



Published in final edited form as:

*Nat Genet.* 2016 June ; 48(6): 624–633. doi:10.1038/ng.3552.

## Genetic variants associated with subjective well-being, depressive symptoms and neuroticism identified through genome-wide analyses

*A full list of authors and affiliations appears at the end of the article.*

### Abstract

We conducted genome-wide association studies of three phenotypes: subjective well-being ( $N=298,420$ ), depressive symptoms ( $N=161,460$ ), and neuroticism ( $N=170,910$ ). We identified three variants associated with subjective well-being, two with depressive symptoms, and eleven with neuroticism, including two inversion polymorphisms. The two depressive symptoms loci replicate in an independent depression sample. Joint analyses that exploit the high genetic correlations between the phenotypes ( $|\rho| \approx 0.8$ ) strengthen the overall credibility of the findings, and allow us to identify additional variants. Across our phenotypes, loci regulating expression in central nervous system and adrenal/pancreas tissues are strongly enriched for association.

### INTRODUCTION

Subjective well-being—as measured by survey questions on life satisfaction, positive affect, or happiness—is a major topic of research within psychology, economics, and epidemiology. Twin studies have found that subjective well-being is genetically correlated with depression (characterized by negative affect, anxiety, low energy, bodily aches and pains, pessimism, and other symptoms) and neuroticism (a personality trait characterized by easily experiencing negative emotions such as anxiety and fear)<sup>1–3</sup>. Depression and neuroticism

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: [http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

CORRESPONDING AUTHORS: Daniel Benjamin ([djbenjam@usc.edu](mailto:djbenjam@usc.edu)) or Meike Bartels ([m.bartels@vu.nl](mailto:m.bartels@vu.nl)) or Philipp Koellinger ([p.d.koellinger@vu.nl](mailto:p.d.koellinger@vu.nl)).

\*These authors contributed equally

#Designed and oversaw the study.

#### URLs:

Genotype-Tissue Expression Portal [www.GTExpress.org](http://www.GTExpress.org)

Social Science Genetic Association Consortium (SSGAC) website: <http://www.thessgac.org/#!data/kuzq8>.

**ACCESSION CODES:** For neuroticism and depressive symptoms, we provide meta-analysis results from the combined analyses for all variants. For subjective well-being, meta-analysis results for all variants are provided for the full sample excluding 23andMe, for which only up to 10,000 SNPs can be reported. Therefore, for the subjective well-being meta-analysis, we provide results for 10,000 SNPs. Meta-analysis results can be downloaded from the SSGAC website.

**AUTHOR CONTRIBUTIONS:** M.B., D.J.B., D.C., J.E.D.N., P.D.K., and R.F.K. designed and oversaw the study. A.O. and B.B. were responsible for quality control and meta-analyses. Bioinformatics analyses were carried out by J.B., T.E., M.A.F., J.R.G., J.L., S.F.W.M., M.N., and H.J.W. Other follow-up analyses were conducted by M.A.F., J.B., P.T., A.O., B.B., and R.K.L. Especially major contributions to the writing and editing were made by M.B., D.J.B., J.B., D.C., J.E.D.N., P.K., A.O., and P.T. All authors contributed to and critically reviewed the manuscript.

**COMPETING FINANCIAL INTERESTS:** The authors declare no competing financial interests.

have received much more attention than subjective well-being in genetic-association studies, but the discovery of associated genetic variants with either of them has proven elusive<sup>4,5</sup>.

In this paper, we report a series of separate and joint analyses of subjective well-being, depressive symptoms, and neuroticism. Our primary analysis is a genome-wide association study (GWAS) of subjective well-being based on data from 59 cohorts ( $N=298,420$ ). This GWAS identifies three loci associated with subjective well-being at genome-wide significance ( $p < 5 \times 10^{-8}$ ). We supplement this primary analysis with auxiliary GWAS meta-analyses of depressive symptoms ( $N=180,866$ ) and neuroticism ( $N=170,910$ ), performed by combining publicly available summary statistics from published studies with new genome-wide analyses of additional data. In these auxiliary analyses we identify two loci associated with depressive symptoms and eleven with neuroticism, including two inversion polymorphisms. In depression data from an independent sample ( $N=368,890$ ), both depressive symptoms associations replicate ( $p=0.004$  and  $p=0.015$ ).

In our two joint analyses, we exploit the high genetic correlation between subjective well-being, depressive symptoms, and neuroticism (i) to evaluate the credibility of the 16 genome-wide significant associations across the three phenotypes, and (ii) to identify novel associations (beyond those identified by the GWAS). For (i), we investigate whether our three subjective well-being-associated SNPs “quasi-replicate” by testing them for association with depressive symptoms and neuroticism. We similarly examine the quasi-replication record of the depressive symptoms and neuroticism loci by testing them for association with subjective well-being. We find that the quasi-replication record closely matches what would be expected given our statistical power if none of the genome-wide significant associations were chance findings. These results strengthen the credibility of (most of) the original associations. For (ii), we use a “proxy phenotype” approach<sup>6</sup>: we treat the set of loci associated with subjective well-being at  $p < 10^{-4}$  as candidates, and we test them for association with depressive symptoms and neuroticism. At the Bonferroni-adjusted 0.05 significance threshold, we identify two loci associated with both depressive symptoms and neuroticism and another two associated with neuroticism.

In designing our study, we faced a tradeoff between analyzing a smaller sample with a homogeneous phenotype measure versus attaining a larger sample by jointly analyzing data from multiple cohorts with heterogeneous measures. For example, in our analysis of subjective well-being, we included measures of both life satisfaction and positive affect, even though these constructs are conceptually distinct<sup>7</sup>. In Supplementary Note and Supplementary Figure 1, we present a theoretical framework for evaluating the costs and benefits of pooling heterogeneous measures. In our context, given the high genetic correlation across measures, the framework predicts that pooling increases statistical power to detect variants. This prediction is supported by our results.

## RESULTS

### GWAS of subjective well-being

Following a pre-specified analysis plan, we conducted a sample-size-weighted meta-analysis ( $N=298,420$ ) of cohort-level GWAS summary statistics. The phenotype measure was life

satisfaction, positive affect, or (in some cohorts) a measure combining life satisfaction and positive affect. We confirmed previous findings<sup>9</sup> of high pairwise genetic correlation between life satisfaction and positive affect using bivariate LD Score regression<sup>10</sup> ( $\hat{\rho} = 0.981$  (SE = 0.065); Supplementary Table 1). Details on the 59 participating cohorts, their phenotype measures, genotyping, quality-control filters, and association models are provided in Online Methods, Supplementary Note, and Supplementary Tables 2–6.

As expected under polygenicity<sup>11</sup>, we observe inflation of the median test statistic ( $\lambda_{GC} = 1.206$ ). The estimated intercept from LD Score regression (1.012) suggests that nearly all of the inflation is due to polygenic signal rather than bias. We also performed family-based analyses that similarly suggest minimal confounding due to population stratification (Online Methods). Using a clumping procedure (Supplementary Note), we identified three approximately independent SNPs reaching genome-wide significance (“lead SNPs”). These three lead SNPs are indicated in the Manhattan plot (Figure 1a) and listed in Table 1. The SNPs have estimated effects in the range 0.015 to 0.018 standard deviations (SDs) per allele (each  $R^2 \approx 0.01\%$ ).

We also conducted separate meta-analyses of the components of our subjective well-being measure, life satisfaction ( $N = 166,205$ ) and positive affect ( $N = 180,281$ ) (Online Methods). Consistent with our theoretical conclusion that pooling heterogeneous measures increased power in our context, the life satisfaction and positive affect analyses yielded fewer signals across a range of  $p$ -value thresholds than our meta-analysis of subjective well-being (Supplementary Table 7).

### GWAS of depressive symptoms and neuroticism

We conducted auxiliary GWAS of depressive symptoms and neuroticism (see Online Methods, Supplementary Note, and Supplementary Tables 8–12 for details on cohorts, phenotype measures, genotyping, association models, and quality-control filters). For depressive symptoms ( $N = 180,866$ ), we meta-analyzed publicly available results from a study performed by the Psychiatric Genomics Consortium (PGC)<sup>12</sup> together with new results from analyses of the initial release of the UK Biobank data (UKB)<sup>13</sup> and the Resource for Genetic Epidemiology Research on Aging (GERA) Cohort<sup>14</sup>. In UKB ( $N = 105,739$ ), we constructed a continuous phenotype measure by combining responses to two questions, which ask about the frequency in the past two weeks with which the respondent experienced feelings of unenthusiasm/disinterest and depression/hopelessness. The other cohorts had ascertained case-control data on major depressive disorder (GERA:  $N_{cases} = 7,231$ ,  $N_{controls} = 49,316$ ; PGC:  $N_{cases} = 9,240$ ,  $N_{controls} = 9,519$ ).

For neuroticism ( $N = 170,910$ ), we pooled summary statistics from a published study by the Genetics of Personality Consortium (GPC)<sup>4</sup> with results from a new analysis of UKB data. The GPC ( $N = 63,661$ ) harmonized different neuroticism batteries. In UKB ( $N = 107,245$ ), our measure was the respondent’s score on a 12-item version of the Eysenck Personality Inventory Neuroticism scale<sup>15</sup>.

In both the depressive symptoms and neuroticism GWAS, the heterogeneous phenotypic measures are highly genetically correlated (Supplementary Table 1). As in our subjective

well-being analyses, there is substantial inflation of the median test statistics ( $\lambda_{GC} = 1.168$  for depressive symptoms,  $\lambda_{GC} = 1.317$  for neuroticism), but the estimated LD Score intercepts (1.008 and 0.998, respectively) suggest that bias accounts for little or none of the inflation.

For depressive symptoms, we identified two lead SNPs, indicated in the Manhattan plot (Fig. 1b). For neuroticism, our meta-analysis yielded 16 loci that are independent according to our locus definition (Fig. 1c). However, 6 of these reside within a well-known inversion polymorphism<sup>16</sup> on chromosome 8. We established that all genome-wide significant signals in the inversion region are attributable to the inversion, and we confirmed that the inversion is associated with neuroticism in both of our neuroticism datasets, the GPC and the UKB (Online Methods and Supplementary Note). In our list of lead SNPs (Table 1), we only retain the most strongly associated SNP from these 6 loci to tag the chromosome 8 inversion.

Another lead SNP associated with neuroticism, rs193236081, is located within a well-known inversion polymorphism on chromosome 17. We established that this association is attributable to the inversion polymorphism (Online Methods and Supplementary Note). Because this inversion yields only one significant locus and is genetically complex<sup>17</sup>, we hereafter simply use its lead SNP as its proxy. Our neuroticism GWAS therefore identified 11 lead SNPs, two of which tag inversion polymorphisms. A concurrent neuroticism GWAS using a subset of our sample reports similar findings<sup>18</sup>.

As shown in Table 1, the estimated effects of all lead SNPs associated with depressive symptoms and neuroticism are in the range 0.020 to 0.031 SDs per allele ( $R^2$  0.02% to 0.04%). In the UKB cohort we estimated the effect of an additional allele of the chromosome 8 inversion polymorphism itself on neuroticism to be 0.035 SDs (Supplementary Table 13). The inversion explains 0.06% of the variance in neuroticism (roughly the same as the total variance explained jointly by the 6 SNPs in the inversion region).

### Genetic overlap across subjective well-being, depressive symptoms, and neuroticism

Figure 2a shows that the three pairwise genetic correlations between our phenotypes, estimated using bivariate LD Score regression<sup>10</sup>, are substantial:  $-0.81$  ( $SE = 0.046$ ) between subjective well-being and depressive symptoms,  $-0.75$  ( $SE = 0.034$ ) between subjective well-being and neuroticism, and  $0.75$  ( $SE = 0.027$ ) between depressive symptoms and neuroticism. Using height as a negative control, we also examined pairwise genetic correlations between each of our phenotypes and height and, as expected, found all three to be modest, e.g., 0.07 with subjective well-being (Supplementary Table 1). The high genetic correlations between subjective well-being, depressive symptoms, and neuroticism may suggest that the genetic influences on these phenotypes are predominantly related to processes common across the phenotypes, such as mood, rather than being phenotype-specific.

## Quasi-replication and Bayesian credibility analyses

We assessed the credibility of our findings using a standard Bayesian framework<sup>19,20</sup> in which a positive fraction of SNPs have null effects and a positive fraction have non-null effects (Online Methods). For each phenotype, the non-null effect sizes are assumed to be drawn from a normal distribution whose variance is estimated from the GWAS summary statistics. As a first analysis, for each lead SNP's association with its phenotype, we calculated the posterior probability of null association after having observed the GWAS results. We found that, for any assumption about the fraction of non-null SNPs in the range 1% to 99%, the probability of true association always exceeds 95% for all 16 loci (and always exceeds 98% for 14 of them).

To further probe the credibility of the findings, we performed “quasi-replication” exercises (Online Methods) in which we tested the subjective well-being lead-SNPs for association with depressive symptoms and neuroticism. We similarly tested the depressive symptoms lead-SNPs and the neuroticism lead-SNPs for association with subjective well-being. Below, we refer to the phenotype for which the lead SNP was identified as the first-stage phenotype and the phenotype used for the quasi-replication as the second-stage phenotype. To avoid sample overlap, for each quasi-replication analysis we omitted any cohorts that contributed to the GWAS of the first-stage phenotype.

Results of the quasi-replication of the three subjective well-being lead-SNPs are shown in Figure 3a. For ease of interpretation, the reference allele for each association in the figure is chosen such that the predicted sign of the second-stage estimate is positive. We find that two out of the three subjective well-being lead-SNPs are significantly associated with depressive symptoms ( $p = 0.004$  and  $p = 0.001$ ) in the predicted direction. For neuroticism, where the second-stage sample size ( $N = 68,201$ ) is about half as large, the subjective well-being-increasing allele has the predicted sign for all three SNPs, but none reach significance.

Figures 3b and 3c show the results for the depressive symptoms and neuroticism lead-SNPs, respectively. In each panel, the blue crosses depict results from the quasi-replications where subjective well-being is the second-stage phenotype. We find that the two depressive symptoms lead-SNPs have the predicted sign for subjective well-being, and one is nominally significant ( $p = 0.04$ ). Finally, of the eleven neuroticism lead-SNPs, nine have the predicted sign for subjective well-being. Four of the eleven are nominally significantly associated with subjective well-being, all with the predicted sign. One of the four is the SNP tagging the inversion on chromosome 8<sup>16</sup>. That SNP's association with neuroticism (and likely with subjective well-being) is driven by its correlation with the inversion (Supplementary Fig. 2).

To evaluate what these quasi-replication results imply about the credibility of the 16 GWAS associations, we compared the observed quasi-replication record to the quasi-replication record expected given our statistical power. We calculated statistical power using our Bayesian framework, under the hypothesis that each lead SNP has a non-null effect on both the first- and second-stage phenotypes. Our calculations take into account both the imperfect genetic correlation between the first- and second-stage phenotypes and inflation of the first-stage estimates due to the well-known problem of winner's curse (Online Methods). Of the 19 quasi-replication tests, our calculations imply that 16.7 would be expected to yield the

anticipated sign and 6.9 would be significant at the 5% level. The observed numbers are 16 and 7. Our quasi-replication results are thus consistent with the hypothesis that none of the 16 genome-wide significant associations are chance findings, and in fact strengthen the credibility of our GWAS results (Supplementary Table 14).

### Lookup of depressive symptoms and neuroticism lead-SNPs

Investigators of an ongoing large-scale GWAS of major depressive disorder ( $N = 368,890$ ) in the 23andMe cohort shared association results for the loci identified in our depressive symptoms and neuroticism analyses (Online Methods and Supplementary Table 15)<sup>21</sup>. Because the depression sample overlaps with our subjective well-being sample, we did not request a lookup of the subjective well-being-associated SNPs.

In Figures 3b and 3c, the results are depicted as green crosses. For interpretational ease, we chose the reference allele so that positive coefficients imply that the estimated effect is in the predicted direction. All 13 associations have the predicted sign. Of the 11 neuroticism polymorphisms, four are significantly associated with depression at the 5% level. Both of the depressive symptoms lead-SNPs replicate ( $p = 0.004$  and  $p = 0.015$ ), with effect sizes (0.007 and  $-0.007$  SDs per allele), close to those predicted by our Bayesian framework (0.008 and  $-0.006$ ) (Supplementary Table 14 and Supplementary Table 15).

Panel A of Table 1 summarizes the results for the 16 lead SNPs identified across our separate GWA analyses of the three phenotypes. The right-most column summarizes the statistical significance of the quasi-replication and depression lookup analyses of each SNP.

### Proxy-phenotype analyses

To identify additional SNPs associated with depressive symptoms, we conducted a two-stage “proxy phenotype” analysis (Online Methods). In the first stage, we ran a new GWAS of subjective well-being to identify a set of candidate SNPs. Specifically, from each locus exhibiting suggestive evidence of association ( $p < 10^{-4}$ ) with subjective well-being, we retained the SNP with the lowest  $p$ -value as a candidate. In the second stage, we tested these candidates for association with depressive symptoms at the 5% significance threshold, Bonferroni-adjusted for the number of candidates. We used an analogous two-stage procedure to identify additional SNPs associated with neuroticism. The first-stage subjective well-being sample differs across the two proxy-phenotype analyses (and from the primary subjective well-being GWAS sample) because we assigned cohorts across the first and second stages so as to maximize statistical power for the overall procedure.

For depressive symptoms, there are 163 candidate SNPs. 115 of them (71%) have the predicted direction of effect on depressive symptoms, 20 are significantly associated at the 5% significance level (19 in the predicted direction), and two remain significant after Bonferroni adjustment. For neuroticism, there are 170 candidate SNPs. 129 of them (76%) have the predicted direction of effect, all 28 SNPs significant at the 5% level have the predicted sign, and four of these remain significant after Bonferroni adjustment (Supplementary Fig. 3 and Supplementary Tables 16 and 17). Two of the four are the SNPs identified in the proxy-phenotype analysis for depressive symptoms.

Table 1 lists the four SNPs in total identified by the proxy-phenotype analyses.

### Biological analyses

To shed some light on possible biological mechanisms underlying our findings, we conducted several analyses.

We began by using bivariate LD Score regression<sup>10</sup> to quantify the amount of genetic overlap between each of our three phenotypes and ten neuropsychiatric and physical health phenotypes. Figures 2b and c display the estimates for subjective well-being and the *negative* of the estimates for depressive symptoms and neuroticism (since subjective well-being is negatively genetically correlated with depressive symptoms and neuroticism). Subjective well-being, depressive symptoms, and neuroticism have strikingly similar patterns of pairwise genetic correlation with the other phenotypes.

Figure 2b shows the results for the five neuropsychiatric phenotypes we examined: Alzheimer's disease, anxiety disorders, autism spectrum disorder, bipolar disorder, and schizophrenia. For four of these phenotypes, genetic correlations with depression (but not neuroticism or subjective well-being) were reported in Bulik-Sullivan et al.<sup>10</sup>. For schizophrenia and bipolar disorder, our estimated correlations with depressive symptoms, 0.33 and 0.26, are substantially lower than Bulik-Sullivan et al.'s point estimates but contained within their 95% confidence intervals. By far the largest genetic correlations we estimate are with anxiety disorders:  $-0.73$  with subjective well-being,  $0.88$  with depressive symptoms, and  $0.86$  with neuroticism. Genetic correlations estimated from GWAS data have not been previously reported for anxiety disorders.

Figure 2c shows the results for five physical health phenotypes that are known or believed to be risk factors for various adverse health outcomes: body mass index (BMI), ever-smoker status, coronary artery disease, fasting glucose, and triglycerides. The estimated genetic correlations are all small in magnitude, consistent with earlier work, although the greater precision of our estimates allows us to reject null effects in most cases. The signs are generally consistent with those of the phenotypic correlations reported in earlier work between our phenotypes and outcomes such as obesity<sup>22</sup>, smoking<sup>23,24</sup>, and cardiovascular health<sup>25</sup>.

Next, to investigate whether our GWAS results are enriched in particular functional categories, we applied stratified LD Score regression<sup>26</sup> to our meta-analysis results. In our first analysis, we report estimates for all 53 functional categories included in the "baseline model"; the results for subjective well-being, depressive symptoms, and neuroticism are broadly similar (Supplementary Tables 18–20) and are in line with what has been found for other phenotypes<sup>26</sup>. In our second analysis, the categories are groupings of SNPs likely to regulate gene expression in cells of a specific tissue. The estimates for subjective well-being, depressive symptoms, and neuroticism are shown in Figure 4a, alongside height, which is again included as a benchmark<sup>27</sup> (see also Supplementary Table 21).

We found significant enrichment of CENTRAL NERVOUS SYSTEM for all three phenotypes and, perhaps more surprisingly, enrichment of ADRENAL/PANCREAS for

subjective well-being and depressive symptoms. The cause of the ADRENAL/PANCREAS enrichment is unclear, but we note that the adrenal glands produce several hormones, including cortisol, epinephrine, and norepinephrine, known to play important roles in the bodily regulation of mood and stress. It has been robustly found that blood serum levels of cortisol in patients afflicted by depression are elevated relative to controls<sup>28</sup>.

While the above analyses utilize the genome-wide data, we also conducted three analyses (Online Methods) restricted to the 16 GWAS and four proxy-phenotype SNPs in Table 1. In brief, we ascertained whether each SNP (or a variant in strong linkage disequilibrium (LD) with it) falls into any of the following three classes: (i) resides in a locus for which genome-wide significant associations with other phenotypes have been reported (Supplementary Table 22), (ii) is nonsynonymous (Supplementary Table 23), and (iii) is an eQTL in blood or in one of 14 other tissues (although the non-blood analyses are based on smaller samples) (Supplementary Table 24). Here we highlight a few particularly interesting results.

We found that five of the 20 SNPs are in loci in which genome-wide significant associations have previously been reported. Two of these five are schizophrenia loci. Interestingly, one of them harbors the gene *DRD2*, which encodes the D<sub>2</sub> subtype of the dopamine receptor, a target for antipsychotic drugs<sup>29</sup> that is also known to play a key role in neural reward pathways<sup>30</sup>. Motivated by these findings, as well as by the modest genetic correlations with schizophrenia reported in Figure 2b, we examined whether the SNPs identified in a recent study of schizophrenia<sup>31</sup> are enriched for association with neuroticism in our non-overlapping UKB sample ( $N = 107,245$ ). We conducted several tests and found strong evidence of such enrichment (Supplementary Note). For example, we found that the  $p$ -values of the schizophrenia SNPs tend to be much lower than the  $p$ -values of a randomly selected set of SNPs matched on allele frequency ( $p = 6.50 \times 10^{-71}$ ).

Perhaps the most notable pattern that emerges from our biological analyses is that the inversions on chromosomes 8 and 17 are implicated consistently across all analyses. The inversion-tagging SNP on chromosome 8 is in LD with SNPs that have previously been found to be associated with BMI<sup>32</sup> and triglycerides<sup>33</sup> (Supplementary Table 22). We also conducted eQTL analyses in blood for the inversion itself and found that it is a significant *cis*-eQTL for 7 genes (Supplementary Table 24). As shown in Figure 4b, all 7 genes are positioned in close proximity to the inversion breakpoints, suggesting that the molecular mechanism underlying the inversion's effect on neuroticism could involve the relocation of regulatory sequences. Two of the genes (*MSRA*, *MTMR9*) are known to be highly expressed in tissues and cell types that belong to the nervous system, and two (*BLK*, *MFHAS1*) in the immune system. In the tissue-specific analyses, we found that the SNP tagging the inversion is a significant eQTL for two genes, *AF131215.9* (in tibial nerve and thyroid tissue analyses) and *NEIL2* (tibial nerve tissue), both of which are also located near the inversion breakpoint.

The SNP tagging the chromosome 17 inversion is a significant *cis*-eQTL for five genes in blood and is an eQTL in all 14 other tissues (Supplementary Table 24). It alone accounts for 151 out of the 169 significant associations identified in the 14 tissue-specific analyses. Additionally, the SNP is in near-perfect LD ( $R^2 > 0.97$ ) with 11 missense variants (Supplementary Table 23) in three different genes, one of which is *MAPT*. *MAPT*, which is

also implicated in both the blood and the other tissue-specific analyses, encodes a protein important in the stabilization of microtubules in neurons. Associations have been previously reported between SNPs in *MAPT* (all of which are in strong LD with our inversion-tagging SNP) and neurodegenerative disorders, including Parkinson's disease<sup>34</sup> and progressive supranuclear palsy<sup>35</sup>, a rare disease whose symptoms include depression and apathy.

## DISCUSSION

The discovery of genetic loci associated with subjective well-being, depression, and neuroticism has proven elusive. Our study identified several credible associations for two main reasons. First, our analyses had greater statistical power than prior studies because ours were conducted in larger samples. Our GWAS findings—three loci associated with subjective well-being, two with depressive symptoms, and eleven with neuroticism—support the view that GWAS can successfully identify genetic associations with highly polygenic phenotypes in sufficiently large samples<sup>5,36</sup>. A striking finding is that two of our identified associations are with inversion polymorphisms.

Second, our proxy-phenotype analyses further boosted power by exploiting the strong genetic overlap between our three phenotypes. These analyses identified two additional loci associated with neuroticism and two with both depressive symptoms and neuroticism. Through our quasi-replication tests, we also demonstrated how studying genetically overlapping phenotypes in concert can provide evidence on the credibility of GWAS findings. Our direct replication of the two genome-wide significant associations with depressive symptoms in an independent depression sample provides further confirmation of those findings (Fig. 2b and Supplementary Table 15).

We were able to assemble much larger samples than prior work in part because we combined data across heterogeneous phenotype measures. Our results reinforce the conclusions from our theoretical analysis that doing so increased our statistical power, but our strategy also has drawbacks. One is that mixing different measures may make any discovered associations more difficult to interpret. Research studying higher quality measures of the various facets of subjective well-being, depressive symptoms, and neuroticism is a critical next step. Our results can help facilitate such work because if the variants we identify are used as candidates, studies conducted in the smaller samples in which more fine-grained phenotype measures are available can be well powered.

Another limitation of mixing different measures is that doing so may reduce the heritability of the resulting phenotype, if the measures are influenced by different genetic factors. Indeed, our estimates of SNP-based heritability<sup>10</sup> for our three phenotypes are quite low: 0.040 (SE = 0.002) for subjective well-being, 0.047 (SE = 0.004) for depressive symptoms, and 0.091 (SE = 0.007) for neuroticism. We correspondingly find that polygenic scores constructed from all measured SNPs explain a low fraction of variance in independent samples: ~0.9% for subjective well-being, ~0.5% for depressive symptoms, and ~0.7% for neuroticism (Online Methods). The low heritabilities imply that even when polygenic scores can be estimated using much larger samples than ours, they are unlikely to attain enough predictive power to be clinically useful.

According to our Bayesian calculations, the true explanatory power (corrected for winner's curse) of the SNP with the largest posterior  $R^2$  is 0.003% for subjective well-being, 0.002% for depressive symptoms, and 0.011% for neuroticism (Supplementary Table 14). These effect sizes imply that in order to account for even a moderate share of the heritability, hundreds or (more likely) thousands of variants will be required. They also imply that our study's power to detect variants of these effect sizes was not high—for example, our statistical power to detect the lead SNP with largest posterior  $R^2$  was only ~13%—which in turn means it is likely that there exist many variants with effect sizes comparable to our identified SNPs that evaded detection. These estimates suggest that many more loci will be found in studies with sample sizes realistically attainable in the near future. Consistent with this projection, when we meta-analyze the 54 SNPs reaching  $p < 10^{-5}$  in our analyses of depressive symptoms together with the 23andMe replication sample for depression, the number of genome-wide significant associations rises from 2 to 5 (Supplementary Table 15).

## ONLINE METHODS

This article is accompanied by a Supplementary Note with further details.

### GWAS of subjective well-being

Genome-wide association analyses were performed at the cohort level according to a pre-specified analysis plan. Genotyping was performed using a range of common, commercially available genotyping arrays. The analysis plan instructed cohorts to upload results imputed using the HapMap2 CEU (r22.b36) reference sample<sup>37</sup>. We meta-analyzed summary association statistics from 59 contributing cohorts with a combined sample size of 298,420 individuals. Before meta-analysis, a uniform set of quality-control (QC) procedures were applied to the cohort-level summary statistics, including but not limited to the EasyQC<sup>38</sup> protocol. All analyses were restricted to European-ancestry individuals.

We performed a sample-size-weighted meta-analysis of the cohort-level summary statistics. To adjust standard errors for non-independence, we inflated them using the square root of the estimated intercept from a LD Score regression<sup>10</sup>. We also performed secondary, separate meta-analyses of positive affect ( $N = 180,281$ ) and life satisfaction ( $N = 166,205$ ) and a post hoc genome-wide analysis of subjective well-being in cohorts with 1000G-imputed data ( $N = 229,883$ ); see Supplementary Figures 4–6.

Detailed cohort descriptions, information about cohort-level genotyping and imputation procedures, cohort-level measures, and quality-control filters are shown in Supplementary Tables 2–6. Supplementary Table 7 reports association results from the following four meta-analyses: the primary subjective well-being analysis, the life satisfaction analysis, the positive affect analysis, and the post hoc subjective well-being analysis. For each phenotype, we provide association results for the set of approximately independent SNPs that attained a  $p$ -value smaller than  $10^{-5}$ . We identify these SNP using the same clumping algorithm as for the lead SNPs, but with the  $p$ -value threshold set at  $10^{-5}$  instead of genome-wide significance.

## GWAS of depressive symptoms and neuroticism

Our auxiliary genome-wide association studies of DS and neuroticism were conducted in 1000G-imputed data, combining new genome-wide association analyses with publicly available summary statistics from previously published studies. We applied a similar QC protocol to that used in our primary subjective well-being analysis. In the DS meta-analysis ( $N = 180,866$ ), we weighted the UKB analysis by sample size and the two case-control studies by effective sample size. In the neuroticism meta-analysis, we performed a sample-size-weighted fixed-effects meta-analysis of the UKB data and the publicly available summary statistics from a previous GWAS of neuroticism.

Detailed cohort descriptions, information about cohort-level genotyping and imputation procedures and quality-control filters are provided in Supplementary Tables 8–12. See Supplementary Figure 7 for quantile-quantile plots of the neuroticism and DS meta-analysis results. Association results for the set of approximately independent set of SNPs that attained a  $p$ -value smaller than  $10^{-5}$  are supplied in Supplementary Table 25.

## Population stratification

To quantify the fraction of the observed inflation of the mean test statistic that is due to bias, we used LD Score regression<sup>10</sup>. The estimated LD Score regression intercepts were all close to 1, suggesting no appreciable inflation of the test statistics attributable to population stratification in any of our subjective well-being, depressive symptoms, or neuroticism meta-analyses (Supplementary Fig. 8). For all three phenotypes, our estimates suggest that less than 2% of the observed inflation of the mean test statistic was accounted for by bias.

In our primary GWAS of subjective well-being, we also used two family-based analyses to test for and quantify stratification biases. These analyses used within-family (WF) estimates, the coefficients from regressing the difference in phenotype across siblings on the difference in siblings' genotype (and controls). These WF estimates are not biased by population stratification because siblings share their ancestry entirely, and therefore differences in siblings' genotypes cannot be due to the siblings being from different population groups. We meta-analyzed association statistics from WF analyses conducted in four cohorts.

In the first analysis, we estimated the fraction of SNPs for which the signs of the WF estimates were concordant with the signs of the estimates obtained from a GWAS identical to our primary subjective well-being GWAS except with the four family cohorts excluded. For the 112,884 approximately independent SNPs considered, we found a sign concordance of 50.83%, which is significantly greater than 50% ( $p = 1.04 \times 10^{-8}$ ). Under the null hypothesis of no population stratification, the observed sign concordance matches the expected rate after winner's curse adjustment nearly perfectly, 50.83% (Supplementary Fig 9).

The second analysis utilized the WF regression coefficient estimates (i.e., not only their signs) to estimate the amount of stratification bias. For each SNP  $j$ , let  $\beta_j$  denote the GWAS estimate, and let  $\beta_{WFj}$  denote the WF estimate. Under the assumption that the causal effect of each SNP is the same within families as in the population, we can decompose the estimates as:

$$\begin{aligned}\hat{\beta}_j &= \beta_j + s_j + U_j \\ \hat{\beta}_{WF,j} &= \beta_j + V_j,\end{aligned}$$

where  $\beta_j$  is the true underlying GWAS parameter for SNP  $j$ ,  $s_j$  is the bias due to stratification (defined to be orthogonal to  $\beta_j$  and  $U_j$ ), and  $U_j$  and  $V_j$  are the sampling variances of the estimates with  $E(U_j) = E(V_j) = 0$ . Whenever  $s_j \neq 0$ , the GWAS estimate of  $\beta_j$  is biased away from the population parameter  $\beta_j$ . The proportion of variance in the GWAS coefficients accounted for by true genetic signals can be written as:

$$\frac{\text{Var}(\beta_j)}{\text{Var}(\beta_j) + \text{Var}(s_j)}.$$

In Supplementary Note, we show that with estimates  $\hat{\beta}_j$  and  $\hat{\beta}_{WF,j}$  (and their standard errors) from independent samples, it is possible to consistently estimate the above ratio. The 95% confidence interval for the ratio implies that between 72% and 100% of the signal in the GWAS estimates is a result of true genetic effects on subjective well-being rather than stratification.

### Analyses of inversion polymorphisms

Two genome-wide significant SNPs for the neuroticism analysis are located within well-known inversion polymorphisms, on chromosomes 8 and 17. Using the genotypic data available for UKB participants, we called the inversion genotypes for UKB participants using a PCA-mixture method. For both inversions, the method clearly distinguishes 3 clusters of genotypes, corresponding to inversion genotypes (Supplementary Fig. 10). We validated the PCA-mixture procedure using existing methods designed to call inversion genotypes<sup>39</sup> (Supplementary Table 26).

For both inversions, we established that the inversion-tagging SNPs were always located in close proximity of the inversion region (Fig. 3b and Supplementary Figs. 10–11). Supplementary Tables 27–28 list the twenty variants that most strongly correlate with the PCs that capture the inversion polymorphisms on chromosome 8 and 17, respectively. In additional analyses, we confirmed that the inversion is associated with neuroticism and subjective well-being in independent cohorts (Supplementary Tables 29–30 and Supplementary Fig. 12–13).

### Proxy-phenotype analyses

In these analyses, we used a two-stage approach that has been successfully applied in other contexts<sup>6</sup>. In the first stage, we conducted a meta-analysis of our first-stage “proxy phenotype” and used our clumping procedure to identify the set of approximately independent SNPs at the  $p$ -value threshold of  $10^{-4}$ . In the second stage, we tested SNPs identified in stage 1 (or high-LD proxies for them) for association with a second-stage phenotype in an independent (non-overlapping) sample. In our analyses, we used our primary phenotype of subjective well-being as the proxy-phenotype. We conducted one

analysis with depressive symptoms as the second-stage phenotype, and one analysis with neuroticism as the second-stage phenotype. In the analyses, we omit cohorts from the first-stage or second-stage as needed to ensure that the samples in the two stages are non-overlapping. Supplementary Table 31 lists the cohort restrictions imposed. These cohort restrictions, as well as the  $p$ -value threshold of  $10^{-4}$ , were chosen before the data were analyzed on the basis of statistical power calculations.

To test for cross-phenotype enrichment, we used a non-parametric procedure that tests whether the lead SNPs are more strongly associated with the second-stage phenotype than randomly chosen sets of SNPs with a similar distribution of allele frequencies (Supplementary Note).

To test the individual lead SNPs for experiment-wide significance, we examined whether any of the lead SNPs (or their high-LD proxies) are significantly associated with the second-stage phenotype at the Bonferroni-adjusted significance level of  $0.05/Y$ .

### Genetic correlations

We used bivariate LD Score regression<sup>10</sup> to quantify the amount of genetic heterogeneity among the phenotypic measures pooled in each of our three separate meta-analyses. For subjective well-being, we estimated a pairwise correlation of 0.981 (SE = 0.065) between life satisfaction and positive affect, 0.897 (SE = 0.017) between “wellbeing” (our measure that combines life satisfaction and positive affect) and life satisfaction, and 1.031 (SE = 0.019) between positive affect and wellbeing. For depressive symptoms, we estimated a genetic correlation of 0.588 (SE = 0.242) between GERA and PGC, 0.972 (SE = 0.216) between GERA and UKB, and 0.797 (SE = 0.108) between UKB and PGC. Finally, we estimated a genetic correlation of 1.11 (SE = 0.14) between the measures of neuroticism in the UKB analyses and the summary statistics from a previously published meta-analysis<sup>4</sup>.

### Bayesian credibility analyses

To evaluate the credibility of our findings, we use a standard Bayesian framework<sup>19</sup> in which our prior distribution for any SNP's effect is:

$$\beta \sim \begin{cases} N(0, \tau_j^2) & \text{with probability } \pi \\ 0 & \text{otherwise.} \end{cases} .$$

Here,  $\pi$  is the fraction of non-null SNPs, and  $\tau_j^2$  is the variance of the non-null SNPs for trait  $j \in \{\text{subjective well-being, depressive symptoms, neuroticism}\}$ . In this framework, credibility is defined as the probability that a given SNP is non-null.

We begin with univariate analyses of the GWAS results that do not incorporate the additional information from the quasi-replication analyses of the 16 lead SNPs reported in Table 1. We use the three subjective well-being-associated SNPs to illustrate our approach, but we use analogous procedures when analyzing depressive symptoms and neuroticism. We calculate credibility for each value  $\pi \in 0.01, 0.02, \dots, 0.99$ . For each assumed value of  $\pi$ , we estimate

$\tau_{SWB}^2$  by maximum likelihood (Supplementary Note). For each SNP, we use Bayes' rule to obtain a posterior estimate of credibility for each of the assumed values of  $\pi$ . Supplementary Figure 14 shows that for all considered values of  $\pi$  and all three SNPs, the posterior probability that the SNP is null is below 1%. Similar analyses of the depressive symptoms and neuroticism SNPs show that the posterior probability never exceeds 5%.

In our joint analyses, we consider two phenotypes with genetic correlation  $r_g$ . We make the simplifying assumption that the set of null SNPs is the same for both phenotypes. The joint distribution of a SNP's effect on the two phenotypes is then given by

$$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \sim \begin{cases} N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_1^2 & \tau_1 \tau_2 r_g \\ \tau_1 \tau_2 r_g & \tau_2^2 \end{bmatrix} \right) & \text{with probability } \pi \\ \begin{bmatrix} 0 \\ 0 \end{bmatrix} & \text{otherwise.} \end{cases}$$

With coefficient estimates,  $\hat{\beta}_1$  and  $\hat{\beta}_2$ , obtained from non-overlapping samples, the variance-covariance matrix of the estimation error will be diagonal. We denote the diagonal entries of this matrix, which represent the variances of the estimation error in the two samples, by  $\sigma_1^2$  and  $\sigma_2^2$ . This gives us the joint prior distribution

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} \sim \begin{cases} N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_1^2 & \tau_1 \tau_2 r_g \\ \tau_1 \tau_2 r_g & \tau_2^2 \end{bmatrix} + \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \right) & \text{with probability } \pi \\ N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \right) & \text{otherwise.} \end{cases}$$

To select parameter values for the prior, we use the estimates of  $r_g$  reported in Supplementary Table 1, and we estimate the parameters  $\pi$ ,  $\tau_1^2$ , and  $\tau_2^2$  from GWAS summary statistics using a maximum likelihood procedure. For this procedure, we make the standard assumption<sup>10,40</sup> that the variance of a SNP's effect size is inversely proportional to the variance of its genotype,  $2 \times \text{MAF} \times (1 - \text{MAF})$ .

The credibility estimates follow from applying Bayes' Rule to calculate either the probability that the SNP is non-null (an event denoted  $C$ ) given only the first-stage estimate,  $P(C|\hat{\beta}_1)$ , or the probability that the SNP is non-null conditional on the results of both the first-stage GWAS and the quasi-replication analysis,  $P(C|\hat{\beta}_1, \hat{\beta}_2)$ . Credibility estimates for our lead SNPs are in Supplementary Table 14.

To calculate the expected record of a replication or quasi-replication study, we assume that the SNP is non-null for both phenotypes. (This is analogous to a standard power calculation for a single phenotype, in which the SNP is assumed to be non-null.) Under this assumption,  $\hat{\beta}_1$  and  $\hat{\beta}_2$  are jointly normally distributed, implying that the conditional distribution of  $\hat{\beta}_2$  given  $\hat{\beta}_1$  is

$$(\hat{\beta}_2 | \hat{\beta}_1, C) \sim N \left[ \frac{\tau_1 \tau_2 r_g}{\tau_1^2 + \sigma_1^2} \hat{\beta}_1, \frac{(\tau_1^2 + \sigma_1^2)(\tau_2^2 + \sigma_2^2) - \tau_1^2 \tau_2^2 r_g^2}{\tau_1^2 + \sigma_1^2} \right].$$

Using this equation, we can calculate the probability that the GWAS estimates will have concordant signs across the two phenotypes, or that the GWAS estimate of the second-stage phenotype will reach some level of significance. These probabilities can be summed over the set of lead SNPs to generate the expected number of SNPs meeting the criterion.

To obtain effect-size estimates for a SNP that are adjusted for the winner's curse (Supplementary Table 32), we use the mean of the posterior distribution of the SNP's effect, conditional on the quasi-replication result and the SNP being non-null. We derive the posterior distribution and expected  $R^2$  in the Supplementary Note.

### Lookup of depressive symptoms and neuroticism-associated SNPs in an independent depression study

We partnered with the investigators of an ongoing large-scale GWAS of major depressive symptoms ( $N = 368,890$ ) to follow up on the associations identified in the depressive symptoms and neuroticism analyses. The participants of the study were all European-ancestry customers of 23andMe, a personal genomics company, who responded to online survey questions about mental health. We did not request results for the SNPs identified in the subjective well-being or proxy-phenotype analyses, since these were both conducted in samples that overlap with 23andMe's depression sample. For details on association models, quality-control filters, and the ascertainment of depression status, we refer to the companion study<sup>21</sup>. The  $p$ -values we report are based on standard errors that have been inflated by the square by the intercept from an LD score regression<sup>10</sup>.

### Polygenic prediction

To evaluate the predictive power of a polygenic score derived from the subjective well-being meta-analysis results, we used two independent hold-out cohorts: the Health and Retirement Study (HRS<sup>41</sup>) and the Netherlands Twin Register (NTR<sup>42,43</sup>). To generate the weights for the polygenic score, we performed meta-analyses of the pooled subjective well-being phenotype excluding each of the holdout cohorts, applying a minimum-sample-size filter of 100,000 individuals (Supplementary Note). The results from these analyses are reported in Supplementary Table 33 and depicted in Supplementary Figure 15.

### Biological annotation

For the biological annotation of the 20 SNPs in Table 1, we generated a list of LD partners for each of the original SNPs. A SNP was considered an LD partner for the original SNP if (i) its pairwise LD with the original SNP exceeded  $R^2 = 0.6$  and (ii) it was located within 250kb of the original SNP. We also generated a list of genes residing within loci tagged by our lead SNPs (Supplementary Table 34).

We used the NHGRI GWAS catalog<sup>44</sup> to determine which of our 20 SNPs (and their LD partners) were in LD with SNPs for which genome-wide significant associations have been

previously reported. Since the GWAS catalog does not always include the most recent GWAS results available, we included additional recent GWAS studies. We used the tool HaploReg<sup>45</sup> to identify nonsynonymous variants in LD with any of the 20 SNPs or their LD partners.

We examined whether the 20 polymorphisms in Table 1 were associated with gene expression levels (Supplementary Table 24 and Supplementary Note). The *cis*-eQTL associations were performed in 4,896 peripheral-blood gene expression and genome-wide SNP samples from two Dutch cohorts measured on the Affymetrix U219 platform<sup>42,43,46</sup>. We also performed eQTL lookups of our 20 SNPs in the Genotype-Tissue Expression Portal<sup>47,48</sup>. We restricted the search to the following trait-relevant tissues: hippocampus, hypothalamus, anterior cingulate cortex (BA24), putamen (basal ganglia), frontal cortex (BA9), nucleus accumbens (basal ganglia), caudate (basal ganglia), cortex, cerebellar hemisphere, cerebellum, tibial nerve, thyroid, adrenal gland, and pituitary.

Finally, using a gene co-expression database<sup>49</sup>, we explored the predicted functions of genes co-locating with the 20 SNPs in Table 1 (Supplementary Table 35).

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Authors

Aysu Okbay<sup>1,2,3,\*</sup>, Bart M.L. Baselmans<sup>4,5,\*</sup>, Jan-Emmanuel De Neve<sup>6,\*</sup>, Patrick Turley<sup>7,\*</sup>, Michel G. Nivard<sup>4,\*</sup>, Mark Alan Fontana<sup>8,\*</sup>, S. Fleur W. Meddens<sup>9,3,10,\*</sup>, Richard Karlsson Linnér<sup>9,3,10,\*</sup>, Cornelius A. Rietveld<sup>1,2,3,\*</sup>, Jaime Derringer<sup>11</sup>, Jacob Gratten<sup>12</sup>, James J. Lee<sup>13</sup>, Jimmy Z. Liu<sup>14</sup>, Ronald de Vlaming<sup>1,2,3</sup>, Tarunveer S. Ahluwalia<sup>15,16,17</sup>, Jadwiga Buchwald<sup>18</sup>, Alana Cavadino<sup>19,20</sup>, Alexis C. Frazier-Wood<sup>21</sup>, Nicholas A. Furlotte<sup>22</sup>, Victoria Garfield<sup>23</sup>, Marie Henrike Geisel<sup>24</sup>, Juan R. Gonzalez<sup>25,26,27</sup>, Saskia Haitjema<sup>28</sup>, Robert Karlsson<sup>29</sup>, Sander W. van der Laan<sup>28</sup>, Karl-Heinz Ladwig<sup>30</sup>, Jari Lahti<sup>31,32,33</sup>, Sven J. van der Lee<sup>2</sup>, Penelope A. Lind<sup>34</sup>, Tian Liu<sup>35,36</sup>, Lindsay Matteson<sup>13</sup>, Evelin Mihailov<sup>37</sup>, Michael B. Miller<sup>13</sup>, Camelia C. Minica<sup>4</sup>, Ilja M. Nolte<sup>38</sup>, Dennis Mook-Kanamori<sup>39,40,41</sup>, Peter J. van der Most<sup>38</sup>, Christopher Oldmeadow<sup>42,43</sup>, Yong Qian<sup>44</sup>, Olli Raitakari<sup>45,46</sup>, Rajesh Rawal<sup>47</sup>, Anu Realo<sup>48,49</sup>, Rico Rueedi<sup>50,51</sup>, Børge Schmidt<sup>24</sup>, Albert V. Smith<sup>52,53</sup>, Evie Stergiakouli<sup>54</sup>, Toshiko Tanaka<sup>55</sup>, Kent Taylor<sup>56</sup>, Juho Wedenoja<sup>18</sup>, Juergen Wellmann<sup>57</sup>, Harm-Jan Westra<sup>58,59</sup>, Sara M. Willems<sup>2</sup>, Wei Zhao<sup>60</sup>, LifeLines Cohort Study<sup>61</sup>, Najaf Amin<sup>2</sup>, Andrew Bakshi<sup>12</sup>, Patricia A. Boyle<sup>62</sup>, Samantha Cherney<sup>63</sup>, Simon R. Cox<sup>64,65</sup>, Gail Davies<sup>64,65</sup>, Oliver S.P. Davis<sup>54</sup>, Jun Ding<sup>44</sup>, Nese Direk<sup>2</sup>, Peter Eibich<sup>66,67</sup>, Rebecca T. Emeny<sup>30,68</sup>, Ghazaleh Fatemifar<sup>69</sup>, Jessica D. Faul<sup>70</sup>, Luigi Ferrucci<sup>71</sup>, Andreas Forstner<sup>72,73</sup>, Christian Gieger<sup>47</sup>, Richa Gupta<sup>18</sup>, Tamara B. Harris<sup>74</sup>, Juliette M. Harris<sup>75</sup>, Elizabeth G. Holliday<sup>42,43</sup>, Jouke-Jan Hottenga<sup>4,5</sup>, Philip L. De Jager<sup>76,77,78</sup>, Marika A. Kaakinen<sup>79,82</sup>, Eero Kajantie<sup>80,81</sup>, Ville Karhunen<sup>82</sup>, Ivana Kolcic<sup>83</sup>, Meena Kumari<sup>84</sup>, Lenore J. Launer<sup>85</sup>, Lude Franke<sup>86</sup>, Ruifang Li-Gao<sup>39</sup>, Marisa Koini<sup>87</sup>, Anu Loukola<sup>18</sup>, Pedro

Marques-Vidal<sup>88</sup>, Grant W. Montgomery<sup>89</sup>, Miriam A. Mosing<sup>90</sup>, Lavinia Paternoster<sup>54</sup>, Alison Pattie<sup>65</sup>, Katja E. Petrovic<sup>87</sup>, Laura Pulkki-Råback<sup>31,33</sup>, Lydia Quaye<sup>75</sup>, Katri Räikkönen<sup>31</sup>, Igor Rudan<sup>91</sup>, Rodney J. Scott<sup>92,43</sup>, Jennifer A. Smith<sup>60</sup>, Angelina R. Sutin<sup>93,55</sup>, Maciej Trzaskowski<sup>94,12</sup>, Anna E. Vinkhuyzen<sup>12</sup>, Lei Yu<sup>95</sup>, Delilah Zabaneh<sup>94</sup>, John R. Attia<sup>42,43</sup>, David A. Bennett<sup>95</sup>, Klaus Berger<sup>57</sup>, Lars Bertram<sup>96,97</sup>, Dorret I. Boomsma<sup>4,5,98</sup>, Harold Snieder<sup>38</sup>, Shun-Chiao Chang<sup>99</sup>, Francesco Cucca<sup>100</sup>, Ian J. Deary<sup>64,65</sup>, Cornelia M. van Duijn<sup>2</sup>, Johan G. Eriksson<sup>101,102,103</sup>, Ute Bültmann<sup>104</sup>, Eco J.C. de Geus<sup>4,5,98</sup>, Patrick J.F. Groenen<sup>3,105</sup>, Vilundur Gudnason<sup>52,53</sup>, Torben Hansen<sup>16</sup>, Catharine A. Hartman<sup>106</sup>, Claire M.A. Haworth<sup>54</sup>, Caroline Hayward<sup>107,108</sup>, Andrew C. Heath<sup>109</sup>, David A. Hinds<sup>22</sup>, Elina Hyppönen<sup>110,20,111</sup>, William G. Iacono<sup>13</sup>, Marjo-Riitta Järvelin<sup>112,113,82,114</sup>, Karl-Heinz Jöckel<sup>24</sup>, Jaakko Kaprio<sup>18,115,116</sup>, Sharon L.R. Kardinaal<sup>60</sup>, Liisa Keltikangas-Järvinen<sup>31</sup>, Peter Kraft<sup>117</sup>, Laura D. Kubzansky<sup>118</sup>, Terho Lehtimäki<sup>119,120</sup>, Patrik K.E. Magnusson<sup>29</sup>, Nicholas G. Martin<sup>121</sup>, Matt McGue<sup>13</sup>, Andres Metspalu<sup>37,122</sup>, Melinda Mills<sup>123</sup>, Renée de Mutsert<sup>39</sup>, Albertine J. Oldehinkel<sup>106</sup>, Gerard Pasterkamp<sup>28,124</sup>, Nancy L. Pedersen<sup>29</sup>, Robert Plomin<sup>125</sup>, Ozren Polasek<sup>83</sup>, Christine Power<sup>20,111</sup>, Stephen S. Rich<sup>126</sup>, Frits R. Rosendaal<sup>39</sup>, Hester M. den Ruijter<sup>28</sup>, David Schlessinger<sup>44</sup>, Helena Schmidt<sup>127,87</sup>, Rauli Svento<sup>128</sup>, Reinhold Schmidt<sup>87</sup>, Behrooz Z. Alizadeh<sup>38,129</sup>, Thorkild I.A. Sørensen<sup>16,54,130</sup>, Tim D. Spector<sup>75</sup>, Andrew Steptoe<sup>23</sup>, Antonio Terracciano<sup>93,55</sup>, A. Roy Thurik<sup>1,3,131,132</sup>, Nicholas J. Timpson<sup>54</sup>, Henning Tiemeier<sup>2,133,134</sup>, André G. Uitterlinden<sup>2,3,135</sup>, Peter Vollenweider<sup>88</sup>, Gert G. Wagner<sup>35,66,136</sup>, David R. Weir<sup>70</sup>, Jian Yang<sup>12,137</sup>, Dalton C. Conley<sup>138</sup>, George Davey Smith<sup>54</sup>, Albert Hofman<sup>2,139</sup>, Magnus Johannesson<sup>140</sup>, David I. Laibson<sup>7</sup>, Sarah E. Medland<sup>34</sup>, Michelle N. Meyer<sup>141,142</sup>, Joseph K. Pickrell<sup>14,143</sup>, Tõnu Esko<sup>37</sup>, Robert F. Krueger<sup>13,#</sup>, Jonathan P. Beauchamp<sup>7,#</sup>, Philipp D. Koellinger<sup>9,3,10,#</sup>, Daniel J. Benjamin<sup>8,#</sup>, Meike Bartels<sup>4,5,98,#</sup>, and David Cesarini<sup>144,145,#</sup>

## Affiliations

<sup>1</sup>Department of Applied Economics, Erasmus School of Economics, Erasmus University Rotterdam, 3062 PA, Rotterdam, The Netherlands <sup>2</sup>Department of Epidemiology, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands <sup>3</sup>Erasmus University Rotterdam Institute for Behavior and Biology, Rotterdam, 3062 PA, The Netherlands <sup>4</sup>Department of Biological Psychology, Vrije Universiteit, Amsterdam, 1081 BT, The Netherlands <sup>5</sup>EMGO+ Institute for Health and Care Research, Amsterdam, 1081 BT, The Netherlands <sup>6</sup>Saïd Business School, University of Oxford, Oxford, OX1 1HP, UK <sup>7</sup>Department of Economics, Harvard University, Cambridge, MA 02138, USA <sup>8</sup>Center for Economic and Social Research, University of Southern California, Los Angeles, CA 90089-3332, USA <sup>9</sup>Department of Complex Trait Genetics, VU University, Center for Neurogenomics and Cognitive Research, Amsterdam, 1081 HV, The Netherlands <sup>10</sup>Amsterdam Business School, University of Amsterdam, Amsterdam, 1018 TV, The Netherlands <sup>11</sup>Psychology, University of Illinois, IL 61820, Champaign, USA <sup>12</sup>Queensland Brain Institute, The University of Queensland, Brisbane, QLD 4072, Australia <sup>13</sup>Department of Psychology, University of Minnesota Twin Cities, Minneapolis, MN 55455, USA

<sup>14</sup>New York Genome Center, New York, NY 10013, USA <sup>15</sup>COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, 2820, Denmark <sup>16</sup>The Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, University of Copenhagen, Faculty of Health and Medical Sciences, Copenhagen, 2100, Denmark <sup>17</sup>Steno Diabetes Center, Gentofte, 2820, Denmark <sup>18</sup>Department of Public Health, University of Helsinki, Helsinki, FI-00014, Finland <sup>19</sup>Centre for Environmental and Preventive Medicine, Wolfson Institute of Preventive Medicine, Queen Mary University of London, London EC1M 6BQ, UK <sup>20</sup>South Australian Health and Medical Research Institute, Adelaide, SA5000, Australia <sup>21</sup>USDA/ARS Children's Nutrition Research Center, Baylor College of Medicine, Houston, TX 77030, USA <sup>22</sup>23andMe, Inc., Mountain View, CA 94041, USA <sup>23</sup>Department of Epidemiology & Public Health, University College London, London WC1E 6BT, UK <sup>24</sup>Institute for Medical Informatics, Biometry and Epidemiology, University Hospital of Essen, Essen, 45147, Germany <sup>25</sup>Centre for Research in Environmental Epidemiology, Institute for Global Health, Barcelona, Spain <sup>26</sup>Universitat Pompeu Fabra, Barcelona, Spain <sup>27</sup>CIBER Epidemiología y Salud Pública, Barcelona, Spain <sup>28</sup>Laboratory of Experimental Cardiology, University Medical Center Utrecht, Utrecht, 3584 CX, The Netherlands <sup>29</sup>Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, 171 77, Sweden <sup>30</sup>Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, 85764, Germany <sup>31</sup>Institute of Behavioural Sciences, P.O. Box 9, 00014 University of Helsinki, Finland <sup>32</sup>Folkhälsan Research Centre, Helsingfors, FI-00014, Finland <sup>33</sup>Helsinki Collegium for Advanced Studies, University of Helsinki, Helsinki 00014, Finland <sup>34</sup>Psychiatric Genetics, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4029, Australia <sup>35</sup>Max Planck Institute for Human Development, Berlin, 14195, Germany <sup>36</sup>Max Planck Institute for Molecular Genetics, Department of Vertebrate Genomics, Berlin, 14195, Germany <sup>37</sup>Estonian Genome Center, University of Tartu, Tartu, 51010, Estonia <sup>38</sup>Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, 9700 RB, The Netherlands <sup>39</sup>Clinical Epidemiology, Leiden University Medical Center, Leiden, 2300 RC, The Netherlands <sup>40</sup>Public Health and Primary Care, Leiden University Medical Center, Leiden, 2300 RC, The Netherlands <sup>41</sup>BESC, King Faisal Specialist Hospital and Research Centre, Riyadh 11211, Saudi Arabia <sup>42</sup>Public Health Stream, Hunter Medical Research Institute, New Lambton, NSW 2305, Australia <sup>43</sup>Faculty of Health and Medicine, University of Newcastle, Newcastle, NSW 2300, Australia <sup>44</sup>Laboratory of Genetics, National Institute on Aging, Baltimore, MD 21224, USA <sup>45</sup>Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, 20521, Finland <sup>46</sup>Department of Clinical Physiology, Turku University Hospital, Turku 20520, Finland <sup>47</sup>Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, 85764, Germany <sup>48</sup>Department of Psychology, University of Tartu, Tartu 50409, Estonia <sup>49</sup>Department of Psychology, University of Warwick, CV4 7AL Coventry, UK <sup>50</sup>Department of Computational

Biology, University of Lausanne, Lausanne, 1011, Switzerland <sup>51</sup>Swiss Institute of Bioinformatics, Lausanne, 1015, Switzerland <sup>52</sup>Icelandic Heart Association, Kopavogur, 201, Iceland <sup>53</sup>Faculty of Medicine, University of Iceland, Reykjavik, 101, Iceland <sup>54</sup>MRC Integrative Epidemiology Unit, University of Bristol, Bristol, BS8 2BN, UK <sup>55</sup>National Institute on Aging, National Institutes of Health, Baltimore, MD 21224, USA <sup>56</sup>Los Angeles Biomedical Research Institute and Department of Pediatrics, Harbor-UCLA, Torrance, 90505 CA, USA <sup>57</sup>Institute of Epidemiology and Social Medicine, University of Muenster, Muenster, 48149, Germany <sup>58</sup>Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA <sup>59</sup>Division of Rheumatology, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA <sup>60</sup>Department of Epidemiology, University of Michigan, Ann Arbor, MI 48109, USA <sup>61</sup>LifeLines Cohort Study, University of Groningen, University Medical Center Groningen, Groningen, 9713 BZ, The Netherlands <sup>62</sup>Department of Behavioral Sciences, Rush University Medical Center, Chicago, IL 60612, USA <sup>63</sup>RAND Corporation, Santa Monica, CA 9041-3208, USA <sup>64</sup>Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, EH8 9JZ, UK <sup>65</sup>Department of Psychology, University of Edinburgh, Edinburgh, EH8 9JZ, UK <sup>66</sup>German Socio-Economic Panel Study, DIW Berlin, Berlin, 10117, Germany <sup>67</sup>Health Economics Research Centre, Nuffield Department of Population Health, University of Oxford, Oxford, OX3 7LF, UK <sup>68</sup>Department of Epidemiology, Geisel School of Medicine at Dartmouth, Lebanon, New Hampshire, USA <sup>69</sup>The Farr Institute of Health Informatics, University College London, London NW1 2DA, UK <sup>70</sup>Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI 48104, USA <sup>71</sup>National Institute on Aging, National Institutes of Health, Baltimore, MD 21224, USA <sup>72</sup>Institute of Human Genetics, University of Bonn, Bonn, 53127, Germany <sup>73</sup>Department of Genomics, Life and Brain Center, University of Bonn, Bonn, 53127, Germany <sup>74</sup>Laboratory of Epidemiology, Demography, National Institute on Aging, National Institutes of Health, Bethesda, MD 20892-9205, United States <sup>75</sup>Department of Twin Research and Genetic Epidemiology, King's College London, London, SE1 7EH, UK <sup>76</sup>Program in Translational NeuroPsychiatric Genomics, Departments of Neurology & Psychiatry, Brigham and Women's Hospital, Boston, MA 02115, USA <sup>77</sup>Harvard Medical School, Boston, MA 02115, USA <sup>78</sup>Program in Medical and Population Genetics, Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA, 02142, USA <sup>79</sup>Department of Genomics of Common Disease, Imperial College London, London, W12 0NN, UK <sup>80</sup>Department of Pediatrics, University of Helsinki, Helsinki, Finland <sup>81</sup>National Institute for Health and Welfare, Helsinki, Finland <sup>82</sup>Center for Life Course Health Research, University of Oulu; Oulu University Hospital, Oulu, Finland <sup>83</sup>Department of Public Health, Faculty of Medicine, University of Split, Croatia, Split 21000, Croatia <sup>84</sup>Institute for Social & Economic Research, University of Essex, Wivenhoe Park CO4 3SQ, UK <sup>85</sup>Neuroepidemiology Section, National Institute on Aging, National Institutes of Health, Bethesda, MD 20892-9205, USA <sup>86</sup>Department of Genetics, University Medical Center Groningen,

University of Groningen, Groningen. 9700 RB, The Netherlands <sup>87</sup>Department of Neurology, General Hospital and Medical University Graz, Graz, 8036, Austria <sup>88</sup>Department of Internal Medicine, Internal Medicine, Lausanne University Hospital (CHUV), Lausanne, 1011, Switzerland <sup>89</sup>Molecular Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4029, Australia <sup>90</sup>Department of Neuroscience, Karolinska Institutet, Retzius Väg 8 171 65 Stockholm, Sweden <sup>91</sup>Centre for Global Health Research, The Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, EH8 9AG, UK <sup>92</sup>Information Based Medicine Stream, Hunter Medical Research Institute, New Lambton, NSW 2305, Australia <sup>93</sup>Department of Behavioral Sciences and Social Medicine, Florida State University College of Medicine, Tallahassee, FL 32306, USA <sup>94</sup>Department of Public Health, Faculty of Medicine, University of Split, Croatia, Split 21000, Croatia <sup>95</sup>Department of Neurological Sciences, Rush University Medical Center, Chicago, IL 60612, USA <sup>96</sup>Lübeck Interdisciplinary platform for Genome Analytics (LIGA), Institutes of Neurogenetics and Integrative & Experimental Genomics, University of Lübeck, Lübeck, 23562, Germany <sup>97</sup>Neuroepidemiology and Ageing Research Unit, School of Public Health, Faculty of Medicine, Imperial College, London SW7 2AZ, UK <sup>98</sup>Neuroscience Campus Amsterdam, Amsterdam, The Netherlands <sup>99</sup>Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA <sup>100</sup>Istituto di Ricerca Genetica e Biomedica (IRGB), Consiglio Nazionale delle Ricerche, c/o Cittadella Universitaria di Monserrato, Monserrato, Cagliari, 9042, Italy <sup>101</sup>Department of General Practice and Primary Health Care, University of Helsinki, Helsinki, 00014, Finland <sup>102</sup>Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland <sup>103</sup>Unit of General Practice, University Central Hospital, Helsinki, Finland <sup>104</sup>Department of Health Sciences, Community & Occupational Medicine, University of Groningen, University Medical Center Groningen, Groningen, 9713 AV, The Netherlands <sup>105</sup>Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, 3062 PA, The Netherlands <sup>106</sup>Department of Psychiatry, University of Groningen, University Medical Center Groningen, 9700 RB, Groningen, The Netherlands <sup>107</sup>MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK <sup>108</sup>Generation Scotland, Centre for Genomics and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK <sup>109</sup>Department of Psychiatry, Washington University School of Medicine, St. Louis, MO 63110, USA <sup>110</sup>Centre for Population Health Research, School of Health Sciences and Sansom Institute, University of South Australia, SA5000, Adelaide, Australia <sup>111</sup>Population, Policy and Practice, UCL Institute of Child Health, London, WC1N 1EH, UK <sup>112</sup>Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment & Health, School of Public Health, Imperial College London, UK <sup>113</sup>Biocenter Oulu, University of Oulu, Oulu, Finland <sup>114</sup>Unit of Primary Care, Oulu University Hospital, Oulu, Finland <sup>115</sup>Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, 00014, Finland <sup>116</sup>Department for Health,

THL-National Institute for Health and Welfare, Helsinki, FI-00271, Finland  
<sup>117</sup>Department of Epidemiology and Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA <sup>118</sup>Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA  
<sup>119</sup>Fimlab Laboratories, Tampere, 33520, Finland <sup>120</sup>Department of Clinical Chemistry, University of Tampere, School of Medicine, Tampere, 33014, Finland  
<sup>121</sup>Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4029, Australia <sup>122</sup>Institute of Molecular and Cell Biology, University of Tartu, Tartu, 51010, Estonia <sup>123</sup>Department of Sociology, University of Oxford, Oxford OX1 3UQ, UK <sup>124</sup>Laboratory of Clinical Chemistry and Hematology, Division Laboratories and Pharmacy, University Medical Center Utrecht, Utrecht, 3584 CX, The Netherlands  
<sup>125</sup>Social, Genetic and Developmental Psychiatry Centre, King's College London, De Crespigny Park SE5 8AF, UK <sup>126</sup>Department of Public Health Sciences, University of Virginia, Charlottesville, VA 22904, USA <sup>127</sup>Research Unit for Genetic Epidemiology, Institute of Molecular Biology and Biochemistry, Center of Molecular Medicine, General Hospital and Medical University, Graz, Graz, 8010, Austria  
<sup>128</sup>Department of Economics, Oulu Business School, Oulu, Finland  
<sup>129</sup>Department of Gastroenterology and Hepatology, University of Groningen, University Medical Center Groningen, Groningen, 9713 GZ, The Netherlands  
<sup>130</sup>Institute of Preventive Medicine, Bispebjerg and Frederiksberg Hospitals, The Capital Region, Frederiksberg, 2000, Denmark <sup>131</sup>Montpellier Business School, Montpellier, 34080, France <sup>132</sup>Panteia, Zoetermeer, 2715 CA, The Netherlands  
<sup>133</sup>Department of Psychiatry, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands  
<sup>134</sup>Department of Child and Adolescent Psychiatry, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands <sup>135</sup>Department of Internal Medicine, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands <sup>136</sup>School of Economics and Management, Berlin University of Technology, Berlin, 10623, Germany  
<sup>137</sup>The University of Queensland Diamantina Institute, The Translational Research Institute, Brisbane, QLD 4102, Australia <sup>138</sup>Department of Sociology, Princeton University, Wallace Hall, Princeton NJ, 08544 <sup>139</sup>Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA  
<sup>140</sup>Department of Economics, Stockholm School of Economics, Stockholm, 113 83, Sweden  
<sup>141</sup>Department of Bioethics, Clarkson University, Schenectady, NY 12308, USA  
<sup>142</sup>Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA  
<sup>143</sup>Department of Biological Sciences, Columbia University, 600 Fairchild Center, New York, NY 10027, USA  
<sup>144</sup>Department of Economics, New York University, New York, NY 10012, USA  
<sup>145</sup>Research Institute for Industrial Economics, Stockholm, 10215, Sweden

## Acknowledgments

This research was carried out under the auspices of the Social Science Genetic Association Consortium (SSGAC). The SSGAC seeks to facilitate studies that investigate the influence of genes on human behavior, well-being, and social-scientific outcomes using large genome-wide association study meta-analyses. The SSGAC also provides opportunities for replication and promotes the collection of accurately measured, harmonized phenotypes across cohorts. The SSGAC operates as a working group within the CHARGE consortium. This research has also been

conducted using the UK Biobank Resource. The study was supported by funding from the U.S. National Science Foundation (EAGER: “Workshop for the Formation of a Social Science Genetic Association Consortium”), a supplementary grant from the National Institute of Health Office of Behavioral and Social Science Research, the Ragnar Söderberg Foundation (E9/11), the Swedish Research Council (421-2013-1061), The Jan Wallander and Tom Hedelius Foundation, an ERC Consolidator Grant (647648 EdGe), the Pershing Square Fund of the Foundations of Human Behavior, and the NIA/NIH through grants P01-AG005842, P01-AG005842-20S2, P30-AG012810, and T32-AG000186-23 to NBER and R01-AG042568-02 to the University of Southern California. We are grateful to Peter M. Visscher for advice, support, and feedback. We thank Samantha Cunningham and Nishanth Galla for research assistance. A full list of acknowledgments is provided in the Supplementary Note.

## References

1. Kendler KS, Myers J. The genetic and environmental relationship between major depression and the five-factor model of personality. *Psychol Med.* 2009; 40:801. [PubMed: 19732485]
2. Weiss A, Bates TC, Luciano M. Happiness is a personal(ity) thing: The genetics of personality and well-being in a representative sample. *Psychol Sci.* 2008; 19:205–210. [PubMed: 18315789]
3. Bartels M, Cacioppo JT, van Beijsterveldt TCEM, Boomsma DI. Exploring the association between well-being and psychopathology in adolescents. *Behav Genet.* 2013; 43:177–190. [PubMed: 23471543]
4. de Moor MHM, et al. Meta-analysis of genome-wide association studies for neuroticism, and the polygenic association with Major Depressive Disorder. *JAMA Psychiatry.* 2015; 72:642–650. [PubMed: 25993607]
5. Hyman S. Mental health: Depression needs large human-genetics studies. *Nature.* 2014; 515:189–191. [PubMed: 25391945]
6. Rietveld CA, et al. Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proc Natl Acad Sci U S A.* 2014; 111:13790–13794. [PubMed: 25201988]
7. Kahneman D, Deaton A. High income improves evaluation of life but not emotional well-being. *Proc Natl Acad Sci U S A.* 2010; 107:16489–16493. [PubMed: 20823223]
8. Kahneman, D.; Riis, J. *The Science of Well-Being.* Uppter, F.; Baylis, N.; Keverne, B., editors. Oxford University Press; 2005. p. 285-301.
9. Bartels M, Boomsma DI. Born to be happy? The etiology of subjective well-being *Behav Genet.* 2009; 39:605–615. [PubMed: 19728071]
10. Bulik-Sullivan BK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet.* 2015; 47:291–295. [PubMed: 25642630]
11. Yang J, et al. Genomic inflation factors under polygenic inheritance. *Eur J Hum Genet.* 2011; 19:807–812. [PubMed: 21407268]
12. Ripke S, et al. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry.* 2013; 18:497–511. [PubMed: 22472876]
13. Sudlow C, et al. UK Biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* 2015; 12:e1001779. [PubMed: 25826379]
14. dbGaP. Resource for Genetic Epidemiology Research on Adult Health and Aging (GERA). 2015. at <[http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs000674.v1.p1](http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000674.v1.p1)>
15. Eysenck, HJ.; Eysenck, SBG. *Manual of the Eysenck Personality Questionnaire.* Hodder and Stroughton; 1975.
16. Tian C, et al. Analysis and application of European genetic substructure using 300 K SNP information. *PLoS Genet.* 2008; 4:e4. [PubMed: 18208329]
17. Steinberg KM, et al. Structural diversity and African origin of the 17q21.31 inversion polymorphism. *Nat Genet.* 2012; 44:872–80. [PubMed: 22751100]
18. Smith DJ, et al. Genome-wide analysis of over 106,000 individuals identifies 9 neuroticism-associated loci. 2015; bioRxiv 032417. doi: 10.1101/032417
19. Meuwissen TH, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics.* 2001; 157:1819–1829. [PubMed: 11290733]
20. Vilhjálmsson BJ, et al. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am J Hum Genet.* 2015; 97:576–592. [PubMed: 26430803]

21. Hyde CL, et al. Common genetic variants associated with major depressive disorder among individuals of European descent. *Nat Genet.*
22. Roberts RE, Kaplan Ga, Shema SJ, Strawbridge WJ. Are the obese at greater risk for depression? *Am J Epidemiol.* 2000; 152:163–170. [PubMed: 10909953]
23. Glassman AJ, et al. Smoking, smoking cessation, and major depression. *J Am Med Assoc.* 1990; 264:1546–1549.
24. Shahab L, West R. Differences in happiness between smokers, ex-smokers and never smokers: Cross-sectional findings from a national household survey. *Drug Alcohol Depend.* 2012; 121:38–44. [PubMed: 21906891]
25. Rugulies R. Depression as a predictor for coronary heart disease: A review and meta-analysis. *Am J Prev Med.* 2002; 23:51–61. [PubMed: 12093424]
26. Finucane HK, et al. Partitioning heritability by functional category using GWAS summary statistics. *Nat Genet.* 2015; 47:1228–1235. [PubMed: 26414678]
27. Wood AR, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet.* 2014; 46:1173–1186. [PubMed: 25282103]
28. Stetler C, Miller GE. Depression and hypothalamic-pituitary-adrenal activation: a quantitative summary of four decades of research. *Psychosom Med.* 2011; 73:114–126. [PubMed: 21257974]
29. Seeman P. Dopamine D2 receptors as treatment targets in schizophrenia. *Clin Schizophr Relat Psychoses.* 2010; 4:56–73. [PubMed: 20643630]
30. Vallone D, Picetti R, Borrelli E. Structure and function of dopamine receptors. *Neurosci Biobehav Rev.* 2000; 24:125–132. [PubMed: 10654668]
31. Ripke S, et al. Biological insights from 108 schizophrenia-associated genetic loci. *Nature.* 2014; 511:421–427. [PubMed: 25056061]
32. Shungin D, et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature.* 2015; 518:187–196. [PubMed: 25673412]
33. Kathiresan S, et al. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet.* 2009; 41:56–65. [PubMed: 19060906]
34. Spencer CCA, et al. Dissection of the genetics of Parkinson's disease identifies an additional association 5' of SNCA and multiple associated haplotypes at 17q21. *Hum Mol Genet.* 2011; 20:345–353. [PubMed: 21044948]
35. Höglinger GU, et al. Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. *Nat Genet.* 2011; 43:699–705. [PubMed: 21685912]
36. Sullivan PF. Don't give up on GWAS. *Mol Psychiatry.* 2012; 17:2–3. [PubMed: 21826059]
37. Consortium TIH. A second generation human haplotype map of over 3.1 million SNPs. *Nature.* 2007; 449:851–861. [PubMed: 17943122]
38. Winkler TW, et al. Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc.* 2014; 9:1192–1212. [PubMed: 24762786]
39. Cáceres A, González JR. Following the footprints of polymorphic inversions on SNP data: from detection to association tests. *Nucleic Acids Res.* 2015; 43:e53. [PubMed: 25672393]
40. Yang J, et al. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet.* 2010; 42:565–569. [PubMed: 20562875]
41. Sonnega A, et al. Cohort Profile: the Health and Retirement Study (HRS). *Int J Epidemiol.* 2014; 43:576–585. [PubMed: 24671021]
42. Willemsen G, et al. The Adult Netherlands Twin Register: Twenty-Five Years of Survey and Biological Data Collection. *Twin Res Hum Genet.* 2013; 16:271–281. [PubMed: 23298648]
43. van Beijsterveldt CEM, et al. The Young Netherlands Twin Register (YNTR): Longitudinal Twin and Family Studies in Over 70,000 Children. *Twin Res Hum Genet.* 2013; 16:252–267. [PubMed: 23186620]
44. Welter D, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* 2014; 42:D1001–1006. [PubMed: 24316577]
45. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* 2012; 40:D930–4. [PubMed: 22064851]

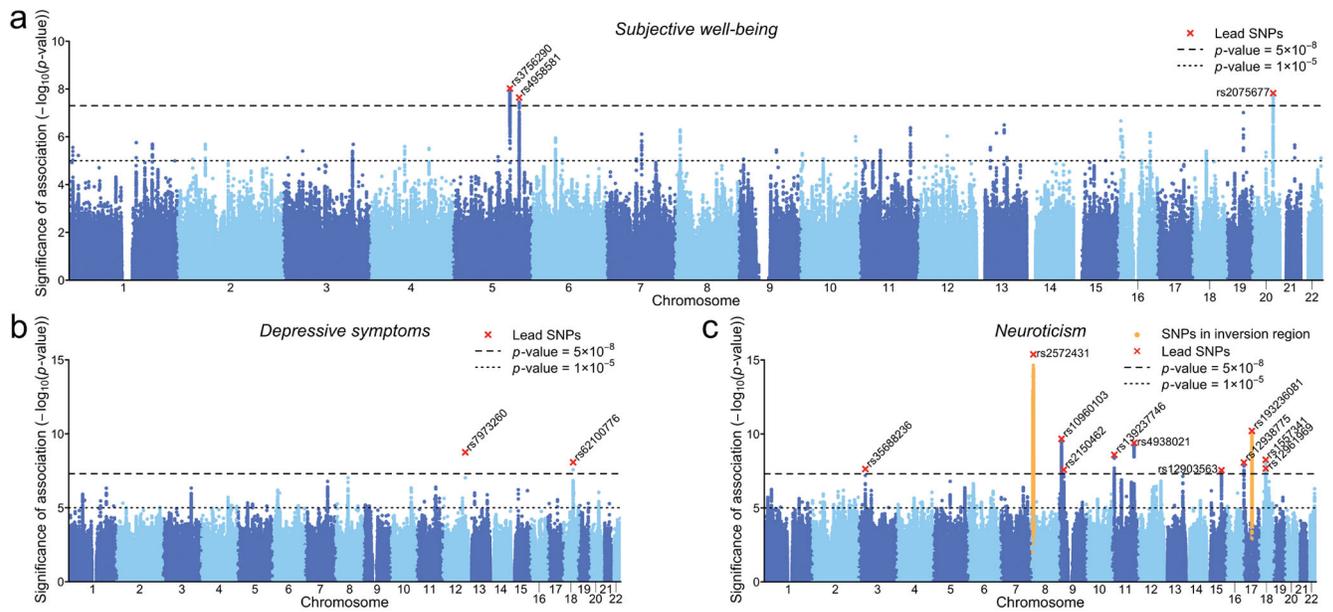
46. Penninx BWJH, et al. The Netherlands study of depression and anxiety (NESDA): Rationale, objectives and methods. *Int J Methods Psychiatr Res.* 2008; 17:121–140. [PubMed: 18763692]
47. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 2013; 45:580–5. [PubMed: 23715323]
48. Ardlie KG, et al. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science.* 2015; 348:648–660. [PubMed: 25954001]
49. Fehrmann RSN, et al. Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat Genet.* 2015; 47:115–125. [PubMed: 25581432]

Author Manuscript

Author Manuscript

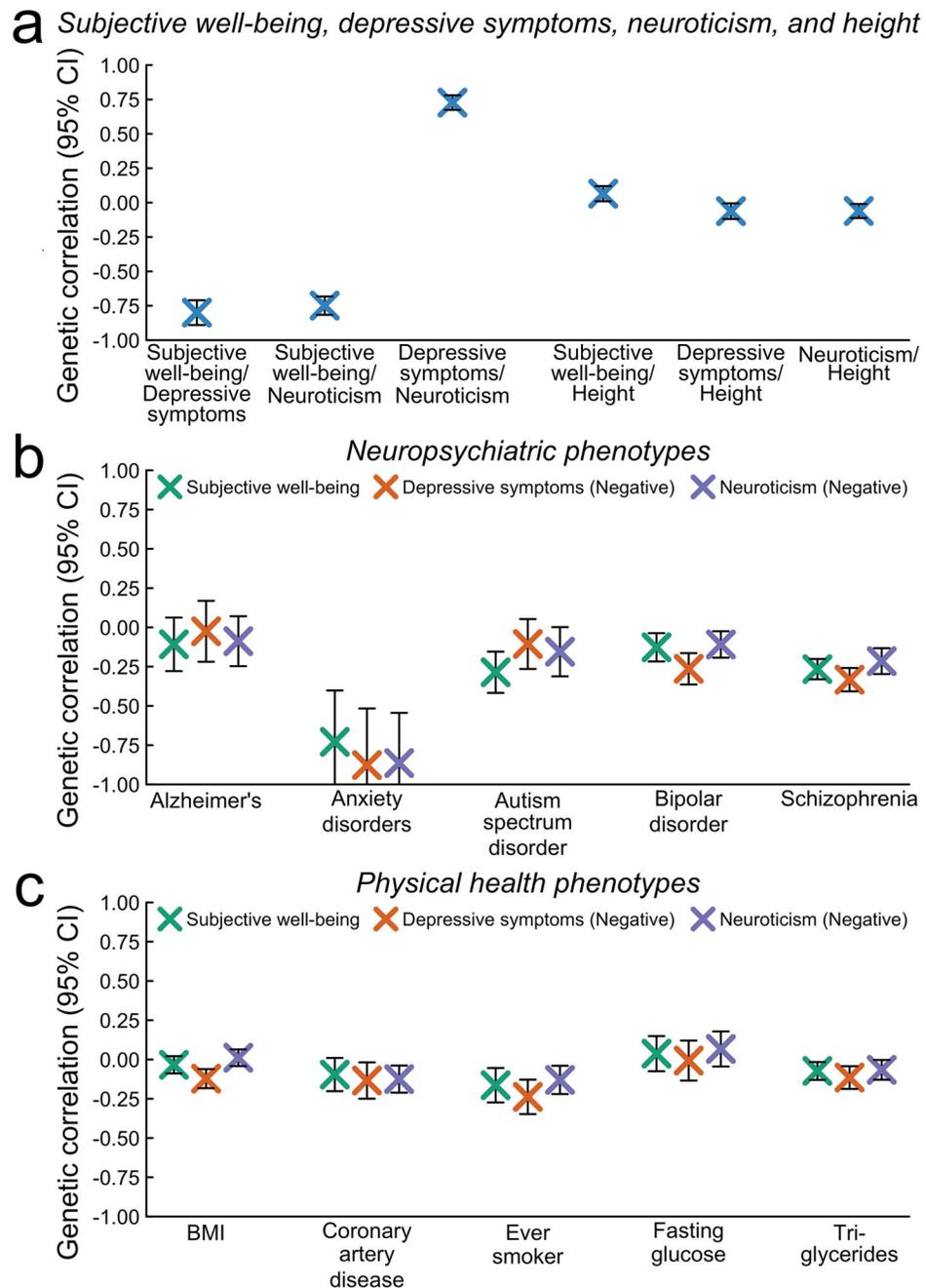
Author Manuscript

Author Manuscript



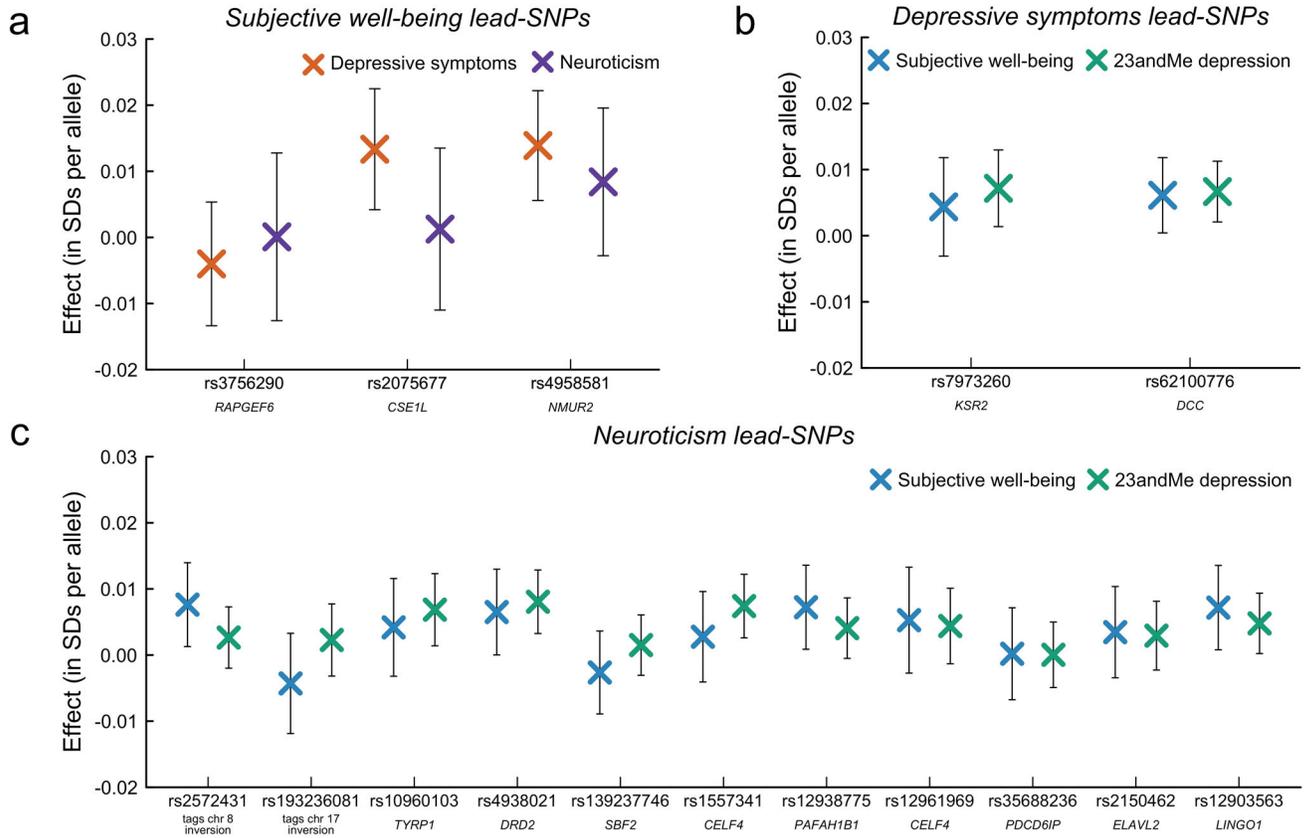
**Fig. 1. Manhattan plots of GWAS results**

(a) Subjective well-being ( $N = 298,420$ ), (b) Depressive symptoms ( $N = 180,866$ ), (c) Neuroticism ( $N = 170,911$ ). The  $x$ -axis is chromosomal position, and the  $y$ -axis is the significance on a  $-\log_{10}$  scale. The upper dashed line marks the threshold for genome-wide significance ( $p = 5 \times 10^{-8}$ ); the lower line marks the threshold for nominal significance ( $p = 10^{-5}$ ). Each approximately independent genome-wide significant association (“lead SNP”) is marked by  $\times$ . Each lead SNP is the lowest  $p$ -value SNP within the locus, as defined by our clumping algorithm (Supplementary Note).



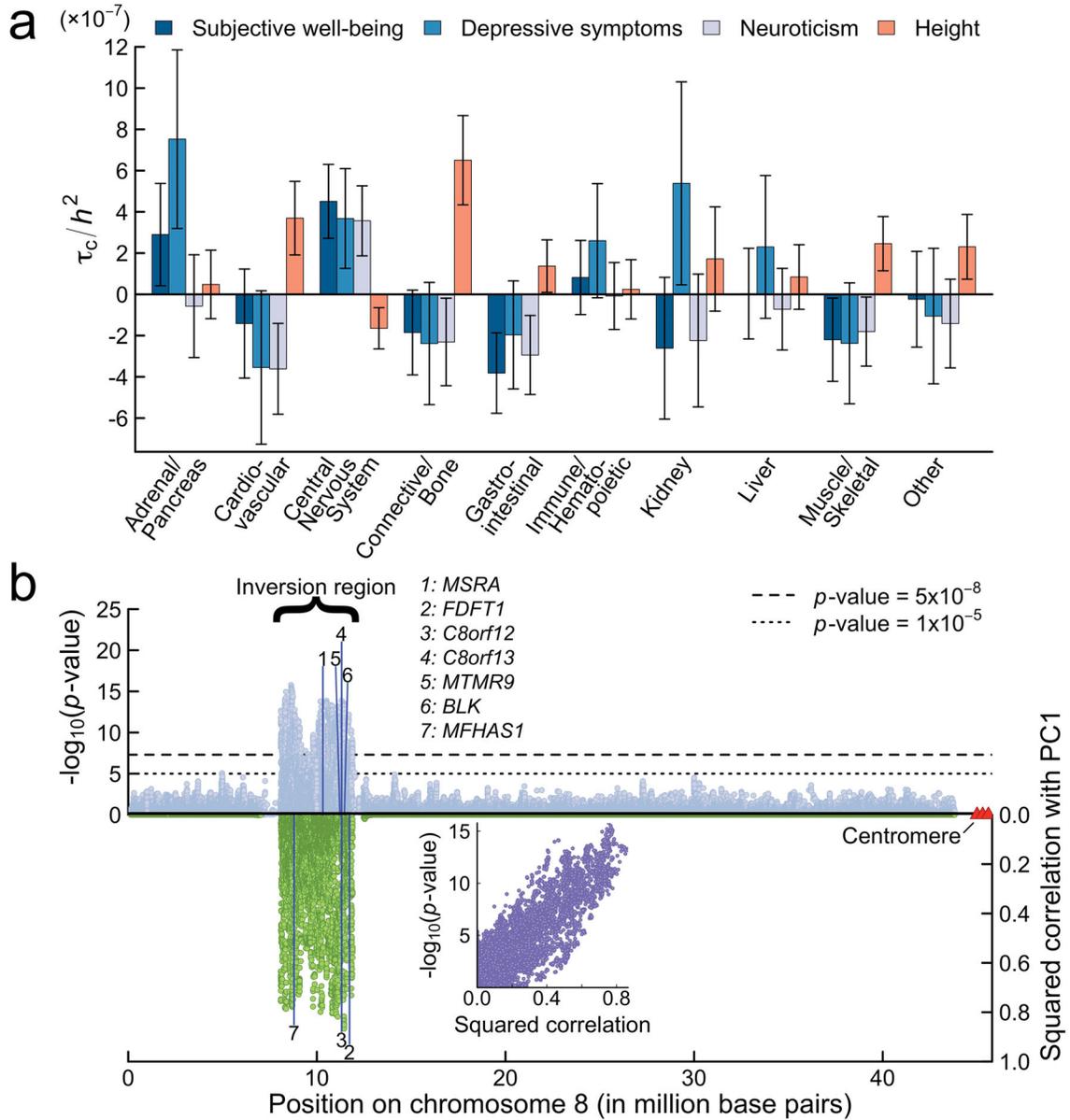
**Fig. 2. Genetic correlations with bars representing 95% confidence intervals**

The correlations are estimated using bivariate LD Score (LDSC) regression. **(a)** Genetic correlations between subjective well-being, depressive symptoms, and neuroticism (“our three phenotypes”), as well as between our three phenotypes and height. **(b)** Genetic correlations between our three phenotypes and selected neuropsychiatric phenotypes. **(c)** Genetic correlations between our three phenotypes and selected physical health phenotypes. In **(b)** and **(c)**, we report the negative of the estimated correlation with depressive symptoms and neuroticism (but not subjective well-being).



**Fig. 3. Quasi-replication and lookup of lead SNPs**

In quasi-replication analyses, we examined whether (a) lead SNPs identified in the subjective well-being meta-analyses are associated with depressive symptoms or neuroticism, (b) lead SNPs identified in the analyses of depressive symptoms are associated with subjective well-being, and (c) lead SNPs identified in the analyses of neuroticism are associated with subjective well-being. The quasi-replication sample is always restricted to non-overlapping cohorts. In a separate lookup exercise, we examined whether lead SNPs for depressive symptoms and neuroticism are associated with depression in an independent sample of 23andMe customers ( $N = 368,890$ ). The results from this lookup are depicted as green crosses in (b) and (c). Bars represent 95% CIs (not adjusted for multiple testing). For interpretational ease, we choose the reference allele so that positive coefficients imply that the estimated effect is in the predicted direction. Listed below each lead SNP is the nearest gene.



**Fig. 4. Results from selected biological analyses**

(a) Estimates of the expected increase in the phenotypic variance accounted for by a SNP due to the SNP's being in a given category ( $\tau_c$ ), divided by the LD Score heritability of the phenotype ( $h^2$ ). Each estimate of  $\tau_c$  comes from a separate stratified LD Score regression, controlling for the 52 functional annotation categories in the “baseline model.” The bars represent 95% CIs (not adjusted for multiple testing). To benchmark the estimates, we compare them to those obtained from a recent study of height<sup>27</sup>. (b) Inversion polymorphism on chromosome 8 and the 7 genes for which the inversion is a significant *cis*-eQTL at FDR < 0.05. The upper half of the figure shows the Manhattan plot for neuroticism for the inversion and surrounding regions. The bottom half shows the squared correlation between the SNPs and the principal component that captures the inversion. The inset plots the

relationship, for each SNP in the inversion region, between the SNP's significance and its squared correlation with the principal component that captures the inversion.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Summary of polymorphisms identified across analyses.

Panel A. Genome-Wide Significant Associations										
Subjective Well-Being (SWB, $N = 298,420$ )										
SNPID	CHR	BP	EA	EAF	Beta	SE	$R^2$	$p$ -value	$N$	Quasi-Repl
rs3756290	5	130,951,750	A	0.24	-0.0177	0.0031	0.011%	$9.6 \times 10^{-9}$	286,851	
rs2075677	20	47,701,024	A	0.76	0.0175	0.0031	0.011%	$1.5 \times 10^{-8}$	288,454	DS**
rs4958581	5	152,187,729	T	0.66	0.0153	0.0027	0.011%	$2.3 \times 10^{-8}$	294,043	DS***
Neuroticism ( $N = 170,908$ )										
SNPID	CHR	BP	EA	EAF	Beta	SE	$R^2$	$p$ -value	$N$	Quasi-Repl
rs2572431#	8	11,105,077	T	0.41	0.0283	0.0035	0.039%	$4.2 \times 10^{-16}$	170,908	SWB*
rs193236081##	17	44,142,332	T	0.77	-0.0284	0.0045	0.028%	$6.3 \times 10^{-11}$	151,297	
rs10960103	9	11,699,270	C	0.77	0.0264	0.0038	0.024%	$2.1 \times 10^{-10}$	165,380	$D^*$ $_{23andMe}$
rs4938021	11	113,364,803	T	0.34	0.0233	0.0037	0.024%	$4.0 \times 10^{-10}$	159,900	$D^{***}$ $_{23andMe}, SWB^*$
rs139237746	11	10,253,183	T	0.51	-0.0204	0.0034	0.021%	$2.6 \times 10^{-9}$	170,908	
rs1557341	18	35,127,427	A	0.34	0.0213	0.0036	0.021%	$5.6 \times 10^{-9}$	165,579	$D^{**}$ $_{23andMe}$
rs12938775	17	2,574,821	A	0.47	-0.0202	0.0035	0.020%	$8.5 \times 10^{-9}$	163,283	SWB*
rs12961969	18	35,364,098	A	0.2	0.0250	0.0045	0.020%	$2.2 \times 10^{-8}$	156,758	
rs35688236	3	34,582,993	A	0.69	0.0213	0.0037	0.019%	$2.4 \times 10^{-8}$	161,636	
rs2150462	9	23,316,330	C	0.26	-0.0217	0.0038	0.018%	$2.7 \times 10^{-8}$	170,907	
rs12903563	15	78,033,735	T	0.50	0.0198	0.0036	0.020%	$2.9 \times 10^{-8}$	157,562	$D^*$ $_{23andMe}, SWB^*$
Depressive Symptoms (DS, $N = 180,866$ )										
SNPID	CHR	BP	EA	EAF	Beta	SE	$R^2$	$p$ -value	$N$	Quasi-Repl/Repl

**Panel A. Genome-Wide Significant Associations**

**Subjective Well-Being (SWB, N = 298,420)**

SNPID	CHR	BP	EA	EAF	Beta	SE	R <sup>2</sup>	p-value	N	Quasi-Repl
rs7973260	12	118,375,486	A	0.19	0.0306	0.0051	0.029%	1.8×10 <sup>-9</sup>	124,498	D* <sub>23andMe</sub>
rs62100776	18	50,754,633	A	0.56	-0.0252	0.0044	0.031%	8.5×10 <sup>-9</sup>	105,739	D** <sub>23andMe,SWB</sub> *

Panel B. SNPs Identified via Proxy-Phenotype Analyses of SWB Loci with p-value < 10<sup>-4</sup>

Depressive Symptoms in Non-Overlapping Cohorts

SNPID	CHR	BP	EA	EAF	Beta <sub>DS</sub>	SE <sub>DS</sub>	R <sup>2</sup>	p <sub>DS</sub>	Bonferroni	N <sub>DS</sub>
rs4346787 <sup>‡</sup>	6	27,491,299	A	0.113	-0.023	0.0059	0.011%	9.8×10 <sup>-5</sup>	0.0160	142,265
rs4481363	5	164,483,794	A	0.524	0.014	0.0038	0.009%	3.1×10 <sup>-4</sup>	0.0499	142,265

Neuroticism in Non-Overlapping Cohorts

SNPID	CHR	BP	EA	EAF	Beta <sub>neuro</sub>	SE <sub>neuro</sub>	R <sup>2</sup>	p <sub>neuro</sub>	Bonferroni	N <sub>neuro</sub>
rs10838738	11	47,663,049	A	0.49	0.0178	0.0039	0.016%	5.0×10 <sup>-6</sup>	0.0009	131,864
rs10774909	12	117,674,129	C	0.52	-0.0150	0.0039	0.011%	1.2×10 <sup>-4</sup>	0.0203	131,235
rs6904596	6	27,491,299	A	0.09	-0.0264	0.0072	0.012%	2.5×10 <sup>-4</sup>	0.0423	116,335
rs4481363	5	164,474,719	A	0.49	0.0151	0.0040	0.011%	1.9×10 <sup>-4</sup>	0.0316	122,592

EA: effect allele. EAF: effect allele frequency. All effect sizes are reported in units of SDs per allele. "Quasi-Repl.": phenotypes for which SNP was found to be nominally associated in quasi-replication analyses conducted in independent samples.

\* significant at the 5%-level,

\*\* significant at the 1%-level,

\*\*\* significant at the 0.1%-level.

# inversion-tagging polymorphism on chromosome 8.

## inversion-tagging polymorphism on chromosome 17.

<sup>‡</sup> proxy for rs6904596 (R<sup>2</sup> = 0.98).