (2015).
50. Wang, Y. *et al.* Leveraging Genomic Annotations and Pleiotropic Enrichment for Improved Replication Rates in Schizophrenia GWAS. *PLoS Genet* **12**, e1005803 (2016).
51. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
52. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997-1004 (1999).
53. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
54. Hibar, D.P. *et al.* Common genetic variants influence human subcortical brain structures. *Nature* **520**, 224-9 (2015).
55. So, H.C., Li, M. & Sham, P.C. Uncovering the total heritability explained by all true susceptibility variants in a genome-wide association study. *Genet Epidemiol* **35**, 447-56 (2011).
56. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369-75, S1-3 (2012).
57. Pruim, R.J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336-7 (2010).
58. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236-41 (2015).
59. Distel, M.A. *et al.* The five-factor model of personality and borderline personality disorder: a genetic analysis of comorbidity. *Biol Psychiatry* **66**, 1131-8 (2009).
60. Ono, Y. *et al.* Genetic structure of the five-factor model of personality in a Japanese twin population. *Keio J Med* **49**, 152-8 (2000).
61. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381**, 1371-9 (2013).
62. Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet* **43**, 977-83 (2011).
63. Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium *et al.* A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry* **18**, 497-511 (2013).
64. Boraska, V. *et al.* A genome-wide association study of anorexia nervosa. *Mol Psychiatry* **19**, 1085-94 (2014).
65. Rosander, P. & Backstrom, M. Personality traits measured at baseline can predict

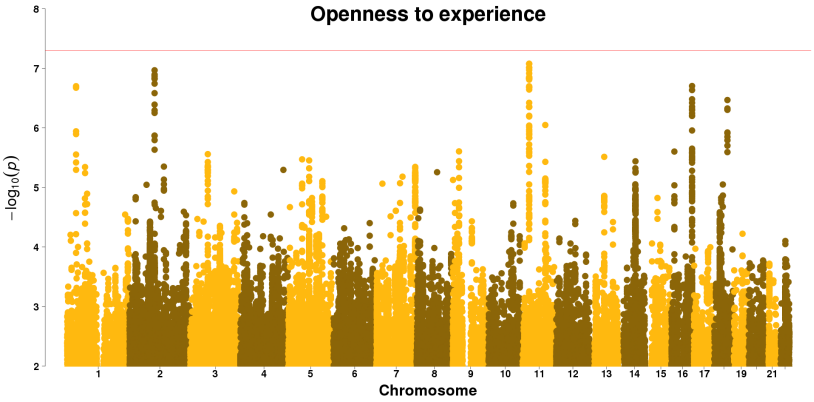
academic performance in upper secondary school three years late. *Scandinavian Journal of Psychology* **55**, 611-618 (2014).
66. Trabzuni, D. *et al.* Quality control parameters on a large dataset of regionally dissected human control brains for whole genome expression studies. *Journal of Neurochemistry* **119**, 275-282 (2011).
67. Franke, B. *et al.* Genetic influences on schizophrenia and subcortical brain volumes: large-scale proof of concept. *Nat Neurosci* **19**, 420-31 (2016).
68. Nyholt, D.R. SECA: SNP effect concordance analysis using genome-wide association summary results. *Bioinformatics* **30**, 2086-8 (2014).

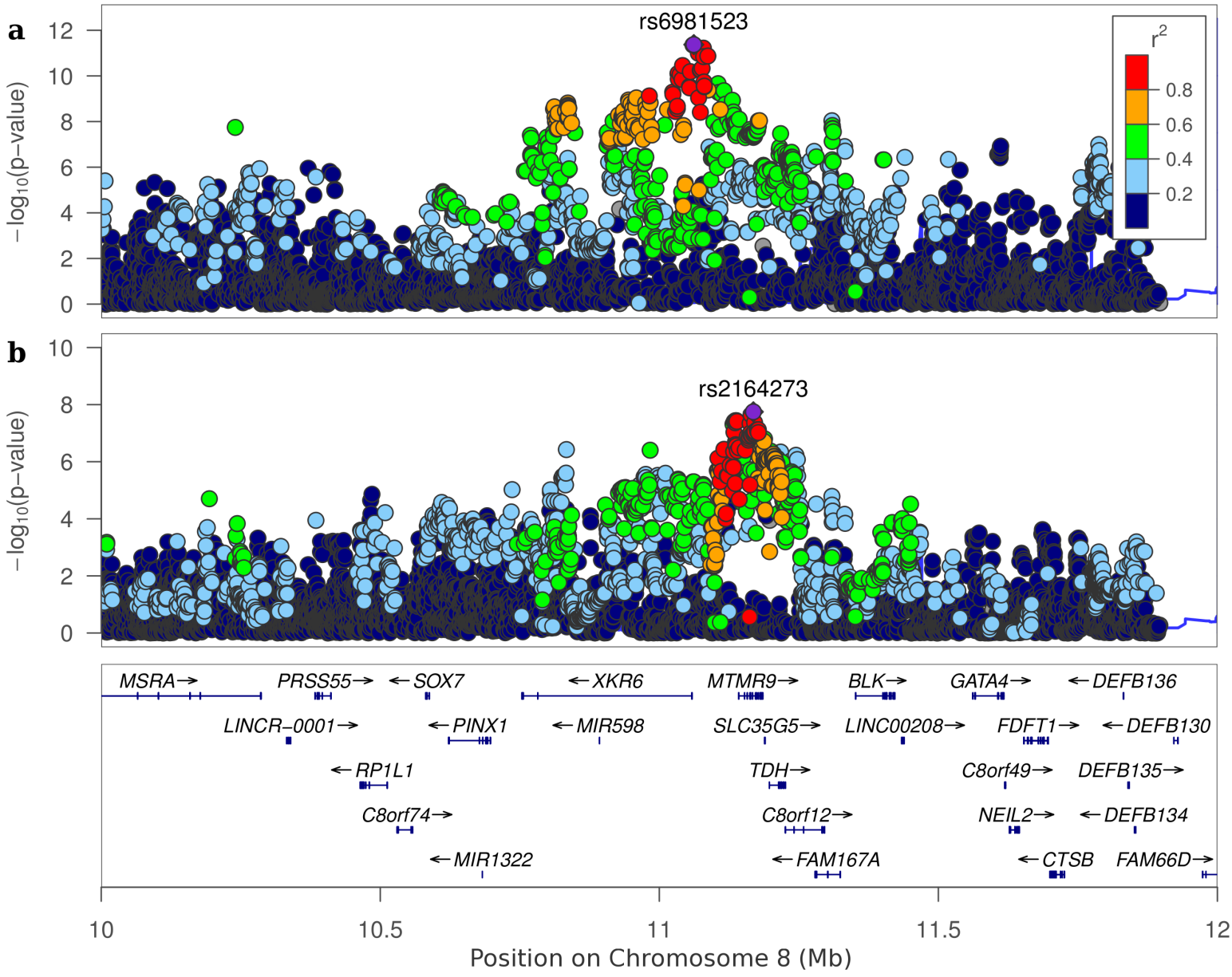

Agreeableness
$-\log_{10}(p)$
Chromosome


Conscientiousness
rs3814424
$-\log_{10}(p)$
Chromosome


Extraversion
rs1426371
rs57590327
rs2164273
rs6481128
rs7498702
$-\log_{10}(p)$
Chromosome


Neuroticism
rs6981523
rs9611519
$-\log_{10}(p)$
Chromosome


Openness to experience
$-\log_{10}(p)$
Chromosome


a
rs6981523
r²
0.8
0.6
0.4
0.2
−log10(p−value)
b
rs2164273
−log10(p−value)
MSRA→
PRSS55→
←SOX7
←XKR6
MTMR9→
BLK→
GATA4→
←DEFB136
LINCR−0001→
←PINX1
←MIR598
SLC35G5→
LINC00208→
FDFT1→
←DEFB130
←RP1L1
TDH→
C8orf49→
DEFB135→
C8orf74→
C8orf12→
NEIL2→
←DEFB134
←MIR1322
←FAM167A
←CTSB
FAM66D→
10
10.5
11
11.5
12
Position on Chromosome 8 (Mb)

a

Genetic correlation

	1	2	3	4	5	6	7	8	9	10	11
1. Agreeableness	1.00	0.23**	0.22**	-0.40**	0.11	-0.03	0.07	-0.29*	0.08	-0.24*	-0.03
2. Conscientiousness	0.23**	1.00	0.15*	-0.18*	-0.19*	-0.13*	-0.18*	-0.28*	-0.10	-0.21*	0.01
3. Extraversion	0.22**	0.15*	1.00	-0.35**	0.34**	-0.01	0.18*	0.02	0.30*	-0.04	-0.05
4. Neuroticism	-0.40**	-0.18*	-0.35**	1.00	-0.15*	0.14	-0.01	0.56**	0.06	0.10	0.15**
5. Openness to experience	0.11	-0.19*	0.34**	-0.15*	1.00	0.36**	0.34**	0.28*	0.19	0.12	0.09
6. Schizophrenia	-0.03	-0.13*	-0.01	0.14	0.36**	1.00	0.65**	0.47**	-0.03	0.11	0.21**
7. Bipolar disorder	0.07	-0.18*	0.18*	-0.01	0.34**	0.65**	1.00	0.52**	0.15	0.07	0.16*
8. Major depression	-0.29*	-0.28*	0.02	0.56**	0.28*	0.47**	0.52**	1.00	-0.12	0.12	0.17
9. Attention-deficit/ hyperactivity disorder	0.08	-0.10	0.30*	0.06	0.19	-0.03	0.15	-0.12	1.00	-0.10	-0.06
10. Autism spectrum disorder	-0.24*	-0.21*	-0.04	0.10	0.12	0.11	0.07	0.12	-0.10	1.00	0.06
11. Anorexia nervosa	-0.03	0.01	-0.05	0.15**	0.09	0.21**	0.16*	0.17	-0.06	0.06	1.00

b