nature genetics

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

**Very few genetic variants have been associated with depression and neuroticism, likely because of limitations on sample size in previous studies. Subjective well-being, a phenotype that is genetically correlated with both of these traits, has not yet been studied with genome-wide data. We conducted genome-wide association studies of three phenotypes: subjective well-being ($n$ = 298,420), depressive symptoms ($n$ = 161,460), and neuroticism ($n$ = 170,911). We identify 3 variants associated with subjective well-being, 2 variants associated with depressive symptoms, and 11 variants associated with neuroticism, including 2 inversion polymorphisms. The two loci associated with depressive symptoms replicate in an independent depression sample. Joint analyses that exploit the high genetic correlations between the phenotypes ($|\hat{\rho}| \approx 0.8$) 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 or pancreas tissues are strongly enriched for association.**



Subjective well-being—as measured by survey questions on life satisfaction, positive affect, or happiness—is a major topic of research in 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)[1–3]. Depression and neuroticism have received much more attention than subjective well-being in genetic association studies, but the discovery of genetic variants associated with either of them has proven elusive[4,5].

Here we report a series of separate and joint analyses of subjective well-being, depressive symptoms, and neuroticism, which identify 16 genome-wide significant associations across the three phenotypes. 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 associations from our initial genome-wide association study (GWAS) and (ii) to identify new associations (beyond those identified by the GWAS). In achieving the first aim, we investigate whether SNPs associated with subjective well-being 'quasi-replicate' by testing them for association with depressive symptoms and neuroticism. We similarly examine the quasi-replication records of the loci associated with depressive symptoms and neuroticism 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 our second aim, we use a 'proxy-phenotype' approach[6]: we treat the set of loci associated with subjective well-being at $P < 1 \times 10^{-4}$ as candidates and test them for association with depressive symptoms and 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[7,8]. In the **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 prespecified analysis plan, we conducted a sample-size-weighted meta-analysis using data from 59 cohorts ($n$ = 298,420 individuals) 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[9] of high pairwise genetic correlation between life satisfaction and positive affect using bivariate LD Score regression[10] ($\hat{\rho}$ = 0.981 (s.e.m. = 0.065); **Supplementary Table 1**). Details on the 59 participating cohorts, their phenotype measures, genotyping, quality control filters, and association models are provided in the Online Methods, **Supplementary Note**, and **Supplementary Tables 2–6**.

As expected under polygenicity[11], we observed 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 suggested minimal confounding due to

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

Received 11 December 2015; accepted 24 March 2016; published online 18 April 2016; corrected after print 27 June 2016 and 29 August 2016; doi:10.1038/ng.3552

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 (**Fig. 1a**) and are listed in **Table 1**. The SNPs had estimated effects in the range of 0.015–0.018 s.d. 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 increases 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 the Online Methods, **Supplementary Note**, and **Supplementary Tables 8–12** for details on the cohorts, phenotype measures, genotyping, association models, and quality control filters). For depressive symptoms ($n = 180{,}866$), we performed meta-analysis on publicly available results from a study performed by the Psychiatric Genomics Consortium (PGC)[12] together with new results from analyses of the initial release of UK Biobank (UKB) data[13] and the Resource for Genetic Epidemiology Research on Aging (GERA) cohort (database of Genotypes and Phenotypes (dbGaP), phs000674.v1.p1). In the UKB cohort ($n = 105{,}739$), we constructed a continuous phenotype measure by combining responses to two questions asking about the frequency in the past 2 weeks with which the respondent experienced feelings of unenthusiasm or disinterest and feelings of depression or 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 ($n = 170{,}911$), we pooled summary statistics from a published study by the Genetics of Personality Consortium (GPC)[4] with results from a new analysis of UKB data. GPC ($n = 63{,}661$) harmonized different neuroticism batteries. In the UKB cohort ($n = 107{,}245$), our measure was the respondent's score on a 12-item version of the Eysenck Personality Inventory Neuroticism[14].

In both the depressive symptoms and neuroticism GWAS, the heterogeneous phenotypic measures were highly genetically correlated (**Supplementary Table 1**). As in our subjective well-being analyses, there was substantial inflation of the median test statistics ($\lambda_{GC} = 1.168$ for depressive symptoms and 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 were independent according to our locus definition (**Fig. 1c**). However, six of these reside within a well-known inversion polymorphism[15] on chromosome 8. We established that all genome-wide significant signals in the inversion region were attributable to the inversion, and we confirmed that the inversion was associated with neuroticism in both of our neuroticism data sets, 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 six 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 was attributable to the inversion polymorphism (Online Methods and **Supplementary Note**).

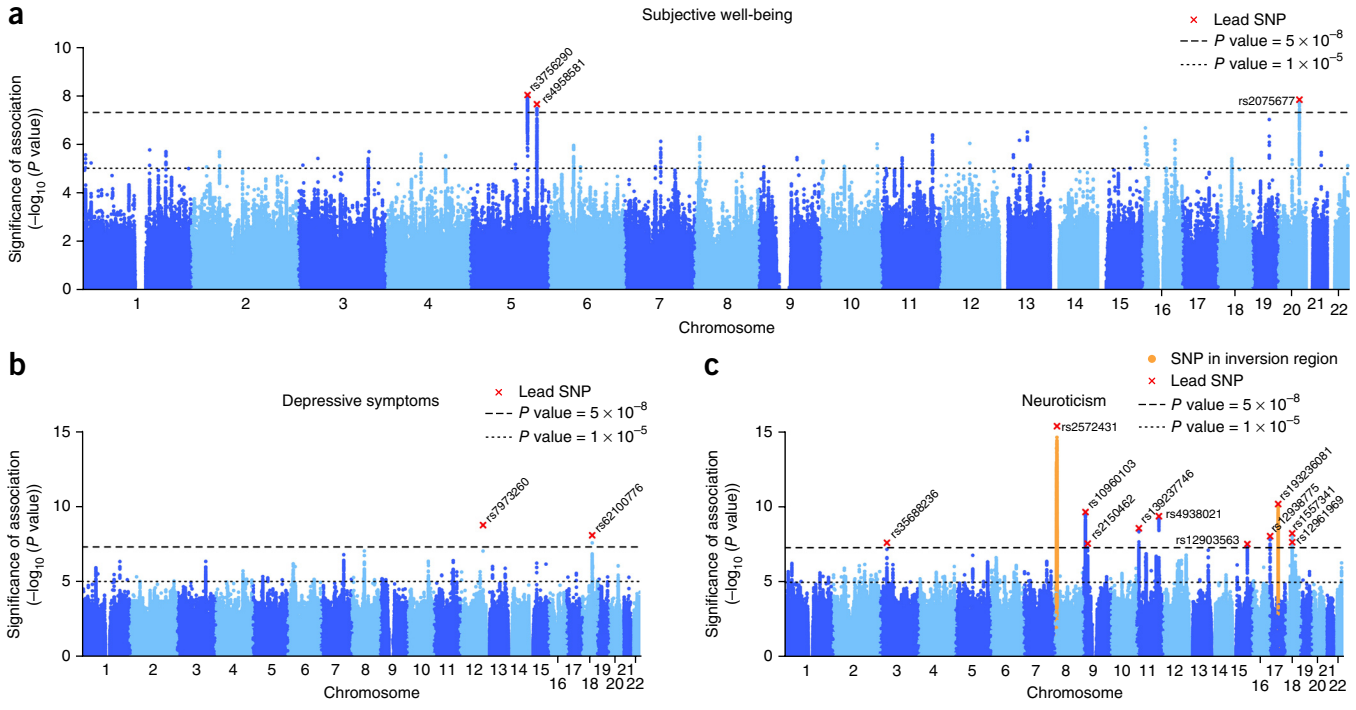


**Figure 1** Manhattan plots of GWAS results. (**a**–**c**) Results are shown for subjective well-being ($n = 298{,}420$) (**a**), depressive symptoms ($n = 180{,}866$) (**b**), and neuroticism ($n = 170{,}911$) (**c**). The *x* axis shows chromosomal position, and the *y* axis shows association significance on a $-\log_{10}$ scale. The upper dashed line marks the threshold for genome-wide significance ($P = 5 \times 10^{-8}$), and the lower dashed line marks the threshold for nominal significance ($P = 1 \times 10^{-5}$). Each approximately independent genome-wide significant association (lead SNP) is marked by a red ×. Each lead SNP is the SNP with the lowest *P* value within the locus, as defined by our clumping algorithm (**Supplementary Note**).

Because this inversion yielded only one significant locus and is genetically complex[16], we hereafter simply use its lead SNP as a proxy for it. Our neuroticism GWAS therefore identified 11 lead SNPs, 2 of which tag inversion polymorphisms. A concurrent neuroticism GWAS using a subset of our sample reported similar findings[17].

The estimated effects of all lead SNPs associated with depressive symptoms and neuroticism were in the range of 0.020–0.031 s.d. per allele ($R^2 \approx 0.02$–$0.04\%$) (**Table 1**). 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 s.d. (**Supplementary Table 13**). The inversion explains 0.06% of variance in neuroticism (roughly the same as the total variance explained jointly by the six SNPs in the inversion region).

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

The three pairwise genetic correlations between our phenotypes, estimated using bivariate LD Score regression[10], are substantial: −0.81 (s.e.m. = 0.046) between subjective well-being and depressive symptoms, −0.75 (s.e.m. = 0.034) between subjective well-being and neuroticism, and 0.75 (s.e.m. = 0.027) between depressive symptoms and neuroticism (**Fig. 2a**). 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, for example, 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 (**Fig. 2**).

## Quasi-replication and Bayesian credibility analyses

We assessed the credibility of our findings using a standard Bayesian framework[18,19] in which a positive fraction of SNPs have null effects and a positive fraction of SNPs 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.

**Table 1 Summary of polymorphisms identified across analyses**

**Genome-wide significant associations**

Subjective well-being ($n$ = 298,420)

| SNP ID | Chr. | Position (bp) | EA | EAF | $\beta$ (s.e.m.) | $R^2$ (%) | $P$ value | $n$ | Quasi-replication[d] |
|---|---|---|---|---|---|---|---|---|---|
| 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 ($n$ = 170,911)

| SNP ID | Chr. | Position (bp) | EA | EAF | $\beta$ (s.e.m.) | $R^2$ (%) | $P$ value | $n$ | Quasi-replication |
|---|---|---|---|---|---|---|---|---|---|
| rs2572431[a] | 8 | 11,105,077 | T | 0.59 | 0.0283 (0.0035) | 0.039 | $4.2 \times 10^{-16}$ | 170,908 | SWB* |
| rs193236081[b] | 17 | 44,142,332 | T | 0.77 | −0.0284 (0.0043) | 0.028 | $6.3 \times 10^{-11}$ | 151,297 | |
| rs10960103 | 9 | 11,699,270 | C | 0.77 | 0.0264 (0.0042) | 0.024 | $2.1 \times 10^{-10}$ | 165,380 | $D^*_{23andMe}$ |
| rs4938021 | 11 | 113,364,803 | T | 0.66 | 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 (0.0037) | 0.021 | $5.6 \times 10^{-9}$ | 165,579 | $D^*_{23andMe}$ |
| rs12938775 | 17 | 2,574,821 | A | 0.53 | −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 (0.0038) | 0.019 | $2.4 \times 10^{-8}$ | 161,636 | |
| rs2150462 | 9 | 23,316,330 | C | 0.74 | −0.0217 (0.0039) | 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 ($n$ = 180,866)

| SNP ID | Chr. | Position (bp) | EA | EAF | $\beta$ (s.e.m.) | $R^2$ (%) | $P$ value | $n$ | Quasi-replication/replication |
|---|---|---|---|---|---|---|---|---|---|
| rs7973260 | 12 | 118,375,486 | A | 0.19 | 0.0306 (0.0051) | 0.029 | $1.8 \times 10^{-9}$ | 124,498 | $D^*_{23andMe}$ |
| rs62100776 | 18 | 50,754,633 | A | 0.56 | −0.0252 (0.0044) | 0.031 | $8.5 \times 10^{-9}$ | 105,739 | $D^*_{23andMe}$, SWB* |

**SNPs identified via proxy-phenotype analyses of subjective well-being loci with $P < 1 \times 10^{-4}$**

Depressive symptoms in non-overlapping cohorts

| SNP ID | Chr. | Position (bp) | EA | EAF | $\beta_{DS}$ ($SE_{DS}$) | $R^2$ (%) | $P_{DS}$ | Bonferroni | $n_{DS}$ |
|---|---|---|---|---|---|---|---|---|---|
| rs4346787[c] | 6 | 27,491,299 | A | 0.113 | −0.023 (0.0059) | 0.011 | $9.8 \times 10^{-5}$ | 0.0160 | 142,265 |
| rs4481363 | 5 | 164,483,794 | A | 0.524 | 0.014 (0.0038) | 0.009 | $3.1 \times 10^{-4}$ | 0.0499 | 142,265 |

Neuroticism in non-overlapping cohorts

| SNP ID | Chr. | Position (bp) | EA | EAF | $\beta_{neuro}$ ($SE_{neuro}$) | $R^2$ (%) | $P_{neuro}$ | Bonferroni | $n_{neuro}$ |
|---|---|---|---|---|---|---|---|---|---|
| rs10838738 | 11 | 47,663,049 | A | 0.49 | 0.0178 (0.0039) | 0.016 | $5.0 \times 10^{-6}$ | 0.0009 | 131,864 |
| rs10774909 | 12 | 117,674,129 | C | 0.52 | −0.0150 (0.0039) | 0.011 | $1.2 \times 10^{-4}$ | 0.0203 | 131,235 |
| rs6904596 | 6 | 27,491,299 | A | 0.09 | −0.0264 (0.0072) | 0.012 | $2.5 \times 10^{-4}$ | 0.0423 | 116,335 |
| rs4481363 | 5 | 164,474,719 | A | 0.49 | 0.0151 (0.0040) | 0.011 | $1.9 \times 10^{-4}$ | 0.0316 | 122,592 |

Chr., chromosome; EA, effect allele; EAF, effect allele frequency (all effect sizes are reported in units of s.d. per allele); SWB, subjective well-being; DS, depressive symptoms; D, depression; SE, standard error.

[a]Inversion-tagging polymorphism on chromosome 8. [b]Inversion-tagging polymorphism on chromosome 17. [c]Proxy for rs6904596 ($r^2$ = 0.98). [d]Phenotypes with which a 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.

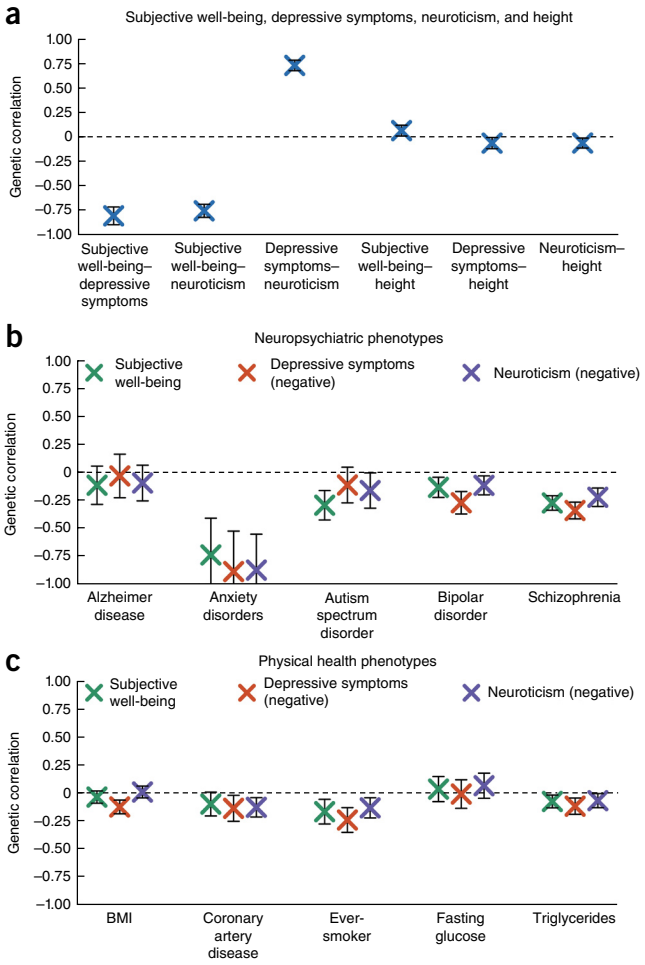


**Figure 2** Genetic correlations. Correlations were estimated using bivariate LD Score (LDSC) regression. (**a**) Genetic correlations between subjective well-being, depressive symptoms, and neuroticism ('our three phenotypes'), as well as between our three phenotypes and height. (**b**) Genetic correlations 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). Error bars represent 95% confidence intervals.

We found that, for any assumption about the fraction of non-null SNPs in the range 1–99%, the probability of true association always exceeded 95% for all 16 loci (and always exceeded 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**. The reference allele for each association in the figure is chosen such that the predicted sign of the second-stage estimate is positive. We found that two of the three subjective well-being lead SNPs were significantly associated with depressive symptoms ($P$ = 0.004 and $P$ = 0.001) in the predicted direction. For neuroticism, where the second-stage sample ($n$ = 68,201) was about one-half as large, the allele increasing subjective well-being had the predicted sign for all three SNPs, but none reached significance.

We also show the results for the depressive symptoms (**Fig. 3b**) and neuroticism (**Fig. 3c**) lead SNPs. In each panel, the blue crosses depict results from the quasi-replications where subjective well-being was the second-stage phenotype. We found that the two depressive symptoms lead SNPs had the predicted sign for subjective well-being, and one was nominally significant ($P$ = 0.04). Finally, of the 11 neuroticism lead SNPs, 9 had the predicted sign for subjective well-being. Four of the 11 had nominally significant association with subjective well-being, all with the predicted sign. One of the four is the SNP tagging the inversion on chromosome 8 (ref. 15). That SNP's association with neuroticism (and likely with subjective well-being) was 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 were 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**; C.L. Hyde, M.W. Nagle, C. Tian, X. Chen, S.A. Paciga *et al.*, unpublished data). Because the depression sample overlaps with our subjective well-being sample, we did not request a lookup of the SNPs associated with subjective well-being.

In **Figure 3b**,**c**, 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 had the predicted sign. Of the 11 neuroticism polymorphisms, 4 were significantly associated with depression at the 5% level. Both of the depressive symptoms lead SNPs replicated ($P$ = 0.004 and $P$ = 0.015), with effect sizes (0.007 and −0.007 s.d. per allele) close to those predicted by our Bayesian framework (0.008 and −0.006) (**Supplementary Tables 14** and **15**).

The top part of **Table 1** summarizes the results for the 16 lead SNPs identified across our separate genome-wide association analyses of the three phenotypes. The rightmost column summarizes the statistical significance of the quasi-replication and depression lookup analyses of each SNP.

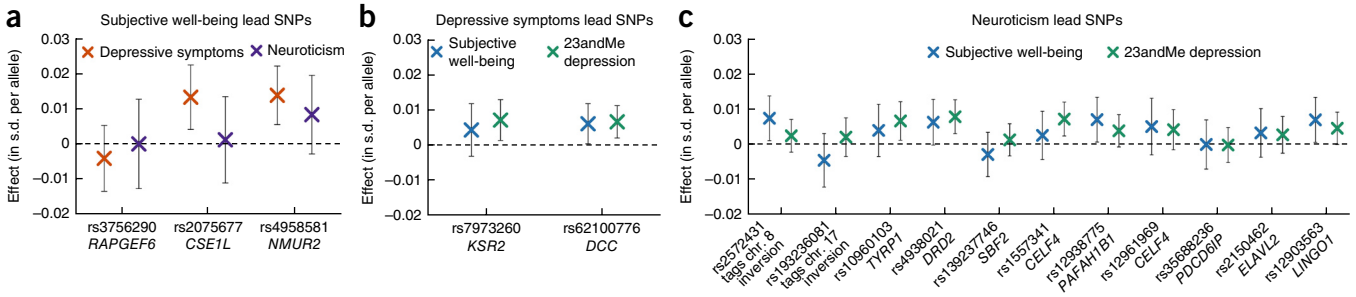


**Figure 3** Quasi-replication and lookup of lead SNPs. (**a**–**c**) In quasi-replication analyses, we examined whether lead SNPs identified in the subjective well-being meta-analyses are associated with depressive symptoms or neuroticism (**a**), whether lead SNPs identified in the analyses of depressive symptoms are associated with subjective well-being (**b**), and whether lead SNPs identified in the analyses of neuroticism are associated with subjective well-being (**c**). The quasi-replication sample was 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**. Error bars represent 95% confidence intervals (not adjusted for multiple testing). For ease of interpretation, we chose the reference allele such that positive coefficients imply that the estimated effect is in the predicted direction. Listed below each lead SNP is the nearest gene or chromosomal region.



## 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 ($P < 1 \times 10^{-4}$) 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 differed 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 were 163 candidate SNPs. Of these, 115 (71%) had the predicted direction of effect on depressive symptoms, 20 were significantly associated at the 5% significance level (19 in the predicted direction), and 2 remained significant after Bonferroni adjustment. For neuroticism, there were 170 candidate SNPs. Of these, 129 (76%) had the predicted direction of effect, all 28 SNPs significant at the 5% level had the predicted sign, and 4 of these remained significant after Bonferroni adjustment (**Supplementary Fig. 3** and **Supplementary Tables 16** and **17**). Two of the four were SNPs also identified in the proxy-phenotype analysis for depressive symptoms.

The four SNPs in total identified by the proxy-phenotype analyses are listed in **Table 1**.

## 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[10] to quantify the amount of genetic overlap between each of our three phenotypes and ten neuropsychiatric and physical health phenotypes. The estimates for subjective well-being and the negative of the estimates for depressive symptoms and neuroticism (as subjective well-being is negatively genetically correlated with depressive symptoms and neuroticism) are shown in **Figure 2b**,**c**. Subjective well-being, depressive symptoms, and neuroticism had strikingly similar patterns of pairwise genetic correlation with the other phenotypes.

We examined five neuropsychiatric phenotypes: Alzheimer disease, anxiety disorders, autism spectrum disorder, bipolar disorder, and schizophrenia (**Fig. 2b**). For four of these phenotypes, genetic correlations with depression (but not neuroticism or subjective well-being) were reported in Bulik-Sullivan *et al.*[10]. For schizophrenia and bipolar disorder, our estimated correlations with depressive symptoms, 0.33 and 0.26, are substantially lower than the point estimates of Bulik-Sullivan *et al.* but are contained within their 95% confidence intervals. By far, the largest genetic correlations we estimated were with anxiety disorders: −0.73 with subjective well-being, 0.88 with depressive symptoms, and 0.86 with neuroticism. To our knowledge, genetic correlations estimated from GWAS data have not previously been reported for anxiety disorders.

We also examined 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 levels, and triglyceride levels (**Fig. 2c**). The estimated genetic correlations were all small in magnitude, consistent with earlier work, although the greater precision of our estimates allowed us to reject null effects in most cases. The signs were generally consistent with those of the phenotypic correlations reported in earlier work between our phenotypes and outcomes such as obesity[20], smoking[21,22], and cardiovascular health[23].

Next, to investigate whether our GWAS results are enriched in particular functional categories, we applied stratified LD Score regression[24] 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 were broadly similar (**Supplementary Tables 18**–**20**) and are in line with what has been found for other phenotypes[24]. In our second analysis, the categories were 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 similar to each other and different from those for height, which was again included as a benchmark[25] (**Fig. 4a** and **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 have important roles in

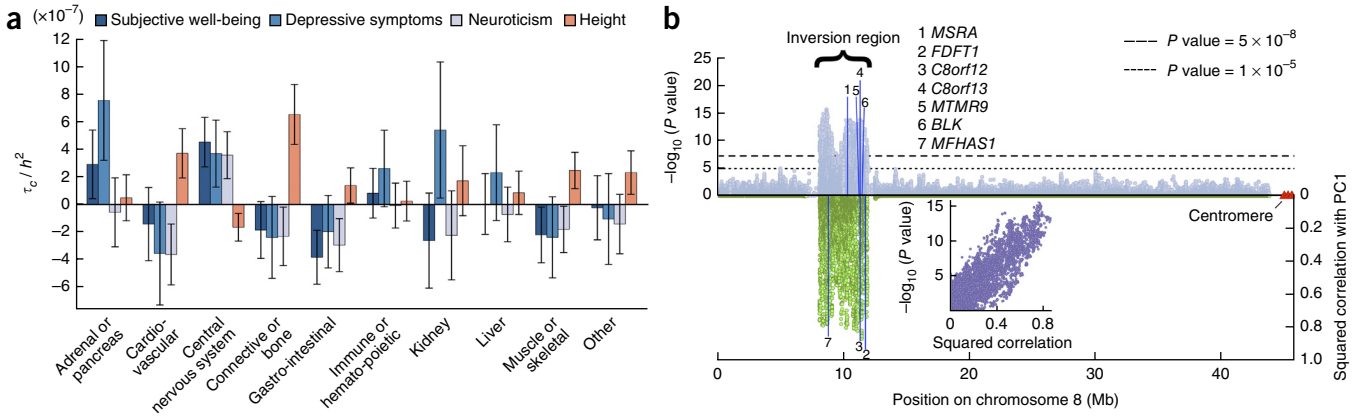


**Figure 4** Results from selected biological analyses. (**a**) Estimates of the expected increase in the phenotypic variance accounted for by a SNP due to the SNP's being in a given category ($\tau_c$), divided by the LD Score heritability of the phenotype ($h^2$). Each estimate of $\tau_c$ comes from a separate stratified LD Score regression, controlling for the 52 functional annotation categories in the baseline model. Error bars represent 95% confidence intervals (not adjusted for multiple testing). To benchmark the estimates, we compare them to those obtained from a recent study of height[25]. (**b**) Inversion polymorphism on chromosome 8 and the seven genes for which the inversion is a significant *cis*-eQTL at false discovery rate (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 (PC1). The inlay 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.

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 those in controls[26].

Whereas the above analyses use genome-wide data, we also conducted three analyses (Online Methods) restricted to the 16 GWAS and 4 proxy-phenotype SNPs in **Table 1**. In brief, we ascertained whether each SNP (or a variant in strong linkage disequilibrium (LD) with it) fell 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 expression quantitative trait locus (eQTL) in blood or in 1 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 5 of the 20 SNPs are in loci in which genome-wide significant associations have previously been reported. Two of these five are loci associated with schizophrenia. Interestingly, one of them harbors the gene *DRD2*, which encodes the $D_2$ subtype of the dopamine receptor, a target for antipsychotic drugs[27] that is also known to have a key role in neural reward pathways[28]. Motivated by these findings, as well as by the modest genetic correlations with schizophrenia (**Fig. 2b**), we examined whether the SNPs identified in a recent study of schizophrenia[29] 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-associated SNPs tended 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 emerged from our biological analyses is that the inversions on chromosomes 8 and 17 were 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[30] and triglyceride levels[31] (**Supplementary Table 22**). We also conducted eQTL analyses in blood for the inversion itself and found that it is a significant *cis*-eQTL for seven genes (**Supplementary Table 24**). All seven genes are positioned in close proximity to the inversion breakpoints (**Fig. 4b**), suggesting that the molecular mechanism underlying the inversion's effect on neuroticism could involve the relocation of regulatory sequences. Two of the genes (*MSRA* and *MTMR9*) are known to be highly expressed in tissues and cell types that belong to the nervous system, and two (*BLK* and *MFHAS1*) are known to be highly expressed 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* (in 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 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 blood and other tissue-specific analyses, encodes a protein important in the stabilization of microtubules in neurons. Associations have previously been reported between SNPs in *MAPT* (all of which are in strong LD with our inversion-tagging SNP) and neurodegenerative disorders, including Parkinson disease[32] and progressive supranuclear palsy[33], 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 previous studies because ours were conducted in larger samples. Our GWAS findings—3 loci associated with subjective well-being, 2 loci associated with depressive symptoms, and 11 loci associated with neuroticism—support the view that GWAS can successfully identify genetic associations with highly polygenic phenotypes in sufficiently large samples[5,34]. 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 additional loci associated 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 these findings (**Fig. 2b** and **Supplementary Table 15**).

We were able to assemble much larger samples than previous work in part because we combined data across heterogeneous phenotype measures. Our results reinforce the conclusions from our theoretical analysis that doing so increases 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[10] for our three phenotypes are quite low: 0.040 (s.e.m. = 0.002) for subjective well-being, 0.047 (s.e.m. = 0.004) for depressive symptoms, and 0.091 (s.e.m. = 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, to account for even a moderate share of 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 the largest posterior $R^2$ value was only ~13%—which in turn means it is likely that there exist many variants with effect sizes comparable to those for 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 perform meta-analysis on the 54 SNPs reaching $P < 1 \times 10^{-5}$ in our analyses of depressive symptoms together with the 23andMe replication sample for depression, the number of genome-wide significant associations increases from two to five (**Supplementary Table 15**).

**URLs.** Genotype-Tissue Expression Portal, http://www.GTExportal.org/; Social Science Genetic Association Consortium (SSGAC), http://www.thessgac.org/#!data/kuzq8.

## METHODS

Methods and any associated references are available in the online version of the paper.

**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.

*Note: Any Supplementary Information and Source Data files are available in the online version of the paper.*


### ACKNOWLEDGMENTS

We are grateful to P.M. Visscher for advice, support, and feedback. We thank S. Cunningham and N. Galla for research assistance. 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 GWAS 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 US 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. A full list of acknowledgments is provided in the **Supplementary Note**.


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

### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.




1. Kendler, K.S. & Myers, J. The genetic and environmental relationship between major depression and the five-factor model of personality. *Psychol. Med.* **40**, 801–806 (2010).
2. Weiss, A., Bates, T.C. & Luciano, M. Happiness is a personal(ity) thing: the genetics of personality and well-being in a representative sample. *Psychol. Sci.* **19**, 205–210 (2008).
3. Bartels, M., Cacioppo, J.T. & van Beijsterveldt, T.C.E.M. & Boomsma, D.I. Exploring the association between well-being and psychopathology in adolescents. *Behav. Genet.* **43**, 177–190 (2013).
4. Genetics of Personality Consortium. Meta-analysis of genome-wide association studies for neuroticism, and the polygenic association with major depressive disorder. *JAMA Psychiatry* **72**, 642–650 (2015).
5. Hyman, S. Mental health: depression needs large human-genetics studies. *Nature* **515**, 189–191 (2014).
6. Rietveld, C.A. *et al.* Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proc. Natl. Acad. Sci. USA* **111**, 13790–13794 (2014).
7. Kahneman, D. & Deaton, A. High income improves evaluation of life but not emotional well-being. *Proc. Natl. Acad. Sci. USA* **107**, 16489–16493 (2010).
8. Kahneman, D. & Riis, J. in *The Science of Well-Being* (eds. Uppter, F., Baylis, N. & Keverne, B.) 285–301 (Oxford University Press, 2005).
9. Bartels, M. & Boomsma, D.I. Born to be happy? The etiology of subjective well-being. *Behav. Genet.* **39**, 605–615 (2009).
10. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
11. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* **19**, 807–812 (2011).

12. Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol. Psychiatry* **18**, 497–511 (2013).
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.* **12**, e1001779 (2015).
14. Eysenck, H.J. & Eysenck, S.B.G. *Manual of the Eysenck Personality Questionnaire* (Hodder and Stroughton, 1975).
15. Tian, C. *et al.* Analysis and application of European genetic substructure using 300 K SNP information. *PLoS Genet.* **4**, e4 (2008).
16. Steinberg, K.M. *et al.* Structural diversity and African origin of the 17q21.31 inversion polymorphism. *Nat. Genet.* **44**, 872–880 (2012).
17. Smith, D.J. *et al.* Genome-wide analysis of over 106,000 individuals identifies 9 neuroticism-associated loci. *bioRxiv* doi:10.1101/032417 (2016).
18. Meuwissen, T.H., Hayes, B.J. & Goddard, M.E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819–1829 (2001).
19. Vilhjálmsson, B.J. *et al.* Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
20. Roberts, R.E., Kaplan, G.A., Shema, S.J. & Strawbridge, W.J. Are the obese at greater risk for depression? *Am. J. Epidemiol.* **152**, 163–170 (2000).
21. Glassman, A.H. *et al.* Smoking, smoking cessation, and major depression. *J. Am. Med. Assoc.* **264**, 1546–1549 (1990).
22. 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.* **121**, 38–44 (2012).
23. Rugulies, R. Depression as a predictor for coronary heart disease. a review and meta-analysis. *Am. J. Prev. Med.* **23**, 51–61 (2002).
24. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
25. Wood, A.R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014).
26. Stetler, C. & Miller, G.E. Depression and hypothalamic-pituitary-adrenal activation: a quantitative summary of four decades of research. *Psychosom. Med.* **73**, 114–126 (2011).
27. Seeman, P. Dopamine $D_2$ receptors as treatment targets in schizophrenia. *Clin. Schizophr. Relat. Psychoses* **4**, 56–73 (2010).
28. Vallone, D., Picetti, R. & Borrelli, E. Structure and function of dopamine receptors. *Neurosci. Biobehav. Rev.* **24**, 125–132 (2000).
29. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
30. Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187–196 (2015).
31. Kathiresan, S. *et al.* Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat. Genet.* **41**, 56–65 (2009).
32. UK Parkinson's Disease Consortium & Wellcome Trust Case Control Consortium 2. Dissection of the genetics of Parkinson's disease identifies an additional association 5′ of *SNCA* and multiple associated haplotypes at 17q21. *Hum. Mol. Genet.* **20**, 345–353 (2011).
33. Höglinger, G.U. *et al.* Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. *Nat. Genet.* **43**, 699–705 (2011).
34. Sullivan, P. Don't give up on GWAS. *Mol. Psychiatry* **17**, 2–3 (2012).

Aysu Okbay[1–3,146], Bart M L Baselmans[4,5,146], Jan-Emmanuel De Neve[6,146], Patrick Turley[7,146], Michel G Nivard[4,146], Mark Alan Fontana[8,146], S Fleur W Meddens[3,9,10,146], Richard Karlsson Linnér[3,9,10,146], Cornelius A Rietveld[1–3,146], Jaime Derringer[11], Jacob Gratten[12], James J Lee[13], Jimmy Z Liu[14], Ronald de Vlaming[1–3], Tarunveer S Ahluwalia[15–17], Jadwiga Buchwald[18], Alana Cavadino[19,20], Alexis C Frazier-Wood[21], Nicholas A Furlotte[22], Victoria Garfield[23], Marie Henrike Geisel[24], Juan R Gonzalez[25–27], Saskia Haitjema[28], Robert Karlsson[29], Sander W van der Laan[28], Karl-Heinz Ladwig[30,31], Jari Lahti[32–34], Sven J van der Lee[2], Penelope A Lind[35], Tian Liu[36,37], Lindsay Matteson[13], Evelin Mihailov[38], Michael B Miller[13], Camelia C Minica[4], Ilja M Nolte[39], Dennis Mook-Kanamori[40–42], Peter J van der Most[39], Christopher Oldmeadow[43,44], Yong Qian[45], Olli Raitakari[46,47], Rajesh Rawal[48], Anu Realo[49,50], Rico Rueedi[51,52], Börge Schmidt[24], Albert V Smith[53,54], Evie Stergiakouli[55], Toshiko Tanaka[56], Kent Taylor[57], Gudmar Thorleifsson[58], Juho Wedenoja[18], Juergen Wellmann[59], Harm-Jan Westra[60,61], Sara M Willems[2], Wei Zhao[62], LifeLines Cohort Study[63], Najaf Amin[2], Andrew Bakshi[12], Sven Bergmann[51,52], Gyda Bjornsdottir[58], Patricia A Boyle[64], Samantha Cherney[65], Simon R Cox[66,67], Gail Davies[66,67], Oliver S P Davis[55], Jun Ding[45], Nese Direk[2], Peter Eibich[68,69], Rebecca T Emeny[30,70], Ghazaleh Fatemifar[71], Jessica D Faul[72], Luigi Ferrucci[56], Andreas J Forstner[73,74], Christian Gieger[48], Richa Gupta[18], Tamara B Harris[75], Juliette M Harris[76], Elizabeth G Holliday[43,44], Jouke-Jan Hottenga[4,5], Philip L De Jager[77–79], Marika A Kaakinen[80,81], Eero Kajantie[82,83], Ville Karhunen[81], Ivana Kolcic[84], Meena Kumari[85], Lenore J Launer[86], Lude Franke[87], Ruifang Li-Gao[40], David C Liewald[66], Marisa Koini[88], Anu Loukola[18], Pedro Marques-Vidal[89], Grant W Montgomery[90], Miriam A Mosing[91], Lavinia Paternoster[55], Alison Pattie[67], Katja E Petrovic[88], Laura Pulkki-Råback[32,34], Lydia Quaye[76], Katri Räikkönen[32], Igor Rudan[92], Rodney J Scott[44,93], Jennifer A Smith[62], Angelina R Sutin[56,94], Maciej Trzaskowski[12], Anna E Vinkhuyzen[12], Lei Yu[95], Delilah Zabaneh[84], John R Attia[43,44], David A Bennett[95], Klaus Berger[59], Lars Bertram[96,97], Dorret I Boomsma[4,5,98], Harold Snieder[39], Shun-Chiao Chang[99], Francesco Cucca[100], Ian J Deary[66,67], Cornelia M van Duijn[2], Johan G Eriksson[101–103], Ute Bültmann[104], Eco J C de Geus[4,5,98], Patrick J F Groenen[3,105], Vilmundur Gudnason[53,54], Torben Hansen[16], Catharine A Hartman[106], Claire M A Haworth[55], Caroline Hayward[107], Andrew C Heath[108], David A Hinds[22], Elina Hyppönen[20,109,110], William G Iacono[13], Marjo-Riitta Järvelin[80,111–113], Karl-Heinz Jöckel[24], Jaakko Kaprio[18,114,115], Sharon L R Kardia[62], Liisa Keltikangas-Järvinen[32], Peter Kraft[116], Laura D Kubzansky[117], Terho Lehtimäki[118,119], Patrik K E Magnusson[29], Nicholas G Martin[120], Matt McGue[13], Andres Metspalu[38,121], Melinda Mills[122], Renée de Mutsert[40], Albertine J Oldehinkel[105], Gerard Pasterkamp[28,123], Nancy L Pedersen[29], Robert Plomin[124], Ozren Polasek[84], Christine Power[20,110], Stephen S Rich[125], Frits R Rosendaal[40], Hester M den Ruijter[28], David Schlessinger[45], Helena Schmidt[88,126], Rauli Svento[127], Reinhold Schmidt[88], Behrooz Z Alizadeh[39,128], Thorkild I A Sørensen[16,55,129], Tim D Spector[76], John M Starr[66,130], Kari Stefansson[54,58], Andrew Steptoe[23], Antonio Terracciano[56,94], Unnur Thorsteinsdottir[54,58], A Roy Thurik[1,3,131,132], Nicholas J Timpson[55], Henning Tiemeier[2,133,134], André G Uitterlinden[2,3,135], Peter Vollenweider[89], Gert G Wagner[36,68,136], David R Weir[72], Jian Yang[12,137], Dalton C Conley[138], George Davey Smith[55], Albert Hofman[2,139], Magnus Johannesson[140], David I Laibson[7], Sarah E Medland[35], Michelle N Meyer[141,142], Joseph K Pickrell[14,143], Tõnu Esko[38], Robert F Krueger[13,147], Jonathan P Beauchamp[7,147], Philipp D Koellinger[3,9,10,147], Daniel J Benjamin[8,147], Meike Bartels[4,5,98,147] & David Cesarini[144,145,147]

[1]Department of Applied Economics, Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, the Netherlands. [2]Department of Epidemiology, Erasmus Medical Center, Rotterdam, the Netherlands. [3]Erasmus University Rotterdam Institute for Behavior and Biology, Rotterdam, the Netherlands. [4]Department of Biological Psychology, Vrije Universiteit, Amsterdam, the Netherlands. [5]EMGO+ Institute for Health and Care Research, Amsterdam, the Netherlands. [6]Saïd Business School, University of Oxford, Oxford, UK. [7]Department of Economics, Harvard University, Cambridge, Massachusetts, USA. [8]Center for Economic and Social Research, University of Southern California, Los Angeles, California, USA. [9]Department of Complex Trait Genetics, Vrije Universiteit, Center for Neurogenomics and Cognitive Research, Amsterdam, the Netherlands. [10]Amsterdam Business School, University of Amsterdam, Amsterdam, the Netherlands. [11]Department of Psychology, University of Illinois, Champaign, Illinois, USA. [12]Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia. [13]Department of Psychology, University of Minnesota Twin Cities, Minneapolis, Minnesota, USA. [14]New York Genome Center, New York, New York, USA. [15]COPSAC (Copenhagen Prospective Studies on Asthma in Childhood), Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark. [16]Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, University of Copenhagen, Faculty of Health and Medical Sciences, Copenhagen, Denmark. [17]Steno Diabetes Center, Gentofte, Denmark. [18]Department of Public Health, University of Helsinki, Helsinki, Finland. [19]Centre for Environmental and Preventive Medicine, Wolfson Institute of Preventive Medicine, Queen Mary University of London, London, UK. [20]South Australian Health and Medical Research Institute, Adelaide, South Australia, Australia. [21]USDA–ARS Children's Nutrition Research Center, Baylor College of Medicine, Houston, Texas, USA. [22]23andMe, Inc., Mountain View, California, USA. [23]Department of Epidemiology and Public Health, University College London, London, UK. [24]Institute for Medical Informatics, Biometry and Epidemiology, University Hospital of Essen, Essen, Germany. [25]Centre for Research in Environmental Epidemiology, Institute for Global Health, Barcelona, Spain. [26]Universitat Pompeu Fabra, Barcelona, Spain. [27]CIBER Epidemiología y Salud Pública, Barcelona, Spain. [28]Laboratory of Experimental Cardiology, University Medical Center Utrecht, Utrecht, the Netherlands. [29]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [30]Institute of Epidemiology II, Mental Health Research Unit, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [31]Department of Psychosomatic Medicine

and Psychotherapy, Klinikum Rechts der Isar, Technische Universität München, Munich, Germany. [32]Institute of Behavioural Sciences, University of Helsinki, Helsinki, Finland. [33]Folkhälsan Research Centre, Helsingfors, Finland. [34]Helsinki Collegium for Advanced Studies, University of Helsinki, Helsinki, Finland. [35]Psychiatric Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [36]Max Planck Institute for Human Development, Berlin, Germany. [37]Max Planck Institute for Molecular Genetics, Department of Vertebrate Genomics, Berlin, Germany. [38]Estonian Genome Center, University of Tartu, Tartu, Estonia. [39]Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [40]Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands. [41]Public Health and Primary Care, Leiden University Medical Center, Leiden, the Netherlands. [42]BESC, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia. [43]Public Health Stream, Hunter Medical Research Institute, New Lambton, New South Wales, Australia. [44]Faculty of Health and Medicine, University of Newcastle, Newcastle, New South Wales, Australia. [45]Laboratory of Genetics, National Institute on Aging, Baltimore, Maryland, USA. [46]Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland. [47]Department of Clinical Physiology, Turku University Hospital, Turku, Finland. [48]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [49]Department of Psychology, University of Tartu, Tartu, Estonia. [50]Department of Psychology, University of Warwick, Coventry, UK. [51]Department of Computational Biology, University of Lausanne, Lausanne, Switzerland. [52]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [53]Icelandic Heart Association, Kopavogur, Iceland. [54]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [55]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK. [56]National Institute on Aging, US National Institutes of Health, Baltimore, Maryland, USA. [57]Los Angeles Biomedical Research Institute and Department of Pediatrics, Harbor-UCLA, Torrence, California, USA. [58]deCODE Genetics/Amgen, Inc., Reykjavik, Iceland. [59]Institute of Epidemiology and Social Medicine, University of Münster, Münster, Germany. [60]Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA. [61]Division of Rheumatology, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA. [62]Department of Epidemiology, University of Michigan, Ann Arbor, Michigan, USA. [63]University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [64]Department of Behavioral Sciences, Rush University Medical Center, Chicago, Illinois, USA. [65]RAND Corporation, Santa Monica, California, USA. [66]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK. [67]Department of Psychology, University of Edinburgh, Edinburgh, UK. [68]German Socio-Economic Panel Study, DIW Berlin, Berlin, Germany. [69]Health Economics Research Centre, Nuffield Department of Population Health, University of Oxford, Oxford, UK. [70]Department of Epidemiology, Geisel School of Medicine at Dartmouth, Lebanon, New Hampshire, USA. [71]Farr Institute of Health Informatics, University College London, London, UK. [72]Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, Michigan, USA. [73]Institute of Human Genetics, University of Bonn, Bonn, Germany. [74]Department of Genomics, Life and Brain Center, University of Bonn, Bonn, Germany. [75]Laboratory of Epidemiology, Demography, National Institute on Aging, US National Institutes of Health, Bethesda, Maryland, USA. [76]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. [77]Program in Translational NeuroPsychiatric Genomics, Departments of Neurology and Psychiatry, Brigham and Women's Hospital, Boston, Massachusetts, USA. [78]Harvard Medical School, Boston, Massachusetts, USA. [79]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA. [80]Department of Genomics of Common Disease, Imperial College London, London, UK. [81]Center for Life Course Health Research, University of Oulu and Oulu University Hospital, Oulu, Finland. [82]Department of Pediatrics, University of Helsinki, Helsinki, Finland. [83]National Institute for Health and Welfare, Helsinki, Finland. [84]Department of Public Health, Faculty of Medicine, University of Split, Split, Croatia. [85]Institute for Social and Economic Research, University of Essex, Wivenhoe Park, UK. [86]Neuroepidemiology Section, National Institute on Aging, US National Institutes of Health, Bethesda, Maryland, USA. [87]Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands. [88]Department of Neurology, General Hospital and Medical University Graz, Graz, Austria. [89]Department of Internal Medicine, Internal Medicine, Lausanne University Hospital (CHUV), Lausanne, Switzerland. [90]Molecular Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [91]Department of Neuroscience, Karolinska Institutet, Stockholm, Sweden. [92]Centre for Global Health Research, Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK. [93]Information-Based Medicine Stream, Hunter Medical Research Institute, New Lambton, New South Wales, Australia. [94]Department of Behavioral Sciences and Social Medicine, Florida State University College of Medicine, Tallahassee, Florida, USA. [95]Department of Neurological Sciences, Rush University Medical Center, Chicago, Illinois, USA. [96]Lübeck Interdisciplinary Platform for Genome Analytics (LIGA), Institute of Neurogenetics and Institute of Integrative and Experimental Genomics, University of Lübeck, Lübeck, Germany. [97]Neuroepidemiology and Ageing Research Unit, School of Public Health, Faculty of Medicine, Imperial College London, London, UK. [98]Neuroscience Campus Amsterdam, Amsterdam, the Netherlands. [99]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. [100]Istituto di Ricerca Genetica e Biomedica (IRGB), Consiglio Nazionale delle Ricerche, Cittadella Universitarià di Monserrato, Monserrato, Italy. [101]Department of General Practice and Primary Health Care, University of Helsinki, Helsinki, Finland. [102]Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland. [103]Unit of General Practice, University Central Hospital, Helsinki, Finland. [104]Department of Health Sciences, Community and Occupational Medicine, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [105]Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, the Netherlands. [106]Department of Psychiatry, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [107]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. [108]Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri, USA. [109]Centre for Population Health Research, School of Health Sciences and Sansom Institute, University of South Australia, Adelaide, South Australia, Australia. [110]Population, Policy and Practice, UCL Institute of Child Health, London, UK. [111]Department of Epidemiology and Biostatistics, MRC–PHE Centre for Environment and Health, School of Public Health, Imperial College London, London, UK. [112]Biocenter Oulu, University of Oulu, Oulu, Finland. [113]Unit of Primary Care, Oulu University Hospital, Oulu, Finland. [114]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland. [115]Department for Health, THL (National Institute for Health and Welfare), Helsinki, Finland. [116]Department of Epidemiology and Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA. [117]Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA. [118]Fimlab Laboratories, Tampere, Finland. [119]Department of Clinical Chemistry, University of Tampere, School of Medicine, Tampere, Finland. [120]Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [121]Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia. [122]Department of Sociology, University of Oxford, Oxford, UK. [123]Laboratory of Clinical Chemistry and Hematology, Division of Laboratories and Pharmacy, University Medical Center Utrecht, Utrecht, the Netherlands. [124]Social, Genetic and Developmental Psychiatry Centre, King's College London, De Crespigny Park, UK. [125]Department of Public Health Sciences, University of Virginia, Charlottesville, Virginia, USA. [126]Research Unit for Genetic Epidemiology, Institute of Molecular Biology and Biochemistry, Center of Molecular Medicine, General Hospital and Medical University Graz, Graz, Austria. [127]Department of Economics, Oulu Business School, Oulu, Finland. [128]Department of Gastroenterology and Hepatology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [129]Institute of Preventive Medicine, Bispebjerg and Frederiksberg Hospitals, Capital Region, Frederiksberg, Denmark. [130]Alzheimer Scotland Dementia Research Centre, University of Edinburgh, Edinburgh, UK. [131]Montpellier Business School, Montpellier, France. [132]Panteia, Zoetermeer, the Netherlands. [133]Department of Psychiatry, Erasmus Medical Center, Rotterdam, the Netherlands. [134]Department of Child and Adolescent Psychiatry, Erasmus Medical Center, Rotterdam, the Netherlands. [135]Department of Internal Medicine, Erasmus Medical Center, Rotterdam, the Netherlands. [136]School of Economics and Management, Berlin University of Technology, Berlin, Germany. [137]University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Queensland, Australia. [138]Department of Sociology, Princeton University, Princeton, New Jersey, USA. [139]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA. [140]Department of Economics, Stockholm School of Economics, Stockholm, Sweden. [141]Department of Bioethics, Clarkson University, Schenectady, New York, USA. [142]Icahn School of Medicine at Mount Sinai, New York, New York, USA. [143]Department of Biological Sciences, Columbia University, New York, New York, USA. [144]Department of Economics, New York University, New York, New York, USA. [145]Research Institute for Industrial Economics, Stockholm, Sweden. [146]These authors contributed equally to this work. [147]These authors jointly directed this work. Correspondence should be addressed to D.J.B. (djbenjam@usc.edu), M.B. (m.bartels@vu.nl) or P.D.K. (p.d.koellinger@vu.nl).

## ONLINE METHODS

**GWAS of subjective well-being.** Genome-wide association analyses were performed at the cohort level according to a prespecified 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 HapMap 2 CEU (r22.b36) reference sample[35]. We performed meta-analysis on 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 procedures were applied to the cohort-level summary statistics, including but not limited to the EasyQC[36] 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 an LD Score regression[10]. 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 1000 Genomes Project–imputed data ($n$ = 229,883) (**Supplementary Figs. 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**. Association results from the following four meta-analyses are reported: the primary subjective well-being analysis, the life satisfaction analysis, the positive affect analysis, and the *post hoc* subjective well-being analysis (**Supplementary Table 7**). For each phenotype, we provide association results for the set of approximately independent SNPs that attained a $P$ value smaller than $1 \times 10^{-5}$. We identify these SNP using the same clumping algorithm as for the lead SNPs but with the $P$-value threshold set at $1 \times 10^{-5}$ instead of genome-wide significance.

**GWAS of depressive symptoms and neuroticism.** Our auxiliary genome-wide association studies of depressive symptoms and neuroticism were conducted in 1000 Genomes Project–imputed data, combining new genome-wide association analyses with publicly available summary statistics from previously published studies. We applied a similar quality control protocol to that used in our primary subjective well-being analysis. In the depressive symptoms 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 ($n$ = 170,911), 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**. Quantile–quantile plots of the neuroticism and depressive symptoms meta-analysis results are shown in **Supplementary Figure 7**. Association results for the set of approximately independent SNPs that attained a $P$ value smaller than $1 \times 10^{-5}$ 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[10]. 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 estimates, the coefficients from regressing the difference in phenotype across siblings on the difference in siblings' genotype (and controls). These within-family 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 performed meta-analysis on association statistics from within-family analyses conducted in four cohorts.

In the first analysis, we estimated the fraction of SNPs for which the signs of the within-family 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 nearly perfectly matches the expected rate after adjustment for winner's curse at 50.83% (**Supplementary Fig. 9**).

The second analysis used the within-family regression coefficient estimates (that is, not only their signs) to estimate the amount of stratification bias. For each SNP $j$, let $\hat{\beta}_j$ denote the GWAS estimate and let $\hat{\beta}_{\mathrm{WF},j}$ denote the within-family 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

$$\hat{\beta}_j = \beta_j + s_j + U_j$$
$$\hat{\beta}_{\mathrm{WF},j} = \beta_j + V_j$$

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 $\mathrm{E}(U_j) = \mathrm{E}(V_j) = 0$. Whenever $s_j \neq 0$, the GWAS estimate of $\hat{\beta}_j$ 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

$$\frac{\mathrm{var}(\beta_j)}{\mathrm{var}(\beta_j) + \mathrm{var}(s_j)}$$

In the **Supplementary Note**, we show that, with estimates $\hat{\beta}_j$ and $\hat{\beta}_{\mathrm{WF},j}$ (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 principal-components analysis (PCA)–mixture method. For both inversions, the method clearly distinguishes three clusters of genotypes, corresponding to inversion genotypes (**Supplementary Fig. 10**). We validated the PCA–mixture procedure using existing methods designed to call inversion genotypes[37] (**Supplementary Table 26**).

For each inversion, we established that the inversion-tagging SNPs were always located in close proximity to the inversion region (**Fig. 3b** and **Supplementary Figs. 10** and **11**). We list the 20 variants that most strongly correlate with the principal components that capture the inversion polymorphisms on chromosomes 8 (**Supplementary Table 27**) and 17 (**Supplementary Table 28**). In additional analyses, we confirmed that the inversion is associated with neuroticism and subjective well-being in independent cohorts (**Supplementary Figs. 12** and **13** and **Supplementary Tables 29** and **30**).

**Proxy-phenotype analyses.** In these analyses, we used a two-stage approach that has been successfully applied in other contexts[6]. 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 $1 \times 10^{-4}$. 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 omitted cohorts from the first or second stage as needed to ensure that the samples in the two stages were non-overlapping. The cohort restrictions imposed are listed in **Supplementary Table 31**. These cohort restrictions, as well as the $P$-value threshold of $1 \times 10^{-4}$, 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 lead SNPs are more strongly associated with the




doi:10.1038/ng.3552

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) were significantly associated with the second-stage phenotype at the Bonferroni-adjusted significance level of 0.05/number of SNPs tested.

**Genetic correlations.** We used bivariate LD Score regression[10] 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 (s.e.m. = 0.065) between life satisfaction and positive affect, 0.897 (s.e.m. = 0.017) between well-being (our measure that combines life satisfaction and positive affect) and life satisfaction, and 1.031 (s.e.m. = 0.019) between positive affect and well-being. For depressive symptoms, we estimated a genetic correlation of 0.588 (s.e.m. = 0.242) between GERA and PGC, 0.972 (s.e.m. = 0.216) between GERA and UKB, and 0.797 (s.e.m. = 0.108) between UKB and PGC. Finally, we estimated a genetic correlation of 1.11 (s.e.m. = 0.14) between the measures of neuroticism in the UKB analyses and the summary statistics from a previously published meta-analysis[4].

**Bayesian credibility analyses.** To evaluate the credibility of our findings, we use a standard Bayesian framework[18] 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 $j \in$ {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 $\pi \in \{0.01, 0.02, \ldots, 0.99\}$. For each assumed value of $\pi$, we estimate $\tau^2_{\text{SWB}}$ 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 $\pi$. For all considered values of $\pi$ and all three SNPs, the posterior probability that the SNP is null is below 1% (**Supplementary Fig. 14**). 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[10,38] that the variance in a SNP's effect size is inversely proportional to the variance of its genotype, $2 \times \text{MAF} \times (1 - \text{MAF})$, where MAF is the minor allele frequency.

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, $P\left(C \mid \hat{\beta}_1\right)$, or the probability that the SNP is non-null conditional on the results of both the first-stage GWAS and the quasi-replication analysis, $P\left(C \mid \hat{\beta}_1, \hat{\beta}_2\right)$. 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

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

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 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 SNPs associated with depressive symptoms and neuroticism 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 because these were both conducted in samples that overlap with the 23andMe's depression sample. For details on association models, quality control filters, and the ascertainment of depression status, we refer to the companion study (C.L. Hyde, M.W. Nagle, C. Tian, X. Chen, S.A. Paciga *et al.* unpublished data). The $P$ values we report are based on standard errors that have been adjusted using the intercept from an LD Score regression[10].

**Polygenic prediction.** To evaluate the predictive power of a polygenic score derived from the subjective well-being meta-analysis results, we used two independent withheld cohorts: the Health and Retirement Study (HRS)[39] and the Netherlands Twin Register (NTR)[40,41]. To generate the weights for the polygenic score, we performed meta-analyses of the pooled subjective well-being phenotype excluding each of the withheld 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 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 to be 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 250 kb 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[42] to determine which of our 20 SNPs (and their LD partners) were in LD with SNPs for which genome-wide

significant associations have previously been reported. Because the GWAS catalog does not always include the most recent GWAS results available, we included additional recent GWAS studies. We used the tool HaploReg[43] 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[40,41,44]. We also performed eQTL lookups of our 20 SNPs in the Genotype-Tissue Expression Portal[45,46]. 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 coexpression database[47], we explored the predicted functions of genes colocating with the 20 SNPs in **Table 1** (**Supplementary Table 35**).

Further details appear in a **Supplementary Note**.


35. International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
36. Winkler, T.W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9**, 1192–1212 (2014).
37. Cáceres, A. & González, J.R. Following the footprints of polymorphic inversions on SNP data: from detection to association tests. *Nucleic Acids Res.* **43**, e53 (2015).
38. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).
39. Sonnega, A. *et al.* Cohort profile: the Health and Retirement Study (HRS). *Int. J. Epidemiol.* **43**, 576–585 (2014).
40. Willemsen, G. *et al.* The Adult Netherlands Twin Register: twenty-five years of survey and biological data collection. *Twin Res. Hum. Genet.* **16**, 271–281 (2013).
41. van Beijsterveldt, C.E.M. *et al.* The Young Netherlands Twin Register (YNTR): longitudinal twin and family studies in over 70,000 children. *Twin Res. Hum. Genet.* **16**, 252–267 (2013).
42. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP–trait associations. *Nucleic Acids Res.* **42**, D1001–D1006 (2014).
43. Ward, L.D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40**, D930–D934 (2012).
44. Penninx, B.W.J.H. *et al.* The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. *Int. J. Methods Psychiatr. Res.* **17**, 121–140 (2008).
45. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
46. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
47. Fehrmann, R.S.N. *et al.* Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat. Genet.* **47**, 115–125 (2015).





doi:10.1038/ng.3552

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

**Aysu Okbay, Bart M L Baselmans, Jan-Emmanuel De Neve, Patrick Turley, Michel G Nivard, Mark Alan Fontana, S Fleur W Meddens, Richard Karlsson Linnér, Cornelius A Rietveld, Jaime Derringer, Jacob Gratten, James J Lee, Jimmy Z Liu, Ronald de Vlaming, Tarunveer S Ahluwalia, Jadwiga Buchwald, Alana Cavadino, Alexis C Frazier-Wood, Nicholas A Furlotte, Victoria Garfield, Marie Henrike Geisel, Juan R Gonzalez, Saskia Haitjema, Robert Karlsson, Sander W van der Laan, Karl-Heinz Ladwig, Jari Lahti, Sven J van der Lee, Penelope A Lind, Tian Liu, Lindsay Matteson, Evelin Mihailov, Michael B Miller, Camelia C Minica, Ilja M Nolte, Dennis Mook-Kanamori, Peter J van der Most, Christopher Oldmeadow, Yong Qian, Olli Raitakari, Rajesh Rawal, Anu Realo, Rico Rueedi, Börge Schmidt, Albert V Smith, Evie Stergiakouli, Toshiko Tanaka, Kent Taylor, Gudmar Thorleifsson, Juho Wedenoja, Juergen Wellmann, Harm-Jan Westra, Sara M Willems, Wei Zhao, LifeLines Cohort Study, Najaf Amin, Andrew Bakshi, Sven Bergmann, Gyda Bjornsdottir, Patricia A Boyle, Samantha Cherney, Simon R Cox, Gail Davies, Oliver S P Davis, Jun Ding, Nese Direk, Peter Eibich, Rebecca T Emeny, Ghazaleh Fatemifar, Jessica D Faul, Luigi Ferrucci, Andreas J Forstner, Christian Gieger, Richa Gupta, Tamara B Harris, Juliette M Harris, Elizabeth G Holliday, Jouke-Jan Hottenga, Philip L De Jager, Marika A Kaakinen, Eero Kajantie, Ville Karhunen, Ivana Kolcic, Meena Kumari, Lenore J Launer, Lude Franke, Ruifang Li-Gao, David C Liewald, Marisa Koini, Anu Loukola, Pedro Marques-Vidal, Grant W Montgomery, Miriam A Mosing, Lavinia Paternoster, Alison Pattie, Katja E Petrovic, Laura Pulkki-Råback, Lydia Quaye, Katri Räikkönen, Igor Rudan, Rodney J Scott, Jennifer A Smith, Angelina R Sutin, Maciej Trzaskowski, Anna E Vinkhuyzen, Lei Yu, Delilah Zabaneh, John R Attia, David A Bennett, Klaus Berger, Lars Bertram, Dorret I Boomsma, Harold Snieder, Shun-Chiao Chang, Francesco Cucca, Ian J Deary, Cornelia M van Duijn, Johan G Eriksson, Ute Bültmann, Eco J C de Geus, Patrick J F Groenen, Vilmundur Gudnason, Torben Hansen, Catharine A Hartman, Claire M A Haworth, Caroline Hayward, Andrew C Heath, David A Hinds, Elina Hyppönen, William G Iacono, Marjo-Riitta Järvelin, Karl-Heinz Jöckel, Jaakko Kaprio, Sharon L R Kardia, Liisa Keltikangas-Järvinen, Peter Kraft, Laura D Kubzansky, Terho Lehtimäki, Patrik K E Magnusson, Nicholas G Martin, Matt McGue, Andres Metspalu, Melinda Mills, Renée de Mutsert, Albertine J Oldehinkel, Gerard Pasterkamp, Nancy L Pedersen, Robert Plomin, Ozren Polasek, Christine Power, Stephen S Rich, Frits R Rosendaal, Hester M den Ruijter, David Schlessinger, Helena Schmidt, Rauli Svento, Reinhold Schmidt, Behrooz Z Alizadeh, Thorkild I A Sørensen, Tim D Spector, John M Starr, Kari Stefansson, Andrew Steptoe, Antonio Terracciano, Unnur Thorsteinsdottir, A Roy Thurik, Nicholas J Timpson, Henning Tiemeier, André G Uitterlinden, Peter Vollenweider, Gert G Wagner, David R Weir, Jian Yang, Dalton C Conley, George Davey Smith, Albert Hofman, Magnus Johannesson, David I Laibson, Sarah E Medland, Michelle N Meyer, Joseph K Pickrell, Tõnu Esko, Robert F Krueger, Jonathan P Beauchamp, Philipp D Koellinger, Daniel J Benjamin, Meike Bartels & David Cesarini**
***Nat. Genet.* 48, 624–633 (2016); published online 18 April 2016; corrected after print 27 June 2016**

In the version of this article initially published, the following co-authors and their affiliations were incorrectly omitted from the author list: Gudmar Thorleifsson, Sven Bergmann, Gyda Bjornsdottir, David C. Liewald, John M. Starr, Kari Stefansson and Unnur Thorsteinsdottir. In addition, the middle initial for co-author Andreas J. Forstner was also omitted. The errors have been corrected in the HTML and PDF versions of the article.

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

**Aysu Okbay, Bart M L Baselmans, Jan-Emmanuel De Neve, Patrick Turley, Michel G Nivard, Mark Alan Fontana, S Fleur W Meddens, Richard Karlsson Linnér, Cornelius A Rietveld, Jaime Derringer, Jacob Gratten, James J Lee, Jimmy Z Liu, Ronald de Vlaming, Tarunveer S Ahluwalia, Jadwiga Buchwald, Alana Cavadino, Alexis C Frazier-Wood, Nicholas A Furlotte, Victoria Garfield, Marie Henrike Geisel, Juan R Gonzalez, Saskia Haitjema, Robert Karlsson, Sander W van der Laan, Karl-Heinz Ladwig, Jari Lahti, Sven J van der Lee, Penelope A Lind, Tian Liu, Lindsay Matteson, Evelin Mihailov, Michael B Miller, Camelia C Minica, Ilja M Nolte, Dennis Mook-Kanamori, Peter J van der Most, Christopher Oldmeadow, Yong Qian, Olli Raitakari, Rajesh Rawal, Anu Realo, Rico Rueedi, Börge Schmidt, Albert V Smith, Evie Stergiakouli, Toshiko Tanaka, Kent Taylor, Gudmar Thorleifsson, Juho Wedenoja, Juergen Wellmann, Harm-Jan Westra, Sara M Willems, Wei Zhao, LifeLines Cohort Study, Najaf Amin, Andrew Bakshi, Sven Bergmann, Gyda Bjornsdottir, Patricia A Boyle, Samantha Cherney, Simon R Cox, Gail Davies, Oliver S P Davis, Jun Ding, Nese Direk, Peter Eibich, Rebecca T Emeny, Ghazaleh Fatemifar, Jessica D Faul, Luigi Ferrucci, Andreas J Forstner, Christian Gieger, Richa Gupta, Tamara B Harris, Juliette M Harris, Elizabeth G Holliday, Jouke-Jan Hottenga, Philip L De Jager, Marika A Kaakinen, Eero Kajantie, Ville Karhunen, Ivana Kolcic, Meena Kumari, Lenore J Launer, Lude Franke, Ruifang Li-Gao, David C Liewald, Marisa Koini, Anu Loukola, Pedro Marques-Vidal, Grant W Montgomery, Miriam A Mosing, Lavinia Paternoster, Alison Pattie, Katja E Petrovic, Laura Pulkki-Råback, Lydia Quaye, Katri Räikkönen, Igor Rudan, Rodney J Scott, Jennifer A Smith, Angelina R Sutin, Maciej Trzaskowski, Anna E Vinkhuyzen, Lei Yu, Delilah Zabaneh, John R Attia, David A Bennett, Klaus Berger, Lars Bertram, Dorret I Boomsma, Harold Snieder, Shun-Chiao Chang, Francesco Cucca, Ian J Deary, Cornelia M van Duijn, Johan G Eriksson, Ute Bültmann, Eco J C de Geus, Patrick J F Groenen, Vilmundur Gudnason, Torben Hansen, Catharine A Hartman, Claire M A Haworth, Caroline Hayward, Andrew C Heath, David A Hinds, Elina Hyppönen, William G Iacono, Marjo-Riitta Järvelin, Karl-Heinz Jöckel, Jaakko Kaprio, Sharon L R Kardia, Liisa Keltikangas-Järvinen, Peter Kraft, Laura D Kubzansky, Terho Lehtimäki, Patrik K E Magnusson, Nicholas G Martin, Matt McGue, Andres Metspalu, Melinda Mills, Renée de Mutsert, Albertine J Oldehinkel, Gerard Pasterkamp, Nancy L Pedersen, Robert Plomin, Ozren Polasek, Christine Power, Stephen S Rich, Frits R Rosendaal, Hester M den Ruijter, David Schlessinger, Helena Schmidt, Rauli Svento, Reinhold Schmidt, Behrooz Z Alizadeh, Thorkild I A Sørensen, Tim D Spector, John M Starr, Kari Stefansson, Andrew Steptoe, Antonio Terracciano, Unnur Thorsteinsdottir, A Roy Thurik, Nicholas J Timpson, Henning Tiemeier, André G Uitterlinden, Peter Vollenweider, Gert G Wagner, David R Weir, Jian Yang, Dalton C Conley, George Davey Smith, Albert Hofman, Magnus Johannesson, David I Laibson, Sarah E Medland, Michelle N Meyer, Joseph K Pickrell, Tõnu Esko, Robert F Krueger, Jonathan P Beauchamp, Philipp D Koellinger, Daniel J Benjamin, Meike Bartels & David Cesarini**
***Nat. Genet.* 48, 624–633 (2016); published online 18 April 2016; corrected after print 27 June 2016; corrected after print 29 August 2016**

In the version of this article initially published, one of the affiliations listed for author Maciej Trzaskowski, to the Department of Public Health, Faculty of Medicine, University of Split, Split, Croatia, was included in error. The correct affiliation for this author is the Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia. The error has been corrected in the HTML and PDF versions of the article.