Supplementary Note to Accompany

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

1. PRIMARY GWAS OF SUBJECTIVE WELL-BEING 4
A. Background 4
B. Overview of subjective well-being analyses 4
C. Participating cohorts 5
D. Study-specific measures 6
E. Genotyping and imputation 6
F. Association analyses 7
G. Quality-control procedures 7
H. Primary meta-analyses 11
I. Post hoc meta-analysis of SWB (1000G) 12
2. TRADEOFF BETWEEN SAMPLE SIZE AND PHENOTYPE HETEROGENEITY 15
A. Theoretical framework 15
B. RMSE criterion 17
C. Technical appendix 20
3. AUXILIARY GWAS OF DEPRESSIVE SYMPTOMS AND NEUROTICISM 22
A. Participating cohorts 22
B. Phenotypes 22
C. Genotyping and imputation 23
D. Association analyses 23
E. Quality control 24
F. Meta-analyses 24
G. Results 24
4. POPULATION STRATIFICATION 27
A. LD Score intercept test 27
B. Within-family sign test 28
C. Within-family regression test 31
D. Discussion 33
5. LONG-RANGE LD REGIONS AND INVERSION POLYMORPHISMS 35
A. Analysis of Chr8 inversion polymorphism 35
B. Analysis of chr17q21 inversion polymorphism 41
C. Long-Range LD region on chromosome 5 42
6. POLYGENIC PREDICTION 44
A. Phenotype measures by cohort 44

B. RESULTS 47
7. PROXY-PHENOTYPE AND GENETIC OVERLAP ANALYSES 48
A. METHODOLOGY FOR PROXY-PHENOTYPE AND CROSS-PHENOTYPE ENRICHMENT ANALYSES 48
B. RESULTS FROM PROXY-PHENOTYPE AND CROSS-PHENOTYPE ENRICHMENT ANALYSES 49
C. GENETIC CORRELATIONS: METHODOLOGY 50
D. GENETIC CORRELATIONS: RESULTS 51
E. POST HOC LOOKUP OF SCHIZOPHRENIA-ASSOCIATED LOCI IN NEUROTICISM 53
8. BAYESIAN CREDIBILITY ANALYSIS 55
A. SINGLE PHENOTYPE FRAMEWORK 55
B. POSTERIOR DISTRIBUTION OF EFFECT ON SECOND-STAGE PHENOTYPE 58
C. BIVARIATE FRAMEWORK 59
D. APPLICATION 62
9. BIOLOGICAL ANNOTATION 65
A. FUNCTIONAL PARTITIONING OF HERITABILITY 65
B. LD WITH NON-SYNONYMOUS VARIANTS 69
C. EQTL ANALYSES 70
D. GWAS CATALOG LOOKUP 72
E. USING CO-EXPRESSION TO PREDICT GENE FUNCTION 73
F. SUMMARY OVERVIEW OF RESULTS 75
10. LOOKUP OF TOP SNPS IN COMPANION STUDY OF DEPRESSION 77
A. BACKGROUND 77
B. META-ANALYSIS OF 54 DS-ASSOCIATED SNPS 77
11. ADDITIONAL INFORMATION 78
A. AUTHOR CONTRIBUTIONS 78
B. ADDITIONAL ACKNOWLEDGMENTS 91
12. REFERENCES 106

1. Primary GWAS of subjective well-being

A. Background

There is much interest in the genetic basis of SWB, with twin studies suggesting that genetic factors may account for as much as 40% of the variance in SWB across individuals[1]. A recent study of the "common narrow heritability" (also called "SNP heritability") of SWB estimated that 5-10% of the variance in SWB can be explained by the cumulative additive effects of genetic variants that are common in the population[2]. Because such common variants are assessed by contemporary genome-wide approaches to genotyping single nucleotide polymorphisms (SNPs), that study concluded that a GWAS on SWB in a sufficiently large sample of individuals may yield reliably associated SNPs.

While some unsuccessful attempts have been conducted to identify genomic regions of interest for SWB[3]—including some candidate gene studies[4,5] and a genome-wide linkage study[6]—as far as we know, there is no large-scale meta-GWAS effort like the one reported here. From the outset we reasoned that the results of this research would have scientific merit regardless of the outcome. If the meta-analysis succeeded in identifying genetic variants, then such information would be an important first step toward understanding the pathways between genes, SWB, and other phenotypes, as well as the complex interplay between these pathways and the environment. If, on the other hand, no robust associations were uncovered, it would allow us to put a much tighter upper bound on the expected effect sizes for common variants associated with complex traits such as SWB.

B. Overview of subjective well-being analyses

The genome-wide association study (GWAS) of subjective well-being (SWB) is based on summary statistics uploaded by cohort-level analysts to a central server. The summary statistics were subsequently quality controlled and meta-analyzed by a central team of analysts. The lead PI of each cohort affirmed that the results contributed to the study were based on analyses approved by the local Research Ethics Committee and/or Institutional Review Board by signing a collaboration agreement which contained the clause "Each Representative Signing this document on behalf of a particular cohort is responsible for ensuring that the Institutional Review Board (IRB) or ethical committee has approved the analysis of well-being in that sample." **Supplementary Note 11** contains information about the cohort-level ethical review boards which approved the analyses; additional information can be obtained from the lead PI of each cohort. All participants provided written informed consent.

SWB is usually defined broadly to include both positive and negative subjective evaluations. Across the many facets of SWB, a distinction is often made between "positive affect" (PA) and "life satisfaction" (LS)[7,8]. PA refers to the frequency and intensity of positive emotions and feeling happy. Typical survey questions used to gauge PA include "During the past week, I was happy?" and "How would you rate your emotional wellbeing at present?" LS refers to a longer-term evaluation of one's life. A typical survey question would be "How satisfied are you with your life as a whole?". The two facets are known to be positively correlated with each other and load on a common genetic factor[9]. For this reason, and to maximize sample size, we decided a priori to make our primary analysis one in which we pool the two measures in a combined analysis, and to report but treat as secondary, analyses of PA and LS considered separately. **Supplementary Note 2** shows that combining the phenotypes in a single analysis does indeed maximize statistical power to detect associations under plausible assumptions.

Although LS and PA are phenotypically distinct, multivariate genetic analyses have found that the variance in LS and the variance in PA are explained by the same underlying common genetic factor[9]. For this reason, and to maximize sample size, we decided prior to conducting the study to make our primary analysis one in which we pool the two measures in a combined analysis, and to report, but treat as secondary, analyses of PA and LS considered separately. This approach is in line with modeling SWB as a "hierarchical construct"[10]. By considering the pooled wellbeing measure as the higher order phenotype, we build on the relatedness of the SWB components (LS and PA). At the same time, by also reporting the LS and PA analyses separately, we allow for the possibility of differences in genetic factors underlying these components of SWB, and we can disentangle whether associations we identify with the combined measure are driven by one of these two facets of SWB.

At the beginning of the study, we circulated an analysis plan describing the cohort-level analyses needed (including the exact specification and restriction to European-descent individuals). The plan asked cohort analysts to upload results by April 16, 2012, but this deadline was not strictly enforced. Final results files were uploaded in October. We subsequently performed a meta-analysis of the results ($N \approx 100{,}000$) and failed to find any genome-wide significant hits. The absence of significant results was not surprising in light of a paper published by PIs in 2013[2] that found that the heritability due to common variants of single-question measures of subjective well-being was around 5%. Given this new information, it was decided at the SSGAC meeting on 15 June 2013 to re-open the discovery phase, invite other cohorts to contribute, and integrate newly available data from currently contributing cohorts. Despite the well-known biases in statistical inference from "optional stopping," we felt it was justified to relax the data-freeze deadline due to the near doubling of the potential sample size.

We accordingly formulated an updated analysis plan, posted at https://osf.io/cq2b5/, which stipulated a data freeze of either 31 December 2014, or the day on which the combined sample size of the uploaded results for PA exceeded $N = 150{,}000$, whichever of the two events occurred earlier. By December 2014, we had attained sample sizes of approximately $N = 117{,}000$ for PA, $N = 85{,}000$ for LS, and $N = 152{,}500$ for the combined well-being (WB) phenotype. At that time, the cohort 23andMe, which had contributed 30,000 observations to the original analysis, indicated that they would be willing to upload new results based on a much larger sample ($N = 90{,}000$) that had since become available. A much larger sample size would also be possible due to the imminent release of the first batch of data from the UK Biobank[11] with $N \approx 120{,}000$ genotyped European-ancestry individuals, roughly 60,000 of whom have answered a high-quality PA question. In December 2014, we therefore decided to further postpone the data freeze until data from 23andMe and UKB became available. This decision was made prior to running any meta-analyses of results uploaded by December 2014.

C. Participating cohorts

Supplementary Table 2 provides study-specific details on all results files from the 59 participating cohorts that passed the quality-control analyses described below. Any cohort with acceptable survey measures of either PA or LS was eligible to participate in the study. Some cohorts had measures of both PA and LS, sometimes measured on more than one occasion. Such cohorts were encouraged to upload three results files: one for PA, one for LS, and one for a combined WB measure constructed by combining the LS and PA responses. Cohorts with multiple measures were encouraged to average responses to reduce the amount of transitory variation in responses. The exact construction of the combined measure was typically

determined in consultation with the cohort analyst. Of the 59 participating cohorts, 7 uploaded LS results only (hereafter, "LS cohorts"), 28 uploaded PA results only ("PA cohorts"), 12 uploaded results for LS and PA but not WB ("LSPA cohorts"), and 12 uploaded LS, PA, and WB results files ("LSPAWB cohorts"), leaving us with a total of 7 + 28 + 12×2 + 12×3 = 95 results files.

D. Study-specific measures

Supplementary Table 3 summarizes the study-specific LS and PA phenotypes. We purposely eschewed limiting the study to a specific questionnaire or survey scale, reasoning that the sample-size gains from an inclusive strategy that permitted some variation in question phrasing would outweigh any loss of power arising due to phenotypic heterogeneity. However, we did not allow questions that asked about happiness or satisfaction in specific domains (e.g., satisfaction with one's health status or financial situation).

Supplementary Table 3 shows that cohorts overwhelmingly used survey questions derived from or adapted from established survey batteries. Of the 31 cohorts with measures of LS, 19 used questions taken or adapted from popular and psychometrically validated life satisfaction or depression scales such as the Satisfaction With Life Scale[12] or the Geriatric Depression Scale[13] (which has a sub-item on life satisfaction suitable for our purposes). Another 7 cohorts used questions adapted from one of the main LS questions of the World Values Survey: "All things considered, how satisfied are you with your life as a whole these days?" Most remaining cohorts asked questions of the general character "How satisfied are you with your life?" Overall, the questions were thus phrased very similarly. The only possible exception of the question used by the 1958 British Birth Cohort ("On balance I look back at my life with a sense of happiness"), which, even though the question contains the word "happiness", we chose to classify as an LS measure because we interpreted the question as an evaluative measure of satisfaction. For LS, there is hence overall little cause for concern that variation in the phrasing of the question introduced substantial phenotypic heterogeneity. The number of response categories varies across questions. However, cohorts with four or fewer response categories account for only 15.4% of the LS sample.

The PA phenotypes analyzed exhibit greater variation and include items such as "During the past week I was happy" and "Do you feel happy most of the time?". Some cohorts also used scores from psychological scales, such as the Subjective Happiness Scale[14] or the well-being trait scale of the Multidimensional Personality Questionnaire[15]. Remaining cohorts used sub-items from a diverse set of psychological questionnaires which includes the Hospital Anxiety and Depression Scale[16], the Scale of Positive and Negative Experience[17], and the Positive Affect and Negative Affect Scale[18]. By far the most common strategy was to use items about positive affect from the Center for Epidemiologic Studies Depression ("CES-D") scale[19], a standard depression battery, which contains a four-item Positive Affect subscale used by some cohorts, whereas others used a single item on this subscale that asks specifically about happiness ("Last week, [how often] were you happy?").

E. Genotyping and imputation

Genotyping was performed using a range of common, commercially available genotyping arrays. **Supplementary Table 4** provides study-specific details on genotyping platform, pre-imputation quality-control filters applied to the genotype data, subject-level exclusion criteria, imputation software used, and the reference sample used for imputation. Because our study was launched before 1000G-imputation became standard practice, the analysis protocol circulated to analysts recommended uploading results

imputed using the HapMap 2 CEU (r22.b36) reference sample[20]. Our analysis plan advised cohorts to exclude from their estimation sample subjects with low overall call rates (< 95%), excess autosomal heterozygosity, or sex mismatch (excessive X-chromosome homozygosity in males). Additionally, the plan advised family-based cohorts to only include one relative from each pedigree or to report standard errors adjusted for the sample relatedness. We encouraged, but did not require, later enrollees into the study to supply us with 1000G-imputed data. **Supplementary Table 4** shows that 40 out of 59 cohorts supplied HapMap2-imputed data. Though the cohorts with 1000G data are a minority, they are larger on average, accounting for more than 70% of our combined sample size.

F. Association analyses

Cohorts were asked to estimate the following regression equation for each SNP:

$$Y = \beta_0 + \beta_1 SNP + \boldsymbol{PC}\,\boldsymbol{\gamma} + \boldsymbol{B}\,\boldsymbol{\alpha} + \boldsymbol{X}\,\boldsymbol{\theta} + \epsilon, \quad (1)$$

where *Y* is an unstandardized outcome variable, *SNP* is the allele dose of the SNP; ***PC*** is a vector of the first four principal components of the variance-covariance matrix of the genotypic data, estimated after the removal of genetic outliers; and ***B*** is a vector of standardized controls, including sex, age and age squared. Cohorts were also asked to include any study-specific covariates such as study site or batch effects, that they considered appropriate. Some cohorts with binary dependent variables uploaded results from a logistic regression model analogous to Equation (1).

G. Quality-control procedures

Generating a Reference File Mapping rsIDs to ChrPosIDs

SNPs imputed using HapMap and 1000G reference panels are ordinarily assigned chromosomal position identifiers ("ChrPosID", a concatenation of a SNP's chromosome number, a colon, and the SNP's base pair position) using different versions of the NCBI build. In several of our analyses, it is desirable to use a harmonized one-to-one mapping from rsID to chromosomal coordinates. We therefore restrict all our SWB analyses to a set of autosomal SNPs with rs identifiers ("rsIDs") that (i) appear in both the HapMap and 1000G reference panels, and (ii) for which a ChrPosID can be generated in build 37 coordinates. To generate this list, we used files that have been made publicly available by the developers of the EasyQC software[21]. All combined data from a number of public sources. For details on construction of several of the key files, see pp. 2-3 in the Supplementary Material of Winkler et al.[21].

In the EasyQC reference files, there are 2,532,578 HapMap SNPs with non-missing information about (i) rsID, (ii) build 36 coordinates, and (iii) European allele frequency in the HapMap Phase 2 CEU reference sample. From this original set of SNPs, we dropped 66 SNPs because their ChrPosID (in build 36 coordinates) was not unique, 31,657 SNPs whose rsIDs could not be located in the 1000G reference file, and an additional 12,921 SNPs for which the allele frequency is not available in the Europeans-only 1000G reference file (the version which excludes X-chromosome markers and monomorphic SNPs). In the final step, we dropped 719 SNPs due to possible allele misalignment and 2,471 SNPs whose allele frequencies differed by more than 0.25 across the HapMap and 1000G European reference samples. This leaves us with 2,484,798 rsIDs with information about allele frequency in a reference sample of Europeans, and a one-to-one mapping to a ChrPosID expressed in build 37 coordinates. In what follows, we refer to this as our reference file.

Pre-QC Verification of Descriptives

For each cohort, we checked whether (i) they had supplied us with complete descriptive statistics, and (ii) the variable coding was in accordance with the analysis plan (e.g., higher values indicating greater WB, PA, or LS). If not, we contacted the cohort to obtain corrected data. Next, as we explain below, we verified to the extent allowed by our data that the reported information was consistent with the uploaded summary statistics. Many of the QC checks were intended to eliminate problems of reverse coding that, if undetected, can substantially reduce power to detect associations.

EasyQC

We used the software EasyQC[21] to check each uploaded results file for quality-control problems. From each uploaded file, we filtered out SNPs in the following order.

i. We first dropped any SNPs in the uploaded results that could not be identified in the reference file, whose construction we described above. In a few cohorts who had imputed their data against the September or December 2013 releases of the 1000 Genomes Phase 1 haplotypes provided by the software IMPUTE2, we also dropped the 730+199 SNPs whose strands are known to have been incorrectly aligned in these releases.[a]

ii. We dropped a SNP if neither an effect nor other allele was supplied, or if either of them takes values other than "A", "C", "G", or "T". We also dropped a SNP if any of the following variables were missing: *p*-value, a coefficient estimate (beta) and its standard error, effect allele frequency, sample size (*N*), and imputation accuracy (for imputed SNPs). We also dropped SNPs if any of the variables reported for the SNP were outside the permissible range of the variable (for example, *p*-values greater than 1 or negative standard errors). As in Okbay et al.[22], we dropped SNPs from cohorts that used likelihood-ratio tests for inference if the reported LR test statistic differed by more than 10% from the squared *t*-statistic constructed by dividing the estimated regression coefficient by the analytical standard error (i.e., the standard error obtained from the information matrix evaluated at the maximum likelihood estimates). This QC screen is based on the observation that both the LR test statistic and the squared *t*-statistic should have the same distribution under the null hypothesis, namely a chi-squared with 1 degree of freedom. Seven cohorts reported *p*-values from likelihood-ratio (LR) tests.

iii. We dropped SNPs with minor allele frequencies below a threshold that varied by sample size. In samples with fewer than 1000 observations, we applied a threshold of 10%. In studies with a sample size between 1000 and 2000, we dropped SNPs with MAF < 5%. In all other samples, we applied a threshold of 3%. **Supplementary Table 5** summarizes these filters and others used in the steps below.

iv. We filtered out SNPs with low imputation accuracy. The definition of the imputation accuracy metric varies by imputation software. If the cohort supplied us with the "Rsq" variable generated by MaCH[23], we dropped SNPs with Rsq < 0.4. If they uploaded the "INFO" variable generated by

[a] The announcement is available on https://mathgen.stats.ox.ac.uk/impute/impute_v2.html#whats_new

IMPUTE[24], we applied a threshold of 0.5. If PLINK's "info" variable was supplied, we applied a threshold of 0.8.

v. We dropped SNPs with low Hardy-Weinberg Equilibrium (HWE) *p*-value (see **Supplementary Table 5** for exact cutoffs used).

vi. We dropped SNPs with call rate below 95%.

vii. We dropped duplicated SNPs based on Build 37 base pair positions obtained by mapping the rsIDs in each results file to the ChrPosIDs in the reference file. We also dropped SNPs that could not be successfully aligned due to mismatch with reference alleles.

Having applied all the filters to the cohort-level summary statistics, we examined how many SNPs were dropped in each filtering step. Whenever an unusual number of markers was being dropped, we flagged the cohort as potentially having an error in the uploaded results file. The issue was discussed with the cohort-level analyst and resolved through a new QC iteration. **Supplementary Table 6** shows, separately for each cohort, the number of SNPs dropped in each filtering step and also shows the estimated genomic control factors. These are in the range 0.871 to 1.099, with a median value of 1.004.

Visual inspection of diagnostic plots

Having processed the data through these filters, we inspected several diagnostic plots.

1. *Allele Frequency Plots (AF Plots):* We looked for errors in allele frequencies and strand orientations by inspecting a plot of sample allele frequencies against the allele frequency in a European reference sample. Any deviations could indicate a number of problems (failure to exclude ethnic outliers, imputation errors, etc.).
2. *P-Z Plots:* We checked that reported *p*-values are consistent with the reported coefficient estimates and their *SE*s.
3. *Q-Q Plots:* We visually inspected the cohort-level Q-Q plots to look for evidence of unaccounted-for stratification.

Potential issues with the plots were always raised with the cohort-level analyst. All analyses are based on results files whose plots did not indicate any quality-control problems.

Additional diagnostics

We supplemented the EasyQC checks with the following diagnostic checks:

i. For cohorts that uploaded results for more than one phenotype, we calculated the correlation between the estimated regression coefficients for the two phenotypes. Since the phenotypes are positively correlated, we expect positively correlated coefficient estimates unless the sample overlap is minimal; negatively correlated betas would be a strong indication of phenotype miscoding or allele misalignment in one of the results files. We found pairwise correlations in the range 0.2 to 0.9, with correlations in the lower end of this range only in those cohorts where the LS and PA questions were asked several years apart.

ii. We conducted a number of additional diagnostic analyses in a restricted set of SNPs comprising the union of a (i) set of 50,000 randomly sampled HapMap2 SNPs and (i) SNPs showing suggestive

evidence of association in the 23andMe data (henceforth the 50K SNP set). The list of 23andMe SNPs was constructed by selecting the lowest *p*-value SNPs from a set of approximately independent loci. These independent loci were determined using a frequently used[25] iterative clumping procedure implemented in Plink[26], as follows. First, the SNP with the smallest *p*-value is identified in the 23andMe results. This SNP is the lead SNP of clump 1. Second, all SNPs whose association *p*-value is lower than 10^{-3} that are within 500 kb of the lead SNP and whose LD with the lead SNP exceeds $R^2 = 0.1$ are identified and assigned to clump 1. We calculate LD using the 1000G phase 1 reference sample composed of Utah Residents (CEPH) with Northern and Western European Ancestry (CEU), Toscani in Italia (TSI), and British in England and Scotland (GBR)[27]. To generate the second clump, the SNP with lowest *p*-value among the SNPs that remain after removal of clump 1 is identified and the same steps are applied to identify the set of SNPs comprising clump 2. The process is repeated until no SNPs with *p*-values below 10^{-3} remain. This process left us with 823 approximately independent loci.

a. For each cohort, we calculated the degree of sign concordance with the list of 23andMe SNPs. Because of the substantial estimation error expected for individual SNPs in single cohorts, sign concordance below 50% is only suggestive of data problems, except in the case of very large cohorts. The overall tendency is for the signs to align more often than expected by chance and in no case did our 95% CIs allow us to reject a sign concordance below 50%.

b. We examined how SE's predicted from the *N*'s and *SD*'s supplied in the descriptive statistics compared to the *SE*'s in the results files. Winkler et al.[21] propose a similar diagnostic (the *SE-N Plots*) which is based on following approximation to the standard error of a coefficient estimated by OLS

(2)
$$SE_j \approx \frac{\hat{\sigma}_Y}{\sqrt{N}} \cdot \frac{1}{\sqrt{2\, MAF_j \left(1 - MAF_j\right)}},$$

where $\hat{\sigma}_Y$ is the standard deviation of the dependent variable (equal to 1 in cohorts that reported standardized regression coefficients), MAF_j is the minor allele frequency of SNP *j*, and *N* is the sample size. We used Equation (2) to generate a predicted standard error for the 50K SNP set, and we then plotted these predicted standard errors against the reported standard errors. We used an analogous equation for cohorts with binary dependent variables that ran logistic regressions. These plots, which we refer to as 50K plots in what follows, were used to check for systematic discrepancies between the predicted and reported standard errors and for outlier *SE*'s. These analyses helped us identify and remedy QC problems or errors in some of the reported summary statistics.

iii. To supplement the 23andMe sign tests, we used bivariate LD score regression[28] to estimate the pairwise genetic correlation, r_g, between the 23andMe sample and each cohort with more than 5,000 observations. We also estimated the r_g between the 23andMe sample and several combinations of small samples that were meta-analyzed (each of the individual cohorts, considered individually, is too small to generate informative estimates of r_g). In the 16 out of 23 instances where the estimator converged, the estimate is positive, as expected. We reject a genetic correlation of zero at the 5% significance level in 8 cases.

Again, any anomaly discovered during the course of applying the quality-control steps described above were raised with analysts. In some cases, multiple iterations with analysts were required before the source of the anomaly was identified, problems fixed, and the results files were cleared for inclusion.

H. Primary meta-analyses

Meta-analyses of SWB, PA and LS

Our analyses are based exclusively on results files that have passed the diagnostic tests described in the previous section. In our primary analyses of the pooled SWB phenotype, we use the WB variable from LSPAWB cohorts. In LSPA cohorts, we used the LS variable, because our analyses suggested its SNP-based heritability was slightly higher, except in cohorts where the LS variable is binary and the PA variable is not. For remaining cohorts, we use whichever results file is available (LS or PA) in the combined analysis. **Supplementary Table 1** also shows that the constructs are highly genetically correlated.

Though we consider them secondary to the SWB analyses, we also performed separate analyses of PA and LS. We thus ran three meta-analyses:

- SWB (*N* = 298,420)
- PA (*N* = 180,281)
- LS (*N* = 166,205)

For each phenotype, we used the software Metal[29] to perform a sample-size weighted meta-analysis of the cohort-level summary statistics. With the exception of one cohort, described below, we do not apply cohort-level genomic control[30] to adjust the standard errors for non-independence. Instead, we meta-analyze the unadjusted cohort-level summary statistics and subsequently inflate the standard errors from the meta-analysis by the square root of the estimated intercept from an LD score regression[28]. The exception is deCODE, whose cohort-level regression estimates are not adjusted for the high level of relatedness in the sample: deCODE's standard procedure is to apply genomic control prior to uploading meta-analysis results. Because the LD score regression intercepts are upward biased in cohorts with related individuals unless the relatedness is accounted for in standard error estimation, we estimate the LD score intercept omitting deCODE from the analysis.

The QQ plots for the SWB, LS and PA analyses are shown in **Supplementary Figures 3a, 3c, and 3d**; Manhattan plots are in **Figure 1** and **Supplementary Figs. 4a-c**. As is expected under polygenicity[31], there is more QQ plot inflation of the median test statistic in the combined analysis ($\lambda_{GC} = 1.206$) than in the separate analyses of LS ($\lambda_{GC} = 1.119$) and PA ($\lambda_{GC} = 1.118$). The estimated LD intercepts suggest that the amount of confounding is minimal: the estimates are 1.012 (SWB), 1.011 (PA) and 1.007 (LS). These analyses suggest that nearly all of the observed inflation is due to polygenicity, a conclusion consistent with results from additional analyses reported in **Supplementary Note 4**.

Supplementary Table 8 reports the set of approximately independent SNPs that reached nominal significance ($p < 10^{-5}$) in each of the three meta-analyses. To determine these independent loci, we used the following algorithm. First, the SNP with the smallest *p*-value is identified in the pooled meta-analysis results. This SNP is the lead SNP of clump 1. Second, we identified all SNPs whose LD with the lead SNP exceeds $R^2 = 0.1$ and assigned them to the clump. We calculate LD using the 1000G phase 1 reference sample composed of Utah Residents (CEPH) with Northern and Western European Ancestry (CEU), Toscani

in Italia (TSI), and British in England and Scotland (GBR). To generate the second clump, the SNP with lowest *p*-value among the SNPs that remain after removal of clump 1 is identified and the same steps are applied to identify the set of SNPs comprising clump 2. The process is repeated until no SNPs with *p*-values below 10^{-5} remain. This process left us with 44 approximately independent loci in the SWB analysis (3 genome-wide significant), 30 in the LS analysis (2 genome-wide significant), and 30 in the PA analysis (none genome-wide significant). If we instead apply cohort-level genomic, all three SNPs which reach genome-wide significance in the SWB analysis remain genome-wide significant, whereas one of the two LS SNPs remains genome-wide significant.

I. Post hoc meta-analysis of SWB (1000G)

When this project was launched in 2012, the standard reference panel for imputation in GWAS meta-analyses was HapMap Phase 2[20], and our primary analysis of SWB was therefore restricted to testing HapMap2 SNPs for association with SWB. As outlined in **Supplementary Note 1.E**, cohorts with 1000G-imputed data nevertheless account for over 70% of the combined sample. Following the suggestion of a referee, we performed a post hoc GWAS of SWB in which we analyzed 1000G SNPs in all cohorts for which such data were available.

The twin goals of this post hoc analysis were to check if the finer resolution of the 1000G-imputed data would allow us to (i) fine-map the three genome-wide significant associations identified in the primary analysis and (ii) identify any novel associations.

Participating cohorts

We included 18 out of the 19 cohorts that used a 1000G reference sample for imputation (listed in **Supplementary Table 4**). The exception is deCODE; the deCODE GWAS was performed in 1000G-imputed data, but only HapMap2 SNPs were tested for association. To further increase the sample size, we reran the SWB GWAS in four of the cohorts (STR1, MCTFR, RS1, RS2 and RS3) for which only results for HapMap2-imputed variants were available when the primary meta-analysis was performed. These four cohorts were included because the QC team had access to individual-level genotypic and phenotypic data, making it feasible to rerun these analyses in 1000G SNPs. Below, we refer to the 24 (19 + 5) cohorts with GWAS results for 1000G SNPs as the 1000G cohorts. Their combined sample size is N = 229,883, or approximately 77% of the sample in the primary SWB analysis.

Quality control

We followed the quality-control procedures similar to those described in **Supplemental Note 1.G** with quality-control parameters shown in Panel B of **Supplementary Table 5**, and the following additional filters:

1. If the data were imputed against the September or December 2013 releases of the 1000 Genomes Phase 1 haplotypes provided by the software IMPUTE2, we drop the 730+199 SNPs whose strands were incorrectly aligned in these releases.
2. We drop indels and structural variants.

To generate a harmonized one-to-one mapping from rsID to chromosomal coordinates for the 1000G variants, we created a new reference file for use by EasyQC[21]. All our quality-control analyses were performed using this file, whose construction is described below.

Our reference file was generated by processing data downloaded on December 22, 2015, from the website of the imputation software MACH[23]**.**

The first file consists of individual-level data on all European-ancestry (EUR) individuals in 1000G Phase 1 Integrated Release Version 3 Haplotypes (hereafter, 1000G Phase 1):

File Name: EUR phase1_release_v3.20101123.snps_indels_svs.genotypes.refpanel.EUR.vcf
URL: http://csg.sph.umich.edu/abecasis/mach/download/1000G.2012-03-14.html

The second file consists of individual-level data on all individuals in the 1000G Phase3 v5 Reference (hereafter, 1000G Phase 3):

File Name: reduced.ALL.phase3_shapeit2_mvncall_integrated_v5.20130502.genotypes.vcf
URL: http://csg.sph.umich.edu/abecasis/mach/download/1000G.Phase3.v5.html

From both data sets, we retain individuals who are members of the CEU (Utah Residents (CEPH) with Northern and Western European Ancestry), TSI (Toscani in Italia), or GBR (British in England and Scotland) populations. Additionally, we used the software Plink to restrict the sample so that it does not include any pairs of individuals whose estimated genomic relatedness exceeds 0.025[26]. As in the main analyses, we also restrict all analyses to autosomal, biallelic SNPs.

Imposing these restrictions in the 1000G Phase 1 sample yields a sample of *N* = 258 CEU/GBR/TSI approximately unrelated individuals and 16,001,120 SNPs. Each SNP in the resulting sample is uniquely identified by its ChrPosID, and each ChrPosID in turn maps to a unique rsID. Imposing the same restrictions in the 1000G Phase 3 sample leaves *N* = 294 CEU/GBR/TSI individuals and 43,805,190 SNPs. We drop 514,910 markers with non-unique ChrPosIDs. After this restriction is imposed, each variant is again uniquely identified by its ChrPosID, and each ChrPosID maps to a unique rsID.

We subsequently impose the following restrictions:

1. We drop 28,518,368 SNPs whose minor allele count is zero in the CEU/GBR/TSI subsample.
2. We drop 3,085 SNPs whose two alleles are not consistent across the Phase 1 and Phase 3 releases.
3. We drop 77,964 SNPs because their ChrPosIDs do not map to the same rsID in the Phase 1 and Phase 3 samples.
4. We drop 10,308 SNPs because the absolute value of the difference in allele frequency between the Phase 1 and Phase 3 sample exceeds 0.25 in the CEU/GBR/TSI subsamples.

These restrictions leave us with our reference file containing 14,680,555 SNPs, which we use both for quality control and pruning. To maximize comparability with the results from the primary analyses, we closely followed the procedures used in the primary SWB analysis. The results from the SNP filtering are in Panel B of **Supplementary Table 6**.

Results of post hoc analysis

We sample-size weighted the cohort-level summary statistics using in the software Metal[29]. As in the main analysis, we adjusted the standard errors using square root of the estimated intercept from an LD score regression[28]. The association analyses were performed using the same association models as the primary

analyses. Restricting the meta-analysis to SNPs with association results for at least *N* = 100,000 individuals leaves us with 9.00M SNPs passing all quality control filters.

The QQ plot for SWB is shown in **Supplementary Figure 3b**. As expected given the smaller sample size, the observed level of inflation ($\lambda_{GC} = 1.124$) is a little lower than estimated in the primary analyses ($\lambda_{GC} = 1.21$). And consistent with the results in the primary analysis, the estimated LD intercept (1.008) suggests a minimal amount of confounding. The combined analyses yielded 76 nominally associated at the *p*-value threshold 10^{-5}, reported in Panel D **Supplementary Table 7** (compared to 44 loci in the main SWB analysis)**.** Two of these 76 SNPs reached genome-wide significance: rs6579956 (4.44×10^{-9}) and rs13185787 (2.13×10^{-8}). Both SNPs lie in a long-range LD region on chromosome 5 (between 129 and 132 Mb[32]; see **Supplementary Note 5**).

The first novel SNP, rs6579956, is in linkage disequilibrium with one of the three original SWB-associated SNPs identified in the primary analysis (rs4958581). rs6579956 is positioned 1090kb downstream from rs4958581 and the two SNPs are only in modest linkage disequilibrium (R^2 = 0.39). The SNPs lie in a region with long intergenic non-coding RNA (lincRNA). Although lincRNAs are not protein-coding, the majority of lincRNAs are thought to have functional consequences, such as regulation of gene expression or conservation of transcript integrity[33].

Following the methodology described in **Supplementary Note 9.C**, we used the Genotype-Tissue Expression Portal (www.GTExportal.org)[34] to test both SNPs for association with gene expression levels across 15 tissues. We found that rs6579956 was significantly associated with gene expression of *IK*, *PCDHA1*, *PCDHA3*, *PCDHA4*, *PCDHA7*, *PCDHA10*, *SLC4A9*, and *TMCO6*, whereas we found no evidence that rs4958581 was associated with gene expression levels. We found no evidence that either SNP was in moderate to high LD with any nonsynonymous SNPs.

The second novel hit rs13185787 lies in the intronic region of *CDC42SE2,* and is in high linkage disequilibrium (R^2 = 0.95) with the other SWB-associated SNP on chromosome five identified in the primary analysis, rs3756290, located in the intronic region of *RAPGEF6.* We conducted analyses identical to those described in the previous paragraph, but found no evidence that either SNP is significantly associated with gene expression levels across 15 tissues tested, nor that either SNP is in moderate to high LD with any nonsynonymous SNPs.

2. Tradeoff between sample size and phenotype heterogeneity

In genome-wide association studies, investigators often combine data from multiple cohorts. When the exact phenotype measures differ across cohorts, researchers face a dilemma: on the one hand, including that cohort in the meta-analysis will increase the total sample size; on the other hand, including the different phenotype measure from that cohort may lead to bias in the GWAS estimates. We refer to this tradeoff between maximizing sample size and having a more precise and uniform measure of the phenotype across cohorts as the "Quantity-Quality Tradeoff". Here, we use a stylized model that provides a framework for thinking about the tradeoff, and can be used to quantify the costs and benefits of different research strategies.

A. Theoretical framework

To facilitate the exposition of the framework, we consider the case of a single SNP x, but all insights apply also to settings where the number of genetic variants tested for association is large.

We assume that that there are two phenotypes, y_1 and y_2, whose genetic correlation is r_g. Denote the effect of SNP x on each of the two traits by β_1 and β_2. In what follows, we assume that the coefficient of interest is β_1. Suppose that we can either (i) estimate β_1 in a sample of size N_1 in which a uniform measure of y_1 is available, or (i) estimate β_1 by pooling data from the first sample with data from an independent sample in which y_2 has been measured in N_2 subjects. In this situation, the fundamental tradeoff is that pooling the samples may reduce estimation error, thus increasing the statistical power to detect associations, but any gains from increased precision must be weighed against the costs of the bias that is introduced by pooling two different phenotypes. Below, we refer to the two samples as samples 1 and 2.

Assuming y_1, y_2, and x are standardized to have zero mean and unit variance, we have:

$$y_1 = x\beta_1 + \varepsilon_1$$
$$y_2 = x\beta_2 + \varepsilon_2.$$

Following Bulik-Sullivan et al.[28], we assume that SNPs' true causal effects are randomly and independently distributed across the genome, and independent of the standardized SNP variable. The latter assumption, which is commonly made[28,35], is equivalent to assuming that the variance of a SNP's effect size is inversely proportional to the variance of its genotype, $2 \times \mathrm{MAF}(1 - \mathrm{MAF})$, implying that rarer variants have larger βs, on average.

The linear predictor of β_2 given β_1 is $\beta_2 = \frac{Cov(\beta_1,\beta_2)}{Var(\beta_1)}\beta_1$. We can therefore express β_2 as

$$\beta_2 = \frac{Cov(\beta_1,\beta_2)}{Var(\beta_1)}\beta_1 + u_2$$
$$= Corr(\beta_1,\beta_2)\frac{\sqrt{Var(\beta_2)}}{\sqrt{Var(\beta_1)}}\beta_1 + u_2$$
$$= r_{g,LD}\sqrt{\frac{h_2^2}{h_1^2}}\beta_1 + u_2$$

where u_2 is uncorrelated with β_1, where $r_{g,LD} = \frac{\rho_g}{\sqrt{h_1^2 \cdot h_2^2}}$ is the genetic correlation between y_1 and y_2 as defined in the LD Score correlation framework[36], h_k^2 is the heritability of y_k and ρ_g is the genetic covariance between y_1 and y_2.

The last equality for β_2 above follows from substituting $corr(\beta_1, \beta_2) = r_{g,LD}$ and $Var(\beta_k) = \frac{h_k^2}{M} E[\ell]$, where $E[\ell]$ is the mean LD Score of the SNPs and *M* is the number of SNPs in the data used to calculate the LD Score (see **Supplementary Note 2.C** for formal proofs). Using the latter equality for *k* = 1,2, it follows that $Var(\beta_2) = \frac{h_2^2}{h_1^2} Var(\beta_1)$. Thus,

$$\sigma_{u_2}^2 \equiv Var(u_2) = Var(\beta_2) - r_{g,LD}^2 \frac{h_2^2}{h_1^2} Var(\beta_1)$$
$$= (1 - r_{g,LD}^2) \frac{h_2^2}{h_1^2} Var(\beta_1) = (1 - r_{g,LD}^2) Var(\beta_2).$$

Now consider meta-analyzing the two estimates, $\widehat{\beta_1}$ and $\widehat{\beta_2}$. From our assumption that the two samples are independent, we have that $cov(\widehat{\beta_1}, \widehat{\beta_2}) = 0$. In an inverse-variance weighted meta-analysis[b], the pooled coefficient estimate is:

$$\hat{\beta}_{meta} = \frac{\frac{1}{\widehat{Var(\beta_1)}} \widehat{\beta_1} + \frac{1}{\widehat{Var(\beta_2)}} \widehat{\beta_2}}{\left(\frac{1}{\widehat{Var(\beta_1)}} + \frac{1}{\widehat{Var(\beta_2)}} \right)} \approx \frac{N_1 \widehat{\beta_1} + N_2 \widehat{\beta_2}}{N_1 + N_2},$$

where the approximate equality follows because $\widehat{Var(\beta_k)} \approx \frac{\sigma_{\varepsilon_k}^2}{N_k Var(x)}$, x has unit variance in both samples, and $\sigma_{\varepsilon_1}^2 \approx \sigma_{\varepsilon_2}^2$ (since $\sigma_{\varepsilon_1}^2 \approx \sigma_{y_1}^2 = 1 = \sigma_{y_2}^2 \approx \sigma_{\varepsilon_2}^2$ since an individual SNP's effect is small). Thus, in our framework, $\hat{\beta}_{meta}$ is a weighted sum of $\widehat{\beta_1}$ and $\widehat{\beta_2}$:

$$\hat{\beta}_{meta} = \frac{N_1(\beta_1 + v_1) + N_2(\beta_2 + v_2)}{N_1 + N_2}$$
$$= \frac{N_1(\beta_1 + v_1) + N_2 \left(r_{g,LD} \sqrt{\frac{h_2^2}{h_1^2}} \beta_1 + u_2 + v_2 \right)}{N_1 + N_2}$$
$$= \frac{N_1 + N_2 r_{g,LD} \sqrt{\frac{h_2^2}{h_1^2}}}{N_1 + N_2} \beta_1 + \frac{N_1 v_1 + N_2 v_2 + N_2 u_2}{N_1 + N_2}$$

[b] The first equality corresponds to an inverse variance-weighted meta-analysis, but the derivations and the results below also hold if we combine the *t*-statistics for $\widehat{\beta_1}$ and $\widehat{\beta_2}$ in a sample-size weighted meta-analysis (indeed the two are asymptotically equivalent).

$$\approx \frac{N_1 + N_2 r_{g,LD}\sqrt{\frac{h_2^2}{h_1^2}}}{N_1 + N_2}\beta_1 + \frac{\sqrt{N_1}\zeta_1 + N_2\sqrt{(1 - r_{g,LD}^2)Var(\beta_2)}\zeta_{u_2} + \sqrt{N_2}\zeta_2}{N_1 + N_2}$$

$$= \frac{N_1 + N_2 r_{g,LD}\sqrt{\frac{h_2^2}{h_1^2}}}{N_1 + N_2}\beta_1 + \frac{\sqrt{N_1 + N_2 + {N_2}^2(1 - r_{g,LD}^2)Var(\beta_2)}}{N_1 + N_2}\zeta,$$

where v_1 and v_2 are the estimation errors in $\widehat{\beta_1}$ and $\widehat{\beta_2}$, and where $\zeta_1, \zeta_2, \zeta_{u_2}$ and ζ are standard normal random variables and $\zeta_1, \zeta_2, \zeta_{u_2}$ are independent. The approximate equality on the fourth line follows because $\sigma_{v_1} \approx \frac{\sigma_{y_1}}{\sqrt{N_1}} = \frac{1}{\sqrt{N_1}}$ and $\sigma_{v_2} \approx \frac{\sigma_{y_2}}{\sqrt{N_2}} = \frac{1}{\sqrt{N_2}}$.

From the formula above, the bias and variance of $\hat{\beta}_{meta}$ are given by

$$\text{Bias}(\hat{\beta}_{meta}) = \frac{N_2\left(1 - r_{g,LD}\sqrt{\frac{h_2^2}{h_1^2}}\right)}{N_1 + N_2}\beta_1$$

and

$$Var(\hat{\beta}_{meta}) = \frac{N_1 + N_2 + {N_2}^2(1 - r_{g,LD}^2)Var(\beta_2)}{(N_1 + N_2)^2}.$$

Finally, when only cohort 1 is included, $\hat{\beta}_{meta} = \widehat{\beta_1}$ and, because $\widehat{\beta_1}$ is an unbiased estimator with variance $\approx 1/N_1$, it follows that $\text{RMSE}(\hat{\beta}_{meta}) = \text{RMSE}(\hat{\beta}_1) \approx 1/\sqrt{N_1}$.

B. RMSE criterion

Here, we describe conditions under which including the second cohort in the meta-analysis would reduce the Root Mean Square Error (RMSE) of the estimator, defined as:

$$\text{RMSE}(\hat{\beta}_{meta}) = \sqrt{\left(\text{Bias}(\hat{\beta}_{meta})\right)^2 + Var(\hat{\beta}_{meta})}.$$

The RMSE jointly captures the accuracy and precision of an estimator and is increasing in the bias and the variance of the estimator; it is a useful measure when the objective is to obtain the best possible estimate of β_1. Also, when $h_2^2 \approx h_1^2$, the predictive power of a polygenic score for y_1 can be maximized by reducing the RMSE of the β's used to construct the score. (However, the RMSE is not always an appropriate criterion: for instance, if $h_2^2 \gg h_1^2$ and $r_{g,LD}^2 \approx 1$, then $\hat{\beta}_{meta}$ will be a very biased estimate of β_1 and its RMSE will be large, but the bias will be consistent across all SNPs and so will not reduce the explanatory power of the associated polygenic score.)

Observe that $E[\hat{\beta}_{meta}] = N_1 + N_2 r_{g,LD}\sqrt{h_2^2/h_1^2}(N_1 + N_2)^{-1}\beta_1 \neq \beta_1$, so $\hat{\beta}_{meta}$ is not an unbiased estimator of β_1. The bias vanishes in the special case when $r_{g,LD} = 1$ and $h_1^2 = h_2^2$. This is reassuring since it corresponds exactly to the case where $y_1$and y_2 have identical genetic architectures.

We thus see that there is a Quantity-Quality Tradeoff: adding sample 2 will increase sample size (N_2 in the denominator) and reduce the variance of $\hat{\beta}_{meta}$ since the denominator of $Var(\hat{\beta}_{meta})$ increases relatively

faster than the numerator as N_2 increases. The variance of $\hat{\beta}_{meta}$ is decreasing in the genetic correlation (as captured by the term $(1 - r_{g,LD}^2)Var(\beta_2)$) and $\hat{\beta}_{meta}$ is not an unbiased estimate of β_1.

$\mathrm{RMSE}(\hat{\beta}_{meta})$ is a function of the following parameters: N_1, N_2, $r_{g,LD}$, β_1, $Var(\beta_2)$, h_1^2 and h_2^2. To illustrate the tradeoff quantitatively, we calibrate these parameters to approximate the decision problem in this paper: we assume that $N_1 = 200{,}000$ and that $h_2^2 = h_1^2$. The assumption that $h_2^2 = h_1^2$ is a reasonable approximation given our estimates of the SNP-based heritability of the LS, PA and WB measures. As shown in **Supplementary Note 2.C**, $Var(\beta_2) = \frac{E[\chi_2^2]-1}{N_2}$, and we use the average sample size and the estimate of $E[\chi_2^2]$ from our GWAS of SWB (uncorrected for χ_2^2 inflation) to obtain $\widehat{Var(\beta_2)} = \frac{1.201-1}{277582} = 7.24 \times 10^{-7}$, which we take as our assumed value of $Var(\beta_2)$.

To evaluate the estimators, we normalize the RMSE by dividing by β_1. We can interpret this quantity as a fraction of β_1, the coefficient of interest. **Supplementary Figure 14** shows how $\mathrm{RMSE}(\hat{\beta}_{meta})/\beta_1$ and $\mathrm{RMSE}(\hat{\beta}_1)/\beta_1$ vary as functions of N_2, for various assumed values of the genetic correlation between y_1 and y_2, $r_{g,LD}$.

In **Supplementary Figure 14a**, the true R^2 of β_1 is assumed to be 0.02%. The straight horizontal line corresponds to the RMSE of $\hat{\beta}_1$ (i.e., when we only include cohort 1), which of course does not depend on N_2. As can be seen, when the genetic correlation is high ($r_{g,LD} = 0.95$ or 0.8), including the second cohort decreases the RMSE and is thus desirable (at least when N_2 is less than 300,000). If the genetic correlation is 0.6, it is desirable to include the second cohort if N_2 is less than ~36,000. If the genetic correlation is 0.3, it only makes sense to include the second cohort if N_2 is less than ~10,000. The intuition behind these results is that when $r_{g,LD}$ is high, little bias is introduced by including the second cohort—even when N_2 is very large—while the larger combined sample size reduces the variance of $\hat{\beta}_{meta}$, thereby reducing its RMSE; when $r_{g,LD}$ is low, however, the bias introduced by the inclusion of the second cohort is higher and its effect comes to dominate the effect of the variance at relatively low values of N_2.

In **Supplementary Figure 14b**, the R^2 of β_1 is assumed to be 0.01%, which corresponds to the estimated R^2 (without correction for the winner's curse) for our three genome-wide significant SNPs for SWB. Notice that the scale of the vertical axis differs in this figure, reflecting the fact that the RMSE is typically higher as a fraction of β_1. This is because the ratio of the bias of $\hat{\beta}_{meta}$ to β_1 is independent of β_1, but the ratio of the standard deviation of $\hat{\beta}_{meta}$ to β_1 is inversely proportional to β_1, so a lower assumed value of β_1 does not affect the bias but increases the standard deviation of $\hat{\beta}_{meta}$ relative to β_1, thus increasing the RMSE relative to β_1. The implications of **Supplementary Figure 14a** are similar to those of **Supplementary Figure 14b**: when the genetic correlation is high ($r_{g,LD} = 0.95$ or 0.8), including the second cohort decreases the RMSE and is thus desirable; when the genetic correlation is 0.6, it is desirable to include the second cohort if N_2 is less than ~90,000; if the genetic correlation is 0.3, it only makes sense to include the second cohort if N_2 is less than ~20,000.

In **Supplementary Figure 14c**, the true R^2 of β_1 is assumed to be 0.001%. Again, the scale of the vertical axis is shifted upward relative to the other panels, reflecting a typically higher RMSE as a fraction of β_1. Even for moderate values of the genetic correlations (e.g., $r_{g,LD} = 0.6$), it is always desirable to include the second cohort (at least when N_2 is less than 300,000). And even for a relatively low genetic correlation of

0.3, it is optimal to include the second cohort as long as N_2 is less than ~200,000. This result follows from the fact that the bias of $\hat{\beta}_{meta}$ scales linearly with β_1 but its variance does not depend on β_1, so the effect of the variance is relatively more important when β_1 is relatively small (e.g., with a true R^2 of 0.001%, or less); hence, though including a second cohort with a very imperfectly correlated phenotype measure increases bias, when β_1 is relatively small that effect is of second order relative to the reduction in the variance of $\hat{\beta}_{meta}$ that results from including the second cohort.

In sum, the solution to the Quantity-Quality Tradeoff depends on a number of key parameters. If we assume that $N_1 = 200{,}000$, that $\frac{h_2^2}{h_1^2} = 1$, and $Var(\beta_2) = 7.24 \times 10^{-7}$, and if the true R^2 of β_1 is 0.02% or 0.01%, it is desirable to include the second cohort with an imperfect phenotype measure if the genetic correlation between that imperfect measure and the phenotype of interest is very high or if the sample of the second cohort is small. If the true R^2 of β_1 is 0.001%, it almost always makes sense to include the second cohort, even if the genetic correlation is moderate (e.g., $r_{g,LD} = 0.3$). These results stem from the fact that, when the genetic correlation is high, N_2 is small, or the true R^2 of β_1 is small, the increase in the bias that results from including the second cohort is small relative to the reduction in the variance.

We conclude by noting that, while we have examined only two phenotypes and two cohorts, the framework we have developed here could be straightforwardly extended to multiple phenotypes and cohorts (although the derivations and the implementation would be tedious because we would have to consider the correlation between every pair of phenotypes and cohorts).

C. Technical appendix

In this sub-section, we derive two identities used in the derivations. Throughout, we maintain the same assumptions that are made by Bulik-Sullivan et al.[28]. In particular, we assume that the SNPs' true causal effects are randomly distributed across the genome, independent across SNPs, and independent of the SNPs' allele frequencies (when the SNPs have been standardized to have unit variance). As in the above derivations, we assume that all variables have been standardized to have zero mean and unit variance (except the error terms).

Proof that $\boldsymbol{Var(\beta_k) = \frac{h_k^2}{M} E[\ell] = \frac{E[\chi_k^2]-1}{N_k}}$

Following Bulik-Sullivan et al.[28], we assume that $y = Xb + u$, where *y* is the phenotype of interest, *X* is a vector of the standardized SNPs, *b* is a vector of the SNP's true average causal effects, and *u* is an error term. (We omit the subscript denoting phenotype *k* for notational simplicity.) Because the SNPs have unit variance, it follows that the true value of the GWAS coefficient for SNP *i* is given by

$$\beta_i = Cov(x_i, y) = Cov(x_i, Xb + u) = \sum_j Cov(x_i, x_j b_j) = \sum_j r_{ij} b_j ,$$

where r_{ij} is the correlation between SNPs *i* and *j*. It follows that the variance of β_i, conditional on the LD structure in the genome of the population of interest, is given by

$$\begin{aligned} Var(\beta_i|\text{LD structure}) &= Cov\left(\sum_j r_{ij} b_j , \sum_l r_{il} b_l \,|\text{LD structure}\right) = \sum_{jl} r_{ij} r_{il} Cov(b_j, b_l) \\ &= \sum_j r_{ij}^2 Var(b_j) = \left(\sum_j r_{ij}^2\right) Var(b_j) = \ell_i \frac{h^2}{M}, \end{aligned}$$

where $\ell_i = \sum_j r_{ij}^2$ is the LD Score of SNP *i*. The second equality follows because r_i. is deterministic conditional on the LD structure, and the first and second equalities on the second line follows from the assumption that the SNPs' true causal effects are randomly distributed, independent across SNPs, and independent of the SNPs' allele frequencies.

By the law of total variance,

$$\begin{aligned} Var(\beta) &= Var(E[\beta_i|\text{LD structure}]) + E[Var(\beta_i|\text{LD structure})] \\ &= 0 + E\left[\ell_i \frac{h^2}{M}\right] = \frac{h^2}{M} E[\ell], \end{aligned}$$

where the first term is equal to zero because $E[b_j] = 0$.

Also, from the LD Score framework, $E[\chi_i^2|\ell_i] \approx \frac{Nh^2}{M}\ell_i + 1$, so $E[\chi^2] \approx \frac{Nh^2}{M} E[\ell] + 1$, where $\chi_i^2 = N\hat{\beta}_i^2$ is the chi square statistics for SNP i from the GWAS. It follows that $Var(\beta) = \boldsymbol{\frac{h^2}{M}} E[\ell] = \frac{E[\chi^2]^{-1}}{N}$. In other words, we can obtain an estimate of $Var(\beta)$ from the mean of the χ^2 statistics from the GWAS and from the GWAS sample size *N*.

Proof that $\boldsymbol{Corr(\beta_1, \beta_2) = \frac{\rho_g}{\sqrt{h_1^2 \cdot h_2^2}} \equiv r_g}$

Observe that

$$
\begin{aligned}
Corr(\beta_1, \beta_2|\text{LD structure}) &= \frac{Cov(\beta_1, \beta_2|\text{LD structure})}{\sqrt{Var(\beta_1|\text{LD structure}) \cdot Var(\beta_2|\text{LD structure})}} \\
&= \frac{Cov\left(\sum_j r_{ij} b_{1j}, \sum_l r_{il} b_{2l}|\text{LD structure}\right)}{\sqrt{\sum_j r_{ij}^2 Var(b_{1j}) \cdot \sum_l r_{il}^2 Var(b_{2l})}} \\
&= \frac{\sum_j r_{ij}^2 Cov(b_{1j}, b_{2j})}{\sqrt{\sum_j r_{ij}^2 Var(b_{1j}) \cdot \sum_l r_{il}^2 Var(b_{2l})}} \\
&= \frac{\left(\sum_j r_{ij}^2\right) Cov(b_1, b_2)}{\left(\sum_j r_{ij}^2\right) \sqrt{Var(b_1) Var(b_2)}} = \frac{Cov(b_1, b_2)}{\sqrt{Var(b_1) Var(b_2)}} \\
&= \frac{\rho_g / M}{\sqrt{h_1^2/M \cdot h_2^2/M}} = \frac{\rho_g}{\sqrt{h_1^2 \cdot h_2^2}},
\end{aligned}
$$

where the second, third and fourth equalities follow because r_i. is deterministic conditional on the LD structure and from the assumption that the SNPs' true causal effects are randomly distributed, independent across SNPs, and independent of the SNPs' allele frequencies. (By assumption, $Cov(b_1, b_2) = \rho_g / M$ and $Var(b_k) = h_k^2 / M$ in the LD Score framework.)

Finally, because $Corr(\beta_1, \beta_2|\text{LD structure}) = \rho_g (h_1^2 \cdot h_2^2)^{-0.5}$ does not depend on the LD structure, it follows that $Corr(\beta_1, \beta_2) = Corr(\beta_1, \beta_2|\text{LD structure}) = \rho_g (h_1^2 \cdot h_2^2)^{-0.5} \equiv r_g$.

3. Auxiliary GWAS of depressive symptoms and neuroticism

A. Participating cohorts

Our meta-analysis of depressive symptoms ("DS") and neuroticism were conducted combining summary statistics from published meta-analyses with new genome-wide analysis. In our first auxiliary GWAS of neuroticism, we combine summary statistics from a meta-analysis conducted by the Genetics of Personality Consortium (GPC)[37] with our own analyses of UKB data. In our analyses of depressive symptoms ("DS"), we combine summary statistics from a GWAS of major depressive disorder performed by the Psychiatric Genomics Consortium (PGC)[38] with own analyses of data from two additional cohorts: the UKB and the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort[39].

Supplementary Table 9 provides a summary overview of the cohorts, their sampling frames, the auxiliary meta-analysis in which the cohort was used, details on how the studies restricted the estimation samples to European-ancestry subjects, and, in the case of publicly available data, the exact data file used in our analyses. PGC and GPC studies are conventional meta-analyses performed by pooling data from multiple cohorts of genotyped European-ancestry subjects. The GERA Cohort[39] is a genotyped subsample of participants in the Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH). GERA cohort members have been linked to administrative health records. Our fourth cohort, UKB, is a British prospective cohort study that targeted individuals aged between 40 and 69 and collected data on them between 2006 and 2010. Of the roughly 500,000 enrollees, data have been released for about 150,000 individuals as part of what is called the interim release[11].

B. Phenotypes

Supplementary Table 10 provides details on the phenotypes used in the auxiliary analyses, their distribution, the size of the final estimation sample, and, when applicable, the exact file with summary statistics from previously published studies used.

Depressive Symptoms

In the UKB, our measure of DS is constructed from responses to the following two questions:

1. "Over the past two weeks, how often have you felt down, depressed or hopeless?"
2. "Over the past two weeks, how often have you had little interest or pleasure in doing things?"

Both questions have five response categories: "Not at all", "Several days", "More than half the days", "Nearly every day" and "Do not know / Prefer not to answer". **Supplementary Table 9** describes how we mapped subjects' responses to these two questions into the continuous DS measure used in our analyses. We selected a relatively simple measure for two main reasons. First, we found exploratory LD score regressions[28] that our measure's SNP-based heritability ($h^2 \approx 0.05$) is not appreciably lower than the SNP-based heritabilities of more detailed measures available in UKB. Second, unlike other more detailed mental-health measures available in UKB data[40], our measure is available for nearly all "White-British ancestry" respondents ($N = 105{,}739$).

In the GERA cohort, our DS measure is binary and constructed using data on patient encounters at Kaiser Permanente Northern California facilities during January 1, 1995, to March 15, 2013. Participants are classified as having major depressive disorder if they had at least two diagnoses of depression on separate

days during this eighteen-year time window according to the ICD 9 CM classification system[41,42]. Our final estimation sample contains 7,231 cases and 49,137 controls, all of whom are of European ancestry.

Finally, the PGC[38] measure varied across cohorts, but classified as cases those individuals with a diagnosis of lifetime MDD satisfying the criteria in the Diagnostic and Statistical Manual of Mental Disorders (see table S2 in Ripke et al. 2013[38]).We use summary statistics from the meta-analysis of 9,240 cases and 9,519 controls.

We used bivariate LD score regression[28] to estimate the pairwise genetic correlations between the three measures used by the three cohorts. Overall, found high genetic correlations (UKB – PGC $\hat{r}_g$ = 0.797 [*SE* = 0.108], UKB – GERA $\hat{r}_g$ = 0.972 [*SE* = 0.216] and PGC – GERA $\hat{r}_g$ = 0.588 [*SE* = 0.242]).

Neuroticism

In the UKB, our measure of neuroticism is the respondent's score on a 12-item version of the Eysenck Personality Inventory Neuroticism scale[43]. This variable is available for 107,245 individuals. GPC[37] used a combination of neuroticism phenotypes harmonized across 29 discovery cohorts. The neuroticism batteries used by the cohorts varied, and included the NEO Personality Inventory, Eysenck Personality Questionnaire, and International Personality Item Pool inventory. The combined sample size of the contributing cohorts is *N* = 63,666.

The genetic correlation for neuroticism between UKB and GPC[37] estimated to be 1.112 (*SE* = 0.143) (**Supplementary Table 1**).

C. Genotyping and imputation

Three of the four studies (PGC being the exception) imputed their data against a 1000G reference panel. The genotyping and imputation of the UKB interim release data have been described extensively elsewhere[44]. The GERA genotype data were imputed using 2,504 individuals from the 1000G Phase 3 reference panel[45]. **Supplementary Table 10** summarizes information about the genotyping and imputation procedures in the cohorts included our auxiliary GWAS meta-analyses.

D. Association analyses

Supplementary Table 11 summarizes the association models used in the auxiliary analyses by listing the estimator used (logistic regression versus linear regression).

In the UKB analyses of DS and neuroticism, we followed the guidelines in the UKB's "exemplar GWAS" (see pp. 12-13 in Marchini[44]. We restrict the estimation sample to the "White-British ancestry" subsample and run linear regressions controlling for 15 principal components, indicator variables for genotyping array, sex, indicator variables for age ranges, and sex-by-age interactions.

The 29 cohorts contributing to the GPC meta-analysis all estimated linear regression models. The exact specifications varied across cohort, see p. 644 in de Moor et al.[37] for details on cohort-level controls. All cohorts controlled for sex and age and most sought to account for stratification by including controls for principal components. The exact number of PCs used appears to have been left at the discretion of the analyst.

In our analyses of the binary DS indicator in GERA, we ran logistic regressions with controls for four principal components of the genotypic data, sex, and 14 indicator variables for age ranges.

Finally, PGC estimated a logistic regression model that included controls for 5 PCs, sex, age, and cohort fixed effects. For details, see Ripke et al.[38]. Unlike the GPC study, the PGC investigators did not meta-analyze summary statistics but instead asked contributing cohorts to upload individual level phenotypic and genetic data.

E. Quality control

To each results file (including those obtained from publicly available summary statistics), we applied the quality-control filters described in **Supplementary Note 1.G** For the three cohorts with 1000G data, we use the reference files provided by EasyQC for 1000G data to harmonize the mapping from rsIDs to ChrPosIDs across results files and check for strand issues.[c] For PGC, we use the reference file whose construction was described in **Supplementary Note 1.G. Supplementary Table 12** reports the number of SNPs dropped in each filtering step.

F. Meta-analyses

Our meta-analysis of neuroticism was performed using sample-size weighting. In the meta-analysis of DS, we weight the UKB by sample size but to improve statistical power, we weight the two case-control studies by effective sample size, defined as:

(3) $$N_{eff} = \frac{4}{N_{cases}^{-1} + N_{controls}^{-1}}$$

as recommended by Willer et al.[29]. We thus ran the following two meta-analyses:

- DS (N = 180,866; N_{eff} = 149,707)
- Neuroticism (N = 170,911)

With the exception of the weighting scheme used in the DS analyses, these meta-analyses were performed exactly as were the SWB analyses. We used the software program METAL[29] to meta-analyze all SNPs that passed the quality-control filters, and we adjusted the standard errors of the resulting meta-analytic estimates by the square root of the estimated LD Score intercept. These adjustment factors were λ = 1.008 for depression and 1 for neuroticism (the actual point estimate being lower than one, λ = 0.9978).

G. Results

Depressive Symptoms

The QQ- and Manhattan plots for DS are provided in **Supplementary Figure 6a** and **Figure 1b**, respectively.

Supplementary Table 25 lists the 54 approximately independent SNPs that reach $p < 10^{-5}$ in our pooled meta-analysis of UKB, PGC and GERA. Two of these SNPs, *rs7973260* ($p = 1.78\times10^{-9}$) and *rs62100776* ($p = 8.45\times10^{-9}$) reach genome-wide significance. The first SNP (*rs7973260*) is available in UKB and PGC, with consistent signs of the effect. There is no reliable proxy tagging the SNP in GERA (the best available proxy has R^2 = 0.21). The second SNP (*rs62100776*) is only available in UKB. For this SNP, a high-LD proxy (*rs8099160*; R^2 = 0.971) is available in all three cohorts. This SNP reaches genome-wide significance in the

[c] http://homepages.uni-regensburg.de/~wit59712/easyqc/1000g/allelefreq.1000G_EUR_p1v3.impute_legends.noMono.noDup.noX.v2.gz. Accessed on 22 June 2015.

combined meta-analysis (p = 2.68×10^{-8}), though the effect in the GERA cohort is in the opposite direction to the estimated effects in UKB and PGC (though not significantly so).

A recent paper reports two genome-wide significant associations with recurrent depression in a case-control study of Chinese women[46]. The original report showed that the associations do not replicate in the PGC sample. We too were unable to replicate the associations in our larger meta-analysis sample.

Neuroticism

Supplementary Figure 6b and **Figure 1c** display the QQ- and Manhattan-plots of the GWAS meta-analysis on neuroticism, respectively.

Supplementary Table 25 shows the 117 approximately independent SNPs reaching p-value < 10^{-5} in the meta-analysis of neuroticism. Sixteen of these reach genome-wide significance.[d] Six of them tag a known inversion on chromosome 8, and one tags a known inversion on chromosome 17 (see **Supplementary Note 5** for additional analyses). None of the 117 SNPs we identify are in linkage disequilibrium with the genome-wide significant hit (rs35855737) reported by GPC[37]. However, in the independent UKB data, we find that the association with rs35855737 has the same sign and has a p-value of 0.064.

Robustness Analysis of Chromosome 18 SNPs

Two of the SNPs on chromosome 18 (rs1557341 and rs12961969) reach genome-wide significance in unconditional analyses and their pairwise linkage disequilibrium is below our cutoff of R^2 = 0.10. They therefore satisfy our definition of approximate independence. However, because the SNPs are in close physical proximity and in weak linkage disequilibrium (R^2 = 0.09), the evidence that they reflect independent genetic signals is less conclusive than for the other SNPs listed in **Table 1**.

To further investigate the robustness of the evidence that the SNPs reflect distinct genetic signals, we conducted two robustness analyses. In each of these, we sought to estimate the *conditional* effects of the two SNPs; that is, the effect of each SNPs conditioning on the allele count of the other.

In our first robustness analysis, we directly estimated the conditional effects in a multivariate association model with both SNPs included. These analyses require access to individual-level genotypic and phenotypic data and were therefore performed in the UKB sample. To maximize comparability with the meta-analysis estimates, we controlled for exactly the same set of control variables used in the main analysis (see **Supplementary Table 11**). Unlike the meta-analysis estimates, we use best-guess data for imputed SNPs instead of stochastic data for the probability of each genotype. For rs1557341, we estimate a conditional effect of 0.054 (SE = 0.0157, p = 6.38×10^{-4}) whereas for rs12961969, we estimate a conditional effect of 0.073 (SE = 0.0186, p = 9.42×10^{-5}). For point of reference, the unconditional effects of rs1557341 and rs12961969 are, respectively, 0.072 (SE = 0.0149, p = 1.20×10^{-6}) and 0.091 (SE = 0.0174, p = 1.38×10^{-7}).

In our second robustness check, we performed conditional and joint (COJO) multiple-SNP analyses[47]. An advantage of COJO is that it is a method that can generate estimates of conditional effects from summary statistics, thus increasing statistical power to detect independent signals. A potential downside of the approach is that in order for conditional estimates to be reliably estimable from summary statistics, an accurate estimate of the pattern of linkage disequilibrium between the SNPs in the samples that generated

the summary statistics is required. In practice, the linkage equilibrium is often estimated from a reference sample, and any differences in LD between the reference sample and the sample used in the GWAS that generated the summary statistics can cause biases.

In our COJO analysis, we used genotype data from the Health and Retirement Study, imputed to the 1000 Genomes Phase I reference panel and converted to best-guess genotypes, to estimate the linkage disequilibrium structure. To generate the sample we used to estimate the linkage equilibrium structure, we began with 8,652 genotyped respondents with European-ancestry individuals. From this sample, we filtered out SNPs with minor allele frequency below 1%, imputation accuracy below 0.7 or with Hardy-Weinberg equilibrium (HWE) *p*-values below 10^{-6}. These filters leave 7,779,969 SNPs. We subsequently estimated pairwise genetic relationships between all individuals using the directly genotyped SNPs, dropping one of each pair individuals with an estimated degree of relatedness above 0.025. This procedure leaves us with our final sample of 8,359 individuals.

Procedurally, COJO is implemented through a stepwise procedure in the software GCTA[48]. To convey the intuition, we give a sketch of the key steps of the algorithm here. To initiate the model, the lowest *p*-value SNP is selected from a given list of SNPs. The conditional effects (and *p*-values) of all remaining SNPs in the list are subsequently estimated using the LD structure from the reference sample. COJO then drops the SNP with greatest *p*-value (unless this SNP is genome-wide significant). The conditional effects of remaining SNPs are then estimated again, and the SNP with greatest *p*-value dropped (again, unless this SNP is genome-wide significant). The process is repeated iteratively until only genome-wide significant SNPs remain.

Applying COJO to the chromosome 18 GWAS summary statistics obtained from the neuroticism meta-analysis, we found that rs1557341 remained genome-wide significant ($p = 5.6\times10^{-9}$). In a joint model including rs1557341 and rs12961969, neither SNP reaches genome-wide significance (rs1557341 $p = 1.35\times10^{-5}$; rs12961969 $p = 5.4\times10^{-5}$). Thus the evidence that these SNPs reflect independent genetic signals is weaker than for the remaining SNPs identified in our main analysis (see **Table 1** in the main text).

4. Population stratification

Unaccounted for population stratification can be an important source of spurious findings in genome-wide association studies. All participating cohorts in our studies were asked to drop genetic outliers from their estimation, and, to guard against stratification by controlling for at least four principal components of the genetic-relatedness matrix (GRM). Several cohorts also used mixed-linear models that are known to be a very powerful method for accounting for stratification. Despite these many safeguards in place, we performed a number of additional analyses (including in the quality-control stage of the study) to test for the presence of any unaccounted-for stratification biases.

We used three methods to assess if the summary statistics from our primary GWAS of SWB exhibited signs of stratification bias. The three methods each rely on a different set of assumptions and thus allow us to be more confident that our results are not attributable to some specific assumption. First, we used the LD Score intercept method described in Bulik-Sullivan et al.[28]. Second, we conduct a GWAS/WF sign test, which examines how often the GWAS coefficients have the same sign as within-family (WF) estimates of the SNP effects. And third, we conduct a GWAS/WF Regression Test, a new method we developed which uses the within-family (WF) estimates to assess the fraction of variance of the GWAS estimates that can be attributed to stratification bias.

A. LD Score intercept test

The LD Score intercept test uses GWAS summary statistics for all measured SNPs. Unlike the Genomic Control (GC) method, which assumes that confounding bias (e.g., due to population stratification and cryptic relatedness) is responsible for inflation in the GWAS chi-square statistics, the LD Score regression method can disentangle inflation that is due to a true polygenic signal throughout the genome—which affects the slope of the LD Score regression—from inflation that is due to confounding biases such as cryptic relatedness and population stratification—which affects the intercept of the regression.

We use the LDSC software[28] to estimate the intercepts in LD Score regressions with the summary statistics of our GWAS of (i) SWB, (ii) DS, (iii) neuroticism, (iv) LS, and (v) PA. We estimated a separate LD Score regression for each of the phenotypes using the summary statistics from the meta-analyses based on all available data (as opposed to the summary statistics from the meta-analyses we conducted for the analyses described in **Supplementary Note 7**, which omit cohorts to avoid overlapping samples).

For each phenotype, we used the "eur_w_ld_chr/" files of LD Scores computed by Finucane et al.[49] and made available on https://github.com/bulik/ldsc/wiki/Genetic-Correlation. These LD Scores were computed with genotypes from the European-ancestry samples in the 1000 Genomes Project using only HapMap3 SNPs. Only HapMap3 SNPs with MAF > 0.01 were included in the LD Score regression.

Because GC will tend to bias the intercept of the LD Score regression downward, we did not apply GC to the summary statistics we used to estimate the LD Score regression[e]. Furthermore, we excluded the

[e] For SWB, LS, and PA, we could access the summary statistics for each cohort and did not apply GC for this analysis. For DS, the meta-analysis included data from the UKB cohort, from the Kaiser cohort, and from the PGC GWAS. We did not apply GC to the UKB, and we were told that GC was not applied to the other cohorts. For neuroticism, the meta-analysis included data from the UKB and the GPC GWAS, and a GC correction factor of 1.18 was applied to one cohort (SardiNIA) in the GPC GWAS prior to the

deCODE cohort from the data for the estimation of the LD Score intercept for SWB and PA, because the cohort-level regression estimates for deCODE did not directly correct for the high level of relatedness in the sample (their standard procedure is to apply GC). Consequently, including the estimates from deCODE would very likely have led to an intercept that is considerably upward biased. (The deCODE cohort is not included in the meta-analyses for DS, neuroticism, and LS.) Our intercept estimates from the LD Score regressions are thus unbiased measures of the amount of stratification there is in the data (or, for SWB, in the data excluding deCODE) that we used for the GWAS of each phenotype.

Supplementary Figure 7 shows LD Score regression plots based on the summary statistics from our three phenotypes. For SWB, we estimated a LD Score intercept of 1.012 (*SE* = 0.008); for DS, 1.008 (*SE* = 0.007); for neuroticism, 0.998 (*SE* = 0.011); for LS, 1.007 (*SE* = 0.007); and for PA, 1.011 (*SE* = 0.008). In all five cases, the intercept estimates are not significantly different from 1. By comparison, the mean χ^2 statistics for all the SNPs in the three LD Score regressions are 1.238 for SWB, 1.168 for DS, 1.317 for neuroticism, 1.138 for LS, and 1.139 for PA. (The mean χ^2 statistics reflect the average strength of the GWAS associations between the SNPs and each phenotype. Under the null hypothesis that there is no confounding bias and that the SNPs have no causal effects on the phenotypes, the mean χ^2 statistics would be one, so mean χ^2 statistics greater than one indicate that some SNPs are associated with the phenotypes.) These estimates imply that about 5% of the observed inflation in the mean χ^2 statistics for SWB, DS, and LS, about 8% of the inflation for PA, and a negligible share of the inflation in the mean χ^2 statistics for neuroticism, is accounted for by confounding bias (due to population stratification, cryptic relatedness, or other confounds) rather than polygenic signal. This suggests that the bulk of the inflation in the chi-square statistics for the five phenotypes is attributable to true polygenic signal throughout the genome, and that population stratification is unlikely to be a major concern for the analyses we present in this paper.

As previously described, we deflated all GWAS results (at the meta-analysis level, not at the cohort level) with correction factors estimated from a LD Score regression to produce our main GWAS estimates (rather than applying the usual GC correction). (For neuroticism, the LD Score intercept is less than one, so we did not apply any correction.) Therefore, the amount of inflation due to confounding bias is likely to be even smaller than these estimates in our main GWAS results (e.g., in the estimates for the genome-wide significant SNPs) for SWB, DS, LS, and PA.

B. Within-family sign test

A simple test relying on less restrictive assumptions than LD Score Regression is a comparison of the sign concordance between our GWAS estimates and a comparable set of within-family (WF) estimates. If the GWAS results are entirely driven by stratification, then the sign of the WF estimates, which are immune to stratification, should be independent of the sign of the GWAS estimates and therefore should only have a concordance of roughly 50%. A larger degree of sign concordance would suggest than at least some of the signal from the GWAS comes from true genetic effects.

---

meta-analysis; that cohort, however, accounts for less than 4% of the individuals in our meta-analysis for neuroticism, so this should induce an downward bias of less than 0.01 in the LD Score intercept for neuroticism.

To be more precise, let $\hat{\beta}_j$ denote the GWAS estimate corresponding to SNP j, and let $\hat{\beta}_{WF,j}$ denote the WF estimate corresponding to SNP j from a regression of sibling differences in phenotypes on sibling differences in genotypes and controls. That is, for phenotype y_i, genotype of SNP j, g_{ji}, and control vector x_i for individual i, we estimate by ordinary least squares:

$$\Delta y_i = \beta_{WF,j}\Delta g_{ij} + \Delta x_i\gamma + \Delta\varepsilon_i$$

where Δ is the sibling-difference operator. Since WF regressions are not biased due to stratification, and under the assumption that the WF effect of each SNP is equal to the population effect, we can decompose the GWAS and WF estimates as

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

where β_j is the true underlying GWAS parameter for SNP j, $\beta_{WF,j}$ is defined as above, s_j is the bias due to stratification (defined to be orthogonal to β_j and U_j), and U_j and V_j are the sampling variances of the estimates with $\mathrm{E}(U_j) = \mathrm{E}(V_j) = 0$. Note that if $\hat{\beta}_j$ and $\hat{\beta}_{WF,j}$ are estimated in independent samples, then U_j and V_j will be uncorrelated.

As our null hypothesis we assume that the GWAS contains no genetic signal. This is equivalent to the assumption $\beta_j = \beta_{WF,j} = 0$ for all j. Under the null, $\hat{\beta}_j = s_j + U_j$ and $\hat{\beta}_{WF,j} = \beta_j + V_j$ are independent and therefore have a sign concordance probability of 50%. This means that among a set of M independent SNPs, the number of SNPs that have a concordant sign, denoted *C*, is distributed

$$C \sim \mathrm{Binomial}(M, 0.5).$$

Using this known distribution, we can measure the observed sign concordance and formally test the null hypothesis that $\beta_j = 0$ for all j. We test this against the alternative hypothesis that $\beta_{WF,j}$ and β_j are not equal to zero and have concordant sign.

For this analysis, WF estimates for SWB were estimated individually in the MCTFR, NTR, STR-Twingene, and STR-SALTY cohorts. (Due to a lack of power, WF estimates for PA, LS, depressive symptoms and neuroticism were not calculated.) These estimates were then combined using a sample-size weighted meta-analysis. This gave us a WF sample size of 4869 sibling pairs (9738 individuals). The GWAS results were obtained by meta-analyzing the main results from this paper, omitting the WF cohorts. This leaves us with a GWAS sample size of 195,341 individuals.

Not every SNP was available in the WF samples. To maximize power, we restrict our SNPs to those that are available in all four WF cohorts. This leaves us with results on 2,039,455 SNPs. Using Plink's[50] clumping algorithm with the 1000G phase 1 reference sample composed of Utah Residents (CEPH) with Northern and Western European Ancestry (CEU), Toscani in Italia (TSI), and British in England and Scotland (GBR), a threshold of $P = 1$, and a LD threshold of $R^2 > 0.1$ in a 1,000,000 kb window, the subset is further reduced to a set of 112,884 approximately independent SNPs.

Power calculation

It is not immediately clear that the sign test would be well-powered using all available SNPs. Restricting to the set of SNPs that meet some significance threshold would leave us with a set of SNPs that individually

would be more likely to have a concordant sign in the absence of stratification. On the other hand, reducing the number of SNPs reduces the precision of the sign test, which increases the required fraction of concordant signs to pass the sign test.

To select the optimal SNP-inclusion threshold, we perform a power calculation for sets of independent SNPs meeting various *p*-value thresholds. In this simulation, we assume that $\beta_j = \beta_{WF,j}$ (i.e., no stratification). We note that for a set of M independent SNPs with true effect sizes of β_j, the expected sign concordance is

$$E(C \mid \beta) = \sum_{j=1}^{M} c\left(\beta_j, \sigma_j, \sigma_{WF,j}\right),$$

where $c\left(\beta_j, \sigma_j, \sigma_{WF,j}\right)$ is defined to be probability that SNP j will have a concordance sign in the WF and GWAS results, and σ_j and $\sigma_{WF,j}$ are the standard errors corresponding to SNP j for the GWAS and WF estimates, respectively. The function $c\left(\beta_j, \sigma_j, \sigma_{WF,j}\right)$ can be calculated as

$$c\left(\beta_j, \sigma_j, \sigma_{WF,j}\right) \equiv \Phi\left(\frac{\beta_j}{\sigma_j}\right)\Phi\left(\frac{\beta_j}{\sigma_{WF,j}}\right) + \left[1 - \Phi\left(\frac{\beta_j}{\sigma_j}\right)\right]\left[1 - \Phi\left(\frac{\beta_j}{\sigma_{WF,j}}\right)\right]$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. By this same notation, we note that the variance of C is

$$\mathrm{Var}(C \mid \beta) = \sum_{j=1}^{M} \mathrm{c}\left(\beta_j, \sigma_j, \sigma_{WF,j}\right)\left[1 - \left(\beta_j, \sigma_j, \sigma_{WF,j}\right)\right].$$

Of course, the true values of β_j are not known. Using the estimated GWAS results, we can calculate the posterior distribution of the effect size for each SNP according to calculations based on the framework in **Supplementary Note 8**. Then, the expected sign concordance is

$$E[E(C \mid \beta)] = E\left[\sum_{j=1}^{M} c\left(\beta_j, \sigma_j, \sigma_{WF,j}\right)\right]$$

$$= \sum_{j=1}^{M} \mathrm{E}\left[c\left(\beta_j, \sigma_j, \sigma_{WF,j}\right)\right],$$

and the expected variance is

$$E[\mathrm{Var}(C \mid \beta)] = E\left\{\sum_{j=1}^{M} c\left(\beta_j, \sigma_j, \sigma_{WF,j}\right)\left[1 - c\left(\beta_j, \sigma_j, \sigma_{WF,j}\right)\right]\right\}$$

$$= \sum_{j=1}^{M} E\left\{c\left(\beta_j, \sigma_j, \sigma_{WF,j}\right)\left[1 - c\left(\beta_j, \sigma_j, \sigma_{WF,j}\right)\right]\right\}.$$

This expression does not have a closed-form solution. However, it can be evaluated by simulation by generating for each SNP a large number of draws from the posterior distribution and evaluating the expectation and variance of the sign concordance for each set of draws. As the number of draws increases,

the mean of the calculated expectation and variance will converge to the true expectation and variance. In our calculations, we simulate the effect sizes with 1000 draws per SNP, assuming a prior with no mass at zero.

Finally, for a set of M independent SNPs, we calculate the minimum number of concordant signs required to pass the sign test at the 5% level of significance. We use a one-sided test since we are only interested in whether the concordance is greater than 50%. This means that the concordance threshold, t, can be calculated using the inverse CDF of a $\text{Binomial}(M, 0.5)$, evaluated at 0.05. Using the approximation that C is approximately normally distributed, the power of the sign test for our M SNPs is

$$Power = 1 - \Phi\left(\frac{E[E(C \mid \beta)] - t}{\sqrt{E[\text{Var}(C \mid \beta)]}}\right).$$

The results of this power calculation are shown in **Supplementary Figure 15**. As we can see, the power of the sign test in increasing as the *p*-value threshold used to filter SNPs increases. That is, it appears that the power gains due to increased precision by adding more SNPs outweighs the power losses due to adding SNPs with less signal. As a result of this analysis, we perform the sign test using the full set of independent SNPs.

Of the 112,884 SNPs used in this analysis, 57,384 of them had a concordant sign in the GWAS and WF results (50.83%). While this concordance may seem very low, it matches the expected sign concordance of 57,418 (50.83%) with a standard deviation of 168 SNPs. Furthermore, this level of concordance passes the sign test with a *p*-value of 1.04×10^{-8}.

Given the strength of these results, we reject the null hypothesis that our GWAS results are driven entirely by stratification. That is, at least some of the signal measured in the GWAS is due to genetic signal. We are not able to decompose the GWAS coefficients, however, into how much of the signal is due to stratification and how much is due to a genetic signal. We again highlight that, while this is a relatively weak conclusion, it is based on a much weaker assumptions than our other tests of stratification: that the WF effect size for each SNP is has the same sign as the effect through the population absent stratification. In contrast, LD Score Regression and the GWAS/WF Regression Test that follows estimate how much of the variance in the GWAS estimates is due to genetic signal versus stratification, though they rely on stronger assumptions.

C. Within-family regression test

In this section, we develop a GWAS/WF Regression Test to measure the amount of stratification in our GWAS estimates for SWB. This serves as a complement to the LD Score-based test, which relies on a different set of assumptions, and as an extension to the sign test. Simply put, it is a regression of the estimates from a WF GWAS onto the GWAS estimates from an independent sample, correcting for estimation error in the GWAS estimates. Since WF estimates are immune to stratification, the slope of this regression may be interpreted as the fraction of variation in the estimates that is explained by true genetic factors, subject to certain assumptions.

As in the sign test, let $\hat{\beta}_j$ denote the GWAS estimate corresponding to SNP j, and let $\hat{\beta}_{WF,j}$ denote the WF estimate corresponding to SNP j from a regression of sibling differences in phenotypes on sibling differences in genotypes and controls. As above, since WF regressions are not biased due to stratification,

and under the assumption that the WF effect of each SNP is equal to the population effect, we can decompose these estimates as

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

where β_j is the true underlying GWAS parameter for SNP j, s_j is the bias due to stratification (defined to be orthogonal to β_j and U_j), and U_j and V_j are the sampling variances of the estimates with $\mathrm{E}(U_j) = \mathrm{E}(V_j) = 0$. Note that if $\hat{\beta}_j$ and $\hat{\beta}_{WF,j}$ are estimated in independent samples, then U_j and V_j will be uncorrelated.

The quantity $\frac{\mathrm{Var}(\beta_j)}{\mathrm{Var}(\beta_j)+\mathrm{Var}(s_j)}$ would be informative about the degree of stratification: it measures the fraction of variance in the GWAS estimates that is due to true signal. To motivate an estimator for this quantity (described below), we first consider a naïve approach: a regression of $\hat{\beta}_{WF,j}$ on $\hat{\beta}_j$. The slope from this regression[f] would be

$$m = \frac{\mathrm{Cov}(\hat{\beta}_j, \hat{\beta}_{WF,j})}{\mathrm{Var}(\hat{\beta}_j)}$$
$$= \frac{\mathrm{Cov}(\beta_j + s_j + U_j, \beta_j + V_j)}{\mathrm{Var}(\beta_j + s_j + U_j)}$$
$$= \frac{\mathrm{Var}(\beta_j)}{\mathrm{Var}(\beta_j) + \mathrm{Var}(s_j) + \mathrm{Var}(U_j)}.$$

This expression is not quite the quantity we want: the term $\mathrm{Var}(U_j)$ in the denominator represents attenuation bias due to estimation error in the GWAS coefficients. To obtain the estimator we want, we can estimate $\mathrm{Var}(U_j)$ as the mean squared standard-error estimate of $\hat{\beta}_j$. That is, defining $\widehat{\sigma^2_{U_j}}$ as the squared standard-error estimate of $\hat{\beta}_j$, we can estimate $\mathrm{Var}(U_j)$ across a set of M SNPs as

$$\widehat{\mathrm{Var}(U_j)} = \frac{1}{M}\sum_{j=1}^{M} \widehat{\sigma^2_{U_j}}\ .$$

We can then use $\widehat{\mathrm{Var}(U_j)}$ to "correct" the quantity m by subtracting this term from the denominator, giving us the corrected slope estimator

$$\widehat{m_c} \equiv \frac{\widehat{Cov(\hat{\beta}_j, \hat{\beta}_{WF,j})}}{\widehat{Var(\hat{\beta}_j)} - \widehat{\mathrm{Var}(U_j)}}.$$

Under the null hypothesis that the GWAS coefficients are not biased due to population stratification (i.e., $\mathrm{Var}(s_j) = 0$), we have $\mathrm{E}[\widehat{m_c}] \to 1$ as $M \to \infty$. In contrast, in the presence of population stratification, $\mathrm{E}[\widehat{m_c}]$

[f] In practice, we used uncentered variances and covariances for this analysis, as both $\hat{\beta}_j$ and $\hat{\beta}_{WF,j}$ should have mean zero across SNPs, and uncentered statistics corresponds to a regression where the regression line passes through the origin. As a robustness check, we also conducted this analysis with centered variances and covariances and the results were nearly identical, as expected. Using uncentered variances and covariances also has the advantage of making this analysis invariant to the choice of reference allele.

will converge to some value less than one. In the extreme case in which the GWAS estimates capture no true signal and are entirely due to population stratification, E$[\widehat{m_c}]$ wil converge to zero.

Following the convention introduced by Bulik-Sullivan et al.[28], we estimate the standard errors using a block jackknife approach, where blocks consist of sets of approximately 2,000 adjacent SNPs. More precisely, each SNP is assigned to one of 1000 equally-sized blocks, with block one consisting of the first 2,039 SNPs on chromosome 1, block two consisting of the next 2,039 SNPs, and so on through the whole genome[g]. Then $\widehat{m_{c,b}}$ is estimated 1,000 times, omitting a different block of SNPs each time. The standard errors are then estimated to be

$$\mathrm{SE}(\widehat{m_c}) = \frac{999}{1000}\sum_b \left(\widehat{m_{c,b}} - \widehat{m_c}\right)^2.$$

As with the sign test, WF estimates were estimated individually in the MCTFR, NTR, STR-Twingene, and STR-Salty cohorts. These estimates were then combined using a sample-size weighted meta-analysis. This gave us a WF sample size of 4,869 sibling pairs (9,738 individuals). In contrast to the GWAS sample used in the sign test, the GWAS results used here omit the WF cohorts listed above and also omit the DeCODE cohort[h]. This leaves us with a GWAS sample size of 195,341 individuals.

Using these data, we estimate a corrected slope of 1.28 with a 95%-confidence interval of [0.72, 1.84]. As a reminder, the slope may be thought of as an estimate of $\frac{\mathrm{Var}(\beta_j)}{\mathrm{Var}(\beta_j)+\mathrm{Var}(s_j)}$, or more intuitively, the proportion of the variance of the GWAS coefficients (if they were estimated in an infinite sample) that is due to genetic signal rather than stratification. Since we do not constrain the slope to be between zero and one, this estimate of $\widehat{m_c}$ does not have a meaningful interpretation. If we censor this estimate to the range of zero to one, it suggests that that there is no stratification in our GWAS estimates. Additionally, the confidence interval suggests that with 95% confidence, at least 72% of the signal from our GWAS estimates is a result of genetic signal as opposed to stratification.

D. Discussion

We have presented three tests of stratification, each using a different set of assumptions and each suggesting that stratification plays a limited role in the estimated variation in our GWAS coefficients. The simplest test, the GWAS/WF sign test, allows us to reject the hypothesis that the GWAS results are driven entirely by stratification. Our two other tests, LD Score Regression and the GWAS/WF Regression Test, allow us to quantify how much stratification may be in our estimates. In both cases, the point estimates suggest that there is very little stratification in our estimates. We note, however, that for the GWAS/WF regression test method of assessing the amount of stratification, the standard errors are moderately large.

---

[g] Chromosome borders were ignored in assigning SNPs to blocks. That is, if as we added adjacent SNPs to a block, we arrived at the end of the chromosome, we continued adding SNPs to that block starting with the beginning of the next chromosome. As a result, some blocks contained SNPs from separate chromosomes.

[h] The deCODE cohort was also omitted from this analysis because their cohort-level regression estimates did not correct for the high level of relatedness in the sample (their standard procedure is to apply GC). As a result, the standard errors of the regression estimates are biased, making it impossible to compute unbiased estimates of Var(U_j).

Although there is now compelling evidence that cohort-level genomic control adjustment is overly conservative[28,31], we examined, for comparability with earlier studies, how many of the 16 genome-wide significant associations survive correction for genomic control. We found that all three SWB-associated SNPs, and both DS-associated SNPs, survived correction for genomic control. We also found that the two inversion-tagging SNPs and three of the other neuroticism lead-SNPs remained genome-wide significant, whereas the remaining 6 neuroticism lead-SNPs did not.

5. Long-range LD regions and inversion polymorphisms

Several of the genome-wide significant signals for neuroticism reside within a well-known inversion polymorphism on chromosome 8[32], and one resides within a well-known inversion polymorphism on chromosome 17[51,52]. In this section, we describe how the chromosome 8 inversion was called, how it was shown to be significantly associated with neuroticism, how that association was replicated in an independent dataset, and how it was annotated. We further describe the calling of a second inversion polymorphism on chromosome 17q21 and its association to neuroticism. As the inversion on chromosome 17 contains multiple allele-specific copy number variants (CNVs) and is generally complex in nature[51], our analyses of this inversion are less exhaustive.

In addition, one of the genome-wide significant signals from the SWB GWAS (rs3756290) resides within a previously identified long-range LD region of the genome, between base pairs 129M and 132M on chromosome 5[32]. Though long-range LD in that region could indicate the presence of an inversion polymorphism, we were unable to confirm the presence of an inversion polymorphism at this location. We therefore conclude that the long-range LD at this location is likely to have a different cause.

Due to genetic recombination during meiosis, the human genome is characterized by strong non-random associations between adjacent SNPs, i.e. linkage disequilibrium (LD). While LD generally exists between loci that are close (< 1Mb apart), some regions of the genome are characterized by long-range LD between loci (up to 4Mb apart). A potential cause for long-range LD is the presence of a so-called *inversion polymorphism.* An inversion polymorphism exists when a part of the genome on a chromosome is present in reverse order in some individuals. Such an inversion prevents recombination, and the inversion genotype in its entirety is transmitted from generation to generation. The lack of recombination in the inverted region causes strong LD between SNPs within the inversion. The presence of substantial long-range LD within an inverted region of the genome can be detected by principal component analysis (PCA)[32,53].

A. Analysis of Chr8 inversion polymorphism

Calling the inversion polymorphism

It is possible to accurately call known inversions with a PCA[54]. To determine whether the known inversion polymorphism on chromosome 8 between 8Mb and 12Mb[32,53] accounts for the associations between several SNPs in this region and neuroticism, we performed a PCA (using flashpca[55]) on 16,535 genotyped chromosome 8 SNPs passing quality control (minor allele frequency ≥ 0.05, call rate ≥ 0.95, and Hardy-Weinberg $p \geq 0.01$) in the UK Biobank cohort[11]. As can be seen from **Supplementary Figure 9**, the first principal component (PC) correlated strongly with SNPs between ~8Mb and ~12Mb on chromosome 8.

As can be seen from **Supplementary Figure 8a**, the histogram of the first PC distinguishes three subgroups of individuals in the UKB cohort and tags the three inversion genotypes: no inversion, heterozygous for the inversion, and homozygous for the inversion. The first PC is a noisy indicator of the inversion genotypes, but if we assume that the noise is normally distributed, the first PC can be modeled as a mixture of three normal distributions. We fitted a three-class normal mixture to the first PC for all UKB participants using the function "normalmixEM" included in the R package "Mixtools"[56]. The mixture results allow us to clearly distinguish three inversion genotypes (**Supplementary Fig. 8a**). We assigned the UKB individuals to the

inversion genotype for which they have the highest posterior probability based on the three-class mixture. The minor allele for the chromosome 8 inversion genotype has an 0.44 allele frequency.

We validated our inversion genotype calls by comparing them to inversion calls made using the InvClust software[57] for 2,000 randomly selected individuals in the UKB that were part of the GWAS on neuroticism. The concordance between inversion calls made with InvClust and the PCA/mixture method was 0.997 (**Supplementary Table 26**), thus raising our confidence in both sets of inversion calls. Given that InvClust is very computationally intensive, we used the calls based on the above-described PCA/mixture method for our main analyses in the UKB.

Tests of association with inversion polymorphism

To examine if the inversion is associated with neuroticism, we regressed continuous standardized neuroticism scores—residualized on the control variables age, sex, age × sex, the 15 top PCs of the genetic relatedness matrix computed from all SNPs, and genotype array—on the inversion genotype calls in the UKB cohort, and found that the inversion genotype is significantly positively associated with neuroticism (β = 0.035, *SE* = 0.0043, $p = 2.01\times10^{-15}$) (see also **Supplementary Table 13**). We also estimated the incremental R^2 of the inversion genotype (defined as the difference between the R^2 from a regression of neuroticism on the control variables and the inversion genotype, and the R^2 from a regression of neuroticism on just the control variables) to be 0.0006. (By comparison, the incremental R^2 of the 6 seemingly independent loci residing within the inversion polymorphism—defined similarly as the difference between the R^2 from a regression on the control variables and the 6 lead SNPs tagging those loci, and the R^2 from a regression on just the control variables—is also equal to 0.0006. We note, however, that this R^2 estimate is not necessarily comparable to the R^2 estimate for the inversion because both may be biased upward by the winner's curse, and the R^2 of the inversion is likely to be attenuated to the extent that our measure of the inversion genotype has measurement error.)

To test whether individual SNPs in the inversion region—defined as the region between base pairs 7,962,590 and 11,962,591, which are the borders reported in Price et al.[32] and in Abdellaoui et al.[58], and correspond to the borders present in the UKB data (**Supplementary Fig. 9**)—are significantly associated with neuroticism independently of the inversion, we performed an additional analysis of the SNPs in the inversion region, with the first PC described above as an additional covariate in the construction of the residualized neuroticism phenotype. No SNP in the inversion region remained genome-wide significant, indicating that the genome-wide signals from the main GWAS in that region are all attributable to the inversion (**Supplementary Fig. 1**). In addition, we repeated the GWAS separately within each of the three subsamples corresponding to the individuals with the three inversion phenotypes. Again, no SNP in the region was genome-wide significant in any of these three GWAS, further indicating that the genome-wide signals in that region are all attributable to the inversion.

Because the sign of the first PC tagging the inversion is arbitrary, we report in **Supplementary Table 27** the direction of the associations between our first PC and 20 SNPs within the inversion region that strongly tag the first PC. This information will allow future researchers to align the sign of the PC capturing the inversion in their dataset to the sign of our first PC, and thus to attempt to independently replicate our results, with the correct signs.

Testing the Chr8 inversion for association with neuroticism and SWB

To investigate if the association replicates in independent samples, we used two approaches.

Test 1: Tests of Association with Inversion-Tagging SNPs. To implement this test, we began by calculating the R^2 (i.e., the squared correlation) between each SNP in the region of the inversion and the first PC that tags the inversion, using European-ancestry in the UKB sample. We retained a list of SNPs whose pairwise R^2 with the inversion exceeded 0.75.

We were able to match 111 of these SNPs to the summary statistics supplied by the GPC. For each of these SNPs, we chose the reference allele to be the allele positively associated with the PC (and hence predicted to be associated with greater neuroticism, since the estimated impact of the PC on neuroticism is positive). Under the null of no association between the inversion and neuroticism in the GPC data, the *t*-statistic for any one of these SNPs as well as the mean *t*-statistic for these SNPs should not be significantly different from zero. We find that the *t*-statistics are in the range 0.65 to 3.85, with an average of 2.09.

Applying the same methodology to a meta-analysis of SWB cohorts not included in the main neuroticism GWAS (see **Supplementary Table 31** for a list of omitted cohorts), we were able to identify 53 matching SNPs (the lower match rate being a consequence of the fact that the primary SWB analysis HapMap Phase 2 SNPs were tested for association). Given the negative correlation between SWB and neuroticism, we expect negative *t*-statistics, and that is indeed what we observe. We find that the *t*-statistics are in the range -1.68 to -4.09, with an average of 2.80. For SWB and neuroticism separately, **Supplementary Table 29** reports results for the 50 matched SNPs whose pairwise R^2 with the PC was the highest.

Although the matched SNPs are all correlated with one another (because they are all highly correlated with the PC capturing the inversion), **Supplementary Figure 11** shows that the *t*-statistics for the SNPs tend to be consistently in the predicted direction throughout the inversion region—and not only at a single location in the inversion region—thereby further suggesting that the significance of the associations between these SNPs and both SWB and neuroticism is truly attributable to the inversion rather than to the effect of a single locus. Altogether, the evidence from Test 1 strongly suggests that the inversion is also significantly associated with neuroticism and SWB in cohorts independent of the sample in which it was identified.

Test 2: Regression of the GPC et al. GWAS Estimates. Our second test relies on a number of structural assumptions. We assume that the first PC based on the SNPs in the inversion region (which captures the inversion, as detailed above) has a direct causal effect on neuroticism and that any given SNP x can also be associated with neuroticism independently of its correlation with the first PC: formally, we write

$$E[n|PC, x] = PC \cdot \beta_1 + x \cdot \beta_2,$$

where n denotes neuroticism, PC is the first PC and where β_2 is independent of *PC* and of SNP *x*'s correlation with *PC*. Thus, if one regresses n on SNP *x*,

$$n = x \cdot \gamma + \epsilon,$$

as in the GPC GWAS, then

$$\hat{\gamma} = \frac{Cov(n, x)}{Var(x)} = \frac{Cov(PC \cdot \beta_1 + x \cdot \beta_2 + \varepsilon, x)}{Var(x)}$$

$$= \frac{Cov(PC, x)}{Var(x)}\beta_1 + \frac{Cov(x, x)}{Var(x)}\beta_2$$
$$= \frac{Corr(PC, x)}{SD(x)}\beta_1 + \beta_2,$$

where *SD(x)* denotes the standard deviation of *x* and where the last equality holds because we standardized *PC* so it has unit variance.

Because β_2 is independent of *PC* and of $Corr(PC, x)$, it follows that regressing $\hat{\gamma}$ on $\frac{Corr(PC,x)}{SD(x)}$,

$$\hat{\gamma} = \frac{\widehat{Corr(PC}, x)}{SD(x)}\beta_1 + u,$$

should yield an unbiased estimate of β_1—i.e. of the effect of *PC* on n.

We therefore estimated a regression of the GWAS estimates reported by GPC (i.e., of $\hat{\gamma}$) on estimates of $\frac{Corr(PC,x)}{SD(x)}$, where $SD(x)$ was estimated from the summary statistics from GPC and where $Corr(PC, x)$ was estimated using the individuals of European ancestry in the UKB as a reference panel (since we cannot estimate this quantity from the GPC summary statistics).

Using all 12,494 SNPs that are in the inversion region on chromosome 8 and for which the GPC study supplied summary statistics, we obtained an estimate of $\widehat{\beta_1} = 0.0063$, indicating that a one standard deviation increase in the PC that tags the inversion increases neuroticism, as measured by GPC, by 0.0063 unit—although the associated standard error, *t*-statistic, and *p*-value from that regression are not valid since the error terms (given by u, which under the model is equal to β_2) will be highly correlated for nearby SNPs.

(Unfortunately, GPC do not report the standard deviation of their measure of neuroticism, so it is not possible to use that to directly translate our estimate of $\widehat{\beta_1}$ into an estimate of the implied R^2 of the PC tagging the inversion in the de GPC data. However, the summary statistics supplied by GPC include data on the GWAS estimates' standard errors and minor allele frequencies for all SNPs; using the approximation[i]

$$\widehat{\sigma_{neuro}} \approx \widehat{\sigma_{\beta_j}} \cdot \sqrt{2 \cdot N \cdot MAF_j \cdot (1 - MAF_j)}$$

(where $\widehat{\sigma_{neuro}}$ is the standard deviation of neuroticism, $\widehat{\sigma_{\beta_j}}$ is the GWAS estimates' standard errors for SNP *j*, and MAF_j is the minor allele frequency of SNP *j*), assuming that the sample size was $N = 60{,}000$[j], and taking the mean of $\widehat{\sigma_{neuro}}$ across all SNPs, we obtain an estimate of $\widehat{\sigma_{neuro}} \approx 1.07$. Based on this, the R^2 of the PC tagging the inversion in the GPC data is approximately $R^2 \approx \left(\frac{\widehat{\beta_1}}{\widehat{\sigma_{neuro}}}\right)^2 = \left(\frac{0.0063}{1.07}\right)^2 \approx 3.5 \times 10^{-5}$.)

Under the assumption that the error terms of the regression (given by u, which under the model is equal to β_2) are uncorrelated for sufficiently distant SNPs, it is possible to obtain unbiased estimates of the

---

[i] This approximation follows from the fact that $\widehat{\sigma_{\beta_j}} = \widehat{\sigma_u}/\sqrt{2 \cdot N \cdot MAF_j \cdot (1 - MAF_j)}$ (where $\widehat{\sigma_u}$ is the standard deviation of the residual from the regression of neuroticism on SNP *j*) and that $\widehat{\sigma_u} \approx \widehat{\sigma_{neuro}}$, since each SNP *j* has a very small effect size.

[j] The GPC GWAS had a total sample size of N = 63,661, but for many SNPs data was missing for some of the 29 cohorts.

standard error of $\widehat{\beta_1}$. This assumption may be reasonable for SNPs in most regions of the genome, since β_2 captures the SNPs' effects independent of *PC*, and distant SNPs' effects should typically be uncorrelated conditional on *PC.* Unfortunately, however, this assumption may fail for SNPs in the inversion region because, as mentioned above, there is strong LD between SNPs within the inversion due to the lack of recombination in the inverted region.

With that caveat in mind, we nonetheless estimated the regression again for several subsets of SNPs that are distant from one another. For the first subset, for each base pair position that is a multiple of 10,000 base pairs and that falls in the range of the inversion on chromosome 8, we included the closest SNP that was in the GPC summary statistics (unless there were no SNPs within 1,000 base pairs of that position in those summary statistics). This left 367 SNPs that are at least 8,000 base pairs away from any other SNP in the subset. We formed the second, third, and fourth subsets analogously, but by including SNPs for base pair positions corresponding to multiples of 25,000, 50,000, and 100,000 base pairs respectively; this yielded sets of 144, 71, and 36 SNPs that are at least 23,000, 48,000, and 98,000 base pairs away from any other SNPs in the respective subsets.

These results are reported in **Supplementary Table 29**. As can be seen, the estimates of $\widehat{\beta_1}$ are around ∼0.0060 and are significantly positive at the 10% level (in one-tailed tests) for all subsets of SNPs, thus further suggesting that the PC capturing the inversion is significantly associated with neuroticism in the GPC data.

Biological annotation

To elucidate the mechanism underlying the association between the inversion polymorphism and neuroticism, we investigated whether the inversion influences gene expression in two independent cohorts (described in more detail below).

We found that the inversion does influence gene expression, especially for genes near the inversion breakpoints. Because the inversion reorients a large portion of DNA (4Mb) and because coding sequences of genes can be separated from functionally active regulatory sequences (like enhancers and modulators) by very long distances, the effect of the inversion on the expression levels of genes flanking the breakpoints of the inversion is a molecular mechanism that could plausibly account for association between the inversion and neuroticism.

Effects of the Chr8 inversion on gene expression

We conducted gene expression analyses of whole peripheral blood from a total of 2,360 unrelated individuals: 1,240 individuals from the Fehrmann cohort measured with the Illumina HT12v3 platform[59]; 229 individuals from the Fehrmann cohort measured with the Illumina H8v2 platform[59]; and 891 individuals from EGCUT cohort measured with Illumina HT12v3[60].

The quality control procedures for both genotype data and gene expression profiles have been described in detail elsewhere[60]. We used the invClust software[57] to call the polymorphic inversion on chromosome 8 in our sample of 2,360. The three cohorts were called separately using default parameters to avoid batch effects due to differences in array design and underlying population structure.

Gene expression data went through comprehensive processing. In short, expression probe intensities were quantile-normalized to the median distribution and subsequently $\log_2$-transformed. Probe and individual

means were centered to zero. Gene expression data were then corrected for possible population structure as well as for possible technical artifacts using principal components derived from the genotypic data (structure) and expression profiles (artifacts). More details can be found in Westra et al.[60].

In order to detect the effect of the inversion on RNA-expression levels of genes within the inversion and its flanking breakpoints, we followed the *cis*-eQTL mapping strategy described in Westra et al.[60]. In brief, we tested for an association between inversion genotype (NI/NI, I/NI, I/I; NI – not inverted, I – inverted) and gene expression value for all genes within the inversion and positioned ±1Mb from the inversion breakpoints (Start: 7,934,925; End: 11,824,441 as called in the Fehrmann and EGCUT cohorts)[k], as detected by the invClust software. We meta-analyzed the three datasets, weighting by sample size. We then permuted the sample labels and repeated this analysis 100 times. We set the FDR to 0.05, using the permuted datasets to ascertain the null distribution of the nominal *p*-value.

We applied the above-described methodology and found that the inversion was significant *cis*-eQTLs for seven genes (eight array probes). Results are presented in **Figure 4b** and in **Supplementary Table 24**. All affected genes are positioned in close proximity to the inversion breakpoints, which points to a clear potential molecular mechanism: the relocation of regulatory sequences.

In total, six of these genes are protein-coding (*BLK, FDFT1, c8orf12/c8orf13, MFHAS1, MSRA, MTMR9*). Here, we briefly describe their known or predicted functions. In order to gain insight into tissue- and cell type-specific expression, we queried the Gene Network co-expression database[61] (described in more detail in **Supplementary Note 9.E**). Here, we find that three genes (*FDFT1, MSRA, MTMR9*) are highly expressed in tissues and cell types that belong to the central nervous system. The other two (*BLK, MFHAS1*) are strongly expressed in tissues and cell types involved in the immune system. Only *c8orf13* and *c8orf12* were not present in that database (see **Supplementary Note 5.A** for a detailed description).

Two genes have clearly described functions in the literature: *FDFT1* encodes an enzyme involved in lipid synthesis and metabolism; it is mainly important for cholesterol synthesis (note that cholesterol is an essential component of the cell membrane and myelin sheaths, and the precursor of vitamin D and all steroid hormones). It is highly expressed in neural stem cells, and a mouse knockout model showed that it might be crucial for the development of the central nervous system[62]. *BLK* is widely involved in immune signaling. It encodes a protein known as “B lymphocyte kinase”, and is important in B-lymphocyte development, differentiation and signaling. However, it may also be involved in the upregulation of insulin synthesis in response to glucose[63]. *BLK* (as well as *c8orf13)* has been linked to the autoimmune diseases Systematic Lupus Erythematosus and Kawasaki disease in several small-scale GWAS[64–66].

The other genes’ functions are less well understood. *MSRA* is highly expressed in the cerebellum, motor neurons, and visual cortex, among other sites. It is highly conserved across species, and encodes methionine sulfoxide reductase: an enzyme that is thought to be important in protein repair following damage caused by oxidative stress[67,68]. *MSRA*-knockout in mice[67] and *MSRA*-overexpression in *Drosophila*[68] have been found to shorten and extend lifespan, respectively. *MTRM9* is highly expressed in many regions

[k] The exact breakpoints of the inversion polymorphism differ between the UKB data and the Fehrmann/EGCUT datasets, because two different methods were used to call the inversion in the two datasets (the PCA/mixture method in the UKB versus InvClust in the Fehrmann/EGCUT dataset). InvClust reports a border whereas the PCA/mixture method does not.

of the brain, as well as in neural stem cells. It is a member of the myotubularin-related proteins family, members of which are highly conserved across evolution. Roughly speaking, these proteins remove phosphate groups from phospholipids (which are integral molecules of the cell membrane). They are thought to be important in the development and maintenance of muscle cells. *MFHAS1* (previously known as *MASL1*) was first discovered due to its high expression in a type of soft-tissue tumor[69]. Its high expression in the spleen, macrophages and dendritic cells hints at importance for immune functioning.

To conclude, we note that it is not directly apparent how each of these genes may contribute to the etiological mechanisms involved in a psychological trait such as neuroticism. However, the high levels of central nervous system expression of some of the genes involved may hint at their hitherto unidentified importance for neural functioning, and thus for neuroticism.

B. Analysis of chr17q21 inversion polymorphism

A second significant signal in the neuroticism GWAS reported in **Table 1**, rs193236081, is located within a complex inversion polymorphism on chromosome 17q21.31[51]. Unlike the straightforward biallelic inversion polymorphism in chromosome 8, this inversion has multiple haplotypes, some of which co-occur with relatively common copy number variations[51]. We here attempt to call the inversion using PCA/Mixture, but we note that this will likely not capture all genetic variation (such as copy number variation) in the region. We associated the called inversion with neuroticism, depressive symptoms and subjective well-being in UK Biobank data.

We performed a PCA (using flashpca[55]) on 11,038 genotyped SNPs on chromosome 17 passing quality control (minor allele frequency ≥ 0.05, call rate ≥ 0.95 and Hardy-Weinberg $p \geq 0.01$) in the UK Biobank cohort. As can be seen from **Supplementary Figure 8b**, the histogram of the first PC distinguishes three subgroups of individuals in the UKB cohort, which suggests this PC tags the three inversion genotypes: no inversion, heterozygous for the inversion, and homozygous for the inversion. We fitted a three-class normal mixture to the first PC for all UKB participants using the function "normalmixEM" included in the R package "Mixtools". The mixture results allow us to clearly distinguish three inversion genotypes (**Supplementary Fig. 8b**). We assigned the UKB individuals to the inversion genotype for which they have the highest posterior probability based on the three-class mixture. The assigned genotypes suggest an inversion minor allele frequency of 0.22.

We then performed a GWAS of neuroticism using the SNPs on chromosome 17 and controlling for the first PC. This analysis revealed that the genome-wide signal, captured by rs193236081, indeed disappears when controlling for the first PC. To further test whether the PC indeed tags this inversion, for each SNP on chromosome 17, we compared its association strength with its squared correlation with the PC. **Supplementary Figure 10** displays the genetic correlation between each SNP on chromosome 17 and the PC tagging the inversion, together with a local Manhattan plot (not controlling for the PC tagging the inversion). The GWAS signals appear to track the strength of association between the SNPs and the inversion.

We then regressed continuous standardized neuroticism scores—residualized on age, sex, age × sex, the 15 top PCs of the genetic relatedness matrix computed from all SNPs in the genome, and genotype array—on the inversion genotype calls in the UKB cohort. We found a significant association between neuroticism and the inversion genotype ($\beta = 0.031$, $SE = 0.005$, $R^2 = 0.00033$, $p = 1.25\times10^{-9}$; see **Supplementary Table**

13). No significant associations were found for either depressive symptoms (p = 0.416) or subjective well-being (p = 0.605).

Because the sign of the first PC tagging the inversion on chromosome 17 is arbitrary, we report in **Supplementary Table 28** the direction of the associations between our first PC and 20 SNPs within the inversion region that strongly tag the first PC. This information will allow future researchers to align the sign of the PC capturing the inversion in their dataset to the sign of our first PC, and thus to attempt to independently replicate our results, with the correct signs. We believe the analysis of the Inversion polymorphism on chromosome 17 provides strong support for the hypothesis that neuroticism is associated to structural variation in this genomic region. However, given the complex nature of the genetic variation in this region, it is difficult to determine whether it is the inversion alleles or CNVs within these alleles that are responsible for the observed association between neuroticism and rs193236081. For enrichment and biological annotation of this region in other sections, we used rs193236081 as a proxy.

Testing the Chr17 inversion for association with neuroticism and SWB

We used Test 1 (see **Supplemental Note 5.A)** to investigate if the inversion on chromosome 17 is associated with neuroticism and SWB. Using the same procedure described in **Supplemental Note 5.A**, we identified 55 SNPs in the PGC data and 8 SNPs in the non-overlapping SWB sample whose pairwise correlation with the PC tagging the inversion exceeded 0.75. Of the 55 SNPs available in PGC, 52 had the expected (positive) sign (as expected, given that the PC capturing the inversion polymorphism is associated greater neuroticisim, see **Supplementary Table 13**). Across the 55 matched SNPs, the average *t*-statistic was 1.75. For SWB, we were only able to identify 8 matching SNPs, whose t-statistics ranged from -0.68 to 1.08, with an average of 0.38 (the opposite sign to that expected). We conclude on the basis of Test 1 that there is suggestive evidence that the inversion on chromosome 17 is also associated with neuroticism in the GPC sample, but not with SWB.

C. Long-Range LD region on chromosome 5

One of the three SWB-associated SNPs, rs3756290, is located in a previously identified long-range LD region of the genome, between base pairs 129M and 132M on chromosome 5[32].

Long-Range LD can arise for various reasons. One possible reason is an inversion (indeed, as we have shown above, some of the significant signals on chromosomes 8 and 17 from our neuroticism GWAS are in long-range LD regions where inversions are present). To examine if SNP rs3756290 also resides within an inversion polymorphism, we conducted analyses similar to those summarized above for the inversions on chromosomes 8 and 17.

First, we performed a PCA (using flashpca[55]) on the 18,997 genotyped chromosome 5 SNPs passing quality control (minor allele frequency ≥ 0.05, call rate ≥ 0.95, and Hardy-Weinberg p ≥ 0.01) in the UK Biobank cohort[11]. Unlike what we found from analogous exercises for the chromosomes 8 and 17 inversions, none of the top 10 PCs clearly distinguishes three genotypes (which could have corresponded to the three inversion genotypes if there were an inversion in the long-range LD region).

Second, to obtain a higher resolution of the structure of the long-range LD region, we performed a localized PCA on the 441 genotyped chromosome 5 SNPs between base pairs 128M and 133M passing the same quality controls (also in the UK Biobank cohort). We found that the top PC from this local PCA

strongly correlates with the top PC from the PCA performed on all of chromosome 5, suggesting that the latter tags the long-range LD region. However, we were not able to clearly distinguish three distinct genotypes from the top PC from either PCA, suggesting that neither clearly tagged an inversion. We interpret these results as consistent with the presence of the long-range LD region between base pairs 129M and 132M on chromosome 5 in our UKB data, but not necessarily with the presence of an inversion in that region. Furthermore, a literature search yielded no previous mention of an inversion at this location.

As has been discussed in previous work[70], other possible causes for long-range LD regions include population admixture, drift and recurrent bottlenecks, epistatic selection, and the hitchhiking of linked sites with a positively-selected mutation.

6. Polygenic prediction

Here we tested how well a polygenic score for SWB, based on the GWA meta-analysis results, could predict life satisfaction (LS), positive affect (PA) and our composite SWB measure in two independent holdout cohorts: the Health and Retirement Study (HRS[71]) and the Netherlands Twin Register (NTR[72,73]). Additionally, we tested how well our polygenic score for SWB could predict different personality traits as measured by the NEO Big Five[74]: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. Furthermore, because SWB has a strong negative phenotypic correlation with depression, we tested whether the polygenic score for SWB could predict DS. In NTR, DS was measured using the Achenbach System of Empirical Based Assessment (ASEBA) DSM-oriented scale[75]. In HRS, the CIDI SF assessment was used to measure DS. Finally, for comparison purposes, we also tested how well our polygenic score could predict height, a phenotype that showed practically no phenotypic correlation with SWB but is also highly polygenic.

A. Phenotype measures by cohort

Life satisfaction and positive affect

HRS: LS and PA questions were administered to 8,248 and 8,285 genotyped participants in four waves. LS was measured using the Satisfaction with Life Scale consisting of five items (e.g., "In most ways my life is close to ideal"), and responses were given on a six-point scale[12]. The PA measure differs across waves. In 2006, it was measured using eight questions (e.g., "During the past thirty days, how much of the time did you feel...extremely happy?") from the Midlife Development Inventory[76], which was adapted from some well-known instruments such as the Affect Balance Scale[77], the University of Michigan's Composite International Diagnostic Interview[78], the Manifest Anxiety Scale[79], the Health Opinion Survey[80], the General Well-Being Schedule[81], and the Center for Epidemiological Studies Depression Scale[19]. Responses were given on a five-point scale. In 2008, 2010 and 2012, it was measured using thirteen questions (five-point scale). Eleven of these questions were obtained from Positive and Negative Affect Schedule—Expanded Form (PANAS-X)[82], and the remaining two were chosen from two other studies in this area[83,84]. For both LS and PA, a score was constructed for each time point by taking the mean across items, and the score was set to missing if more than half of the items were unanswered.

NTR: LS and PA data were available in 8,165 and 6,320 genotyped participants. LS was measured longitudinally using the Satisfaction with Life Scale consisting of five items (e.g., "My life is going more or less as I wanted") with responses given on a seven-point scale, resulting in a minimum score of five and a maximum score of 35[12]. PA was also measured longitudinally using four questions that were adapted from the Subjective Happiness Scale[14] (e.g., "On the whole, I am a happy person") with responses on a seven-point scale, resulting in a minimum score of four and a maximum score of 28. Seventy percent of the PA and LS samples consist of participants of the Adults Netherland Twin Register (ANTR[72]), and the remaining thirty percent originate from the longitudinal survey study of the Young Netherlands Twin Register (YNTR[73]). The participants from ANTR were invited to participate in surveys including PA questions in 2002 and 2009, and LS questions in 2002, 2009 and 2013. The participants from YNTR were invited to participate in both surveys at ages 14, 16, 18, and 20, followed by another invitation to participate in the LS survey at age 22. For both LS and PA, a score was constructed for each time point by taking the sum across items,

and was set to missing if more than two of the items were unanswered. When subjects have a missing score on one item of the scale, this missing value was replaced by the subject's mean scale score.

Personality

HRS: The Big Five personality traits (neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness) were measured in four waves between 2006 and 2012, with 26 items in 2006-2008 and 31 in 2010-2012. The original 26 items in 2006 and 2008 were obtained from the MIDUS survey[76]. Extraversion (*N* = 8,271), Agreeableness (*N* = 8,271), and Conscientiousness (*N* = 8,268) were measured with five items, Openness to Experience (*N* = 8,253) with seven, and Neuroticism (*N* = 8,264) with four. In 2010 and 2012, five items from the International Personality Item Pool[85] were added to the "Conscientiousness" sub-dimension. The responses were given on a four-point scale[74]. For each trait, scores were constructed by taking the mean of all items in the respective category after recoding opposite-stated items. A score was set to missing if more than half of the items in that category had missing values.

NTR: Personality data were available for 7,596 genotyped participants. The Big Five personality traits were measured in three waves (2004, 2009, 2013) by using the NEO-FFI[74], a sixty-item personality questionnaire consisting of five subscales: neuroticism, extraversion, openness, agreeableness and conscientiousness. The responses were given on a five-point scale (0-4). Subscale scores were constructed for each time point by taking the sum across the twelve subscale-specific items (after recoding opposite-stated items), and was set to missing if ten or more items of the total scale were unanswered. When subjects have fewer than ten missing items, scores were replaced by a score of two (which is neutral given the 0-4 scale).

Depressive symptoms

HRS: DS was measured in ten waves (1995-2012) for 8,617 individuals using the CIDI-SF questionnaire, which implements the diagnostic criteria of DSM. The CIDI-SF starts with a small number of screen questions that are used to skip out participants least likely to have depressive symptoms. If the screen questions are endorsed for the necessary intensity and duration, participants are asked about seven symptoms. Summary scores equal to the number of symptoms (ranging from zero to seven) are assigned to all respondents that endorse one of the screen questions. If one or more symptom questions are unanswered, we set the summary score to missing. Instead of assigning a diagnosis of depression based on the number of symptoms, we followed the strategy of assigning a probability of "caseness" to each summary score as detailed by Nelson et al. (1998)[86]. The probabilities of caseness that correspond to the summary scores are derived from the US National Comorbidity Survey[78] and reflect the probability that a respondent with a certain response profile would meet the diagnostic criteria if given the full CIDI interview. If a respondent was screened out of the interview because of not endorsing any of the screen questions for the necessary duration or intensity, s/he was assigned a probability of caseness equal to zero.

NTR: DS was obtained from the DSM-oriented Depression subscale of the age-appropriate survey from the ASEBA taxonomy[75] and were available in 7,126 participants. To measure DS, fourteen questions were used (e.g., "Doesn't eat well") and responses were given on a three-point scale ranging from zero ("not true") to two ("very true"). The DSM-oriented subscale was constructed for each time point by taking the sum across the fourteen subscale-specific items and was set to missing if more than twenty percent of the total survey items were unanswered. When the missingness rate is less that twenty percent for a participant, the missings were replaced by the participant's mean score. Seventy percent of the sample consists of

participants of the ANTR and were invited to participate in surveys including the age-appropriate survey of the ASEBA taxonomy in 1991, 1995, 1997, 2000, 2009 and 2013. The remaining thirty percent of the participants originate from the longitudinal survey study of the YNTR and were invited to participate at ages 14, 16, 18, 20, and 23.

Height

HRS: Height data were based on self-report, available for 8,650 HRS participants. All participants were older than eighteen years old.

NTR: Height data were based on self-report and laboratory measurement and were available for 8,619 NTR participants. Over the past 25 years, participant have been invited to self-report height in 1991, 1993, 1995, 1997, 2000, 2002, 2004, 2009, 2011, and 2013. Furthermore, subgroups have been invited to take part in one or multiple of the seven laboratory studies. The height measure is based on an average score over all repeated assessments. When a reported or measured height deviated more than two cm's from earlier reported height, the score for that specific occasion was set to missing.

The final phenotypes for LS, PA, Big Five personality traits, and DS in HRS and NTR were constructed by first grouping the respondents based on which combination of waves they have responded in, and then taking the standardized residuals from a regression, within each group, of the phenotypic score on sex, age, age^2, and all interactions. For the individuals who have responded in multiple waves, the average phenotypic score was used for obtaining the residuals. The composite SWB measure (N_{HRS} = 8,226, N_{NTR} = 6,314) was created by taking the average of the residualized PA and LS scores when both are available, and set to missing if either PA or LS is missing. Height was similarly constructed, but no grouping was made based on number of responses. In HRS, the first height measurement available and in NTR, the mean height measurement was residualized on sex, birth year, birth year squared, and interactions.

We ran meta-analyses of the pooled SWB phenotype excluding each of the holdout cohorts, applying a minimum sample size filter of 100,000 individuals. Using these summary statistics, we constructed two sets of polygenic scores: (1) LDpred polygenic scores (LD-PGS) using LDpred effect sizes[87], and (2) linear polygenic scores (Lin-PGS) using the effect sizes from the original meta-analyses[88]. LDpred adjusts the effect sizes from the meta-analysis for the effects of linkage disequilibrium (LD) using an external reference panel to estimate the LD structure among SNPs. As the LD reference panel, we used in both HRS and NTR the cohort-specific genotype data imputed to 1000 Genomes Phase 1 reference panel and converted to hard calls. Since HapMap3 SNPs are known to be imputed reliably, the LD-PGS are based on HapMap3 SNPs only.[l] For the results with LD-PGS and Lin-PGS to be comparable, we also restricted the Lin-PGS to HapMap3 SNPs. We constructed all scores in PLINK using allelic dosages of genotypes imputed to 1000G Phase 1. For LD-PGS we set the fraction of causal SNPs to 1, and the Lin-PGS were obtained using all HapMap3 SNPs, without applying a *p*-value threshold. 1,059,092 SNPs were used to construct the scores in HRS, and 1,059,064 SNPs in NTR.

[l] Using only the genotyped SNPs to construct LD-PGS as was done in earlier work[87] was not possible for NTR since genotyping was done with multiple different chips, making it difficult to pool non-imputed data.

B. Results

We first regressed each residualized phenotype on 10 principal components. In HRS, the principal components were computed using HapMap3 SNPs with a minor allele frequency greater than 0.01, and restricting the sample to individuals of European ancestry. In NTR, 3 PCs were included that reflected the Dutch population structure and 7 PCs were included that reflected chip effects. For an extensive description of how these principal components were computed in NTR, see Abdellaoui et al. (2013)[58].

Next, we ran the same regressions adding the score as a covariate and computed the increase in R^2. We used Ordinary Least Squares (OLS) regression in both HRS and NTR despite NTR being a family cohort since R^2 is not biased by family structure. The incremental R^2's for each phenotype and polygenic score are shown in **Supplementary Table 33**. To obtain 95% confidence intervals (CI) around the incremental R^2's, bootstrapping was performed with 1000 repetitions.

The results of the polygenic score analyses are depicted in **Supplementary Figure 13**. The sample-size-weighted mean predictive power (incremental R^2 after controlling for age, age^2, sex, and PCs) for SWB is 0.9%, which is statistically distinguishable from zero but small relative to the predictive power attained by polygenic scores estimated in comparable sample sizes for other phenotypes (e.g., educational attainment). Similarly, the mean predictive power for its components, LS and PA, are both 0.7%. SWB also significantly predicts DS (0.5%) and the personality traits Neuroticisms (0.7%) and Extraversion (0.4%). In contrast, for the other phenotypes, the mean predictive power is in all cases smaller than 0.3% and in many cases not statistically distinguishable from zero. We compared the LDpred results with the linear approach. The results are similar but with the expected somewhat lower estimates for the linear method.

7. Proxy-phenotype and genetic overlap analyses

In this section, we describe and report the results from tests of joint enrichment that allow us to formally test if the SNPs showing the strongest evidence of association with one phenotype (for example, SWB), are more strongly associated with another phenotype (for example, DS) than expected by chance. The analyses are motivated by the evidence of the strong genetic correlations between SWB, DS, and neuroticism[89–91], including the results shown in **Supplementary Table 1 and Figure 2**. We also use bivariate LD score regression to estimate genetic correlations between each of our phenotypes – SWB, DS and neuroticism – are associated with a host of neuropsychiatric and physical health variables.

A. Methodology for proxy-phenotype and cross-phenotype enrichment analyses

We use a two-stage approach that has been successfully applied in other contexts[92]. In the first stage, we conduct a meta-analysis of a first-stage "proxy phenotype" (e.g., SWB). In the second stage, we test the "lead/lead-proxy SNPs"—the SNPs showing strongest evidence of association with the first-stage phenotype—for association with a second-stage phenotype (e.g., DS) in an independent (non-overlapping) sample. Note that in the analyses described in this section, relative to the GWAS on SWB, DS, and neuroticism reported in **Supplementary Note 2**, we omit cohorts from the first-stage or second-stage as needed to ensure that the samples in the two stages are non-overlapping.

In total, we perform three lookup exercises; see **Supplementary Table 31** for a summary overview of the analyses, including cohort restrictions used to eliminate overlap between the stage-one and stage-two samples. In our analysis of DS, we apply the effective sample-size weighting scheme described in **Supplementary Note 3.F** to the two case-control studies (GERA[39] and PGC[38]) and continue to weight UKB by its sample size. As in the main analyses, we perform sample-size-weighted meta-analyses of SWB and neuroticism.

For convenience, in what follows we adopt the convention of naming each lookup analysis in the format "First-stage phenotype → Second-stage phenotype". We conducted three lookup exercises. In our first lookup exercise, the first- and second-stage phenotypes are, respectively, SWB and DS, or simply SWB → DS. Our second lookup is SWB → Neuroticism, and our third lookup is SWB → Height, where we treat Height as a negative control.

We omit from the meta-analysis of the second-stage phenotype SNPs missing from a substantial fraction of individuals; see the notes in **Supplementary Table 31** for details. For example, in the analysis where the second-stage phenotype is DS, we only consider SNPs available in all three DS cohorts (GERA[39], PGC[38] and UKB[11]). And in the analysis where the second-stage is neuroticism, we only consider SNPs available in our two neuroticism cohorts, PGC[37] and UKB[11], with a minimum total sample size of N = 90,000. Below, we describe the methodology we used to construct the lead SNPs, and the tests of enrichment we performed.

Generating lead SNPs

Throughout, we apply a uniform methodology to define the lead SNPs that are subsequently tested for association, both jointly and individually, with the second-stage phenotype. For brevity, we illustrate the methodology used to construct our list of lead/lead-proxy SNPs using the example of SWB. However, the procedure used in the other two lookups is nearly identical, as explained in the relevant subsections below.

We began by identifying a set of approximately independent "SWB-associated SNPs" from the first-stage meta-analysis (or more generally, "first-stage-phenotype-associated SNPs"). We applied the clumping methodology described in Online Methods, but choosing the *p*-value threshold for the index SNPs to be 10^{-4}. The more liberal *p*-value threshold was chosen prior to the study based on power calculations. As in our main analyses, we used the 1000G phase 1 reference sample[27] composed of Utah Residents (CEPH) with Northern and Western European Ancestry (CEU), Toscani in Italia (TSI), and British in England and Scotland (GBR) for clumping and for estimating linkage disequilibrium.

Applying the clumping procedure to the SWB meta-analysis results from the SWB → DS lookup generated 223 approximately independent SWB-associated lead SNPs. Of these, 85 were available in all three DS cohorts used in the second-stage analyses, whereas 148 were not. For each of these 148 SNPs, we examined if there are any SNPs satisfying the following conditions: (i) the SNP is in high LD ($R^2 > 0.8$) with the SWB-associated SNP, and (ii) the SNP is available both in the SWB meta-analysis and in all three cohorts contributing to the meta-analysis of DS. A proxy-lead SNP satisfying these criteria was available for 78 out of 148 SNPs (mean $R^2 = 0.96$, range 0.81 to 1.00). Whenever more than one proxy is available for a SNP, we chose as our proxy the SNP whose R^2 with the SWB-associated SNP was the greatest. Our final list of lead SNPs in the first lookup exercise therefore contains 85+78 = 163 SNPs.

Testing lead/proxy-lead SNPs for Enrichment

Because SWB, DS, and neuroticism phenotypes are all highly polygenic, it is of limited interest to test the null hypothesis that the *p*-value distribution of the lead/lead-proxy SNPs is uniform. We instead perform a non-parametric test of joint enrichment that probes whether the lead SNPs are more strongly associated with the second-stage phenotype than randomly chosen sets of SNPs with minor allele frequencies within one percentage point of the lead/proxy-lead SNP. To perform our test, we generated 1,000 matched SNPs for each of the *Y* lead/lead-proxy SNPs (e.g., *Y* = 163 in the SWB → DS analysis).

We then ranked the *Y*×1000 + *Y* SNPs by *p*-value and conducted a Mann-Whitney test[93] of the null hypothesis that the *p*-value distribution of the *Y* lead/lead-proxy SNPs are drawn from the same distribution as the *Y*×1000 matched SNPs. To test the individual lead SNPs for experiment-wide significance, we examine whether any of the lead SNPs (or their high-LD proxies) are significantly associated with the second-stage phenotype at the Bonferroni-corrected significance level of 0.05/*Y*. Throughout, we adopt the convention of classifying an effect size as "in the predicted direction" if either (i) the signs are concordant and the two phenotypes are estimated to have a positive genetic correlation, or (ii) the signs are discordant and the phenotypes are estimated to have a negative genetic correlation.

B. Results from proxy-phenotype and cross-phenotype enrichment analyses

Are SWB-Associated SNPs Enriched for Depression?

Supplementary Figure 2a is a two-way scatterplot of the *z*-statistics of the lead/lead-proxy SNPs in SWB (horizontal axis) against DS (vertical axis). To aid interpretation, we choose the reference allele to be the SWB-increasing variant, so all *z*-statistics are by construction positive for the first-stage phenotype. On the basis of the negative genetic correlation reported in **Supplementary Table 1** ($\hat{\rho} = -0.81$), we expect plotted points to lie disproportionately below the dashed horizontal line at zero (i.e. negative *z*-statistics). That is indeed what we find: 116 out of 163 (71%) signs are in the expected direction. Moreover, for 19 out

of the 20 SNPs that are nominally significantly associated ($p < 0.05$) in the analysis of DS, the association is in the predicted direction.

Pictured with black circles in the scatterplot are the three lead/proxy-lead SNPs that reach p-value < 10^{-7} in the SWB meta-analysis. Two of these are nominally associated with DS: rs12517563 ($p = 0.007$) and rs2075677 ($p = 0.0149$). Pictured with red circles are the two SNPs that are significantly associated with depressive symptoms at the Bonferroni-corrected p-value threshold of 0.05/163 = 0.00037. These are rs6904596 ($p = 9.78\times10^{-5}$) and rs4481363 ($p = 3.06\times10^{-4}$). The direction of the association with depressive symptoms is in the predicted direction for all four SNPs (rs12517563, rs2075677, rs6904596, rs4481363): the SWB-increasing allele is estimated to reduce depression risk. **Supplementary Table 16** lists the association results for the lead/proxy-lead SWB-associated SNPs in the first-stage SWB meta-analysis and the second-stage depressive symptoms meta-analysis conducted in an independent sample. The SNPs are ordered by p-value attained in the SWB analysis (from smallest to largest). Among SWB-associated SNPs with p-value < 10^{-5}, 80% have signs in the predicted direction. Our test of joint enrichment rejects the null of no enrichment relative to the expected level for a randomly sampled set of SNPs matched on allele frequency ($p = 0.033$).

Are SWB-Associated SNPs Enriched for Neuroticism?

Applying the clumping algorithm in Online Methods, we identified 170 lead/lead-proxy SNPs from the first-stage analysis of SWB. The results from this lookup analysis are summarized in **Supplementary Figure 2b**, where the reference allele is again chosen to be the SWB-increasing allele. Given the negative genetic correlation reported in **Supplementary Table 1** ($\hat{\rho} = -0.75$), we expect z-statistics disproportionately below the dashed horizontal line. Indeed, 129 out of 170 signs (76%) are in the predicted direction in the neuroticism results. Moreover, all 28 SNPs that are nominally significant in the neuroticism analysis have the predicted sign. None of the three SNPs reaching p-value < 10^{-7} in the first-stage analysis are associated with neuroticism. However, four SNPs are significant at the Bonferroni-corrected significance threshold 0.05/173 = 0.00029. These are *rs10838738* ($p = 2.6\times10^{-5}$), *rs6904596* ($p = 4.2\times10^{-5}$), *rs4481363* ($p = 5.7\times10^{-5}$) and *rs10774909* ($p = 7.3\times10^{-5}$). In all four cases, the effects are in the expected direction. For complete results, see **Supplementary Table 17**. Finally, our test of joint enrichment rejects the null of no enrichment relative to the expected level for a randomly sampled set of SNPs matched on allele frequency ($p = 10^{-4}$).

Negative-Control Analyses: Are SWB-Associated SNPs Enriched for Height?

For our negative-control analyses, our first-stage analyses of SWB were performed omitting cohorts that contributed to GIANT consortium's year-2010 study of height[94], leaving us with a first-stage discovery sample of N = 229,853. Applying our methodology gives 181 lead/lead-proxy SNPs. Our second-stage lookup is conducted using publicly available summary statistics from the height GWAS (N = 133,859). We find no evidence that the proportion of SNPs for which the allele estimated to increase SWB is also the allele estimated to increase height is statistically distinguishable from 50% ($p = 0.373$), and the Mann-Whitney test of joint enrichment fails to reject the null hypothesis ($p = 0.454$).

C. Genetic correlations: methodology

From twin and family studies, there is evidence of genetic overlap between subjective well-being phenotypes, depression, and neuroticism[89–91]. We obtained evidence from the GWAS data by using bivariate LD score regression[28,36] to estimate pairwise genetic correlations between SWB, DS, and

neuroticism. Treating height as a negative control, we also estimated the genetic correlation between height[95] and each of SWB, DS, and neuroticism. We also estimated genetic correlations between our three phenotypes and ten neuropsychiatric and physical health variables.

LD Score regression produces unbiased estimates even in the presence of sample overlap and only requires summary statistics and a reference panel from which to estimate each SNP's "LD score" (the amount of genetic variation tagged by a SNP). We used the file of LD Scores computed by Finucane et al.[49] using genotypic data from a European-ancestry population (eur_w_ld_chr, see https://github.com/bulik/ldsc/wiki/Genetic-Correlation, accessed September 15, 2015).

LD Score regression exploits the fact that under some reasonable assumptions, the following moment condition holds for a polygenic trait:

$$E\left[z_{1j}z_{2j}\right] = Intercept + \frac{\sqrt{N_1N_2}}{M} Cov_g \ell_j,$$

where z_{kj} is the *z*-statistic of SNP *j* from the GWAS of trait *k* (*k* = 1, 2), *Intercept* is the regression intercept, N_k is the sample size of the GWAS of trait *k*, *M* is the number of SNPs included in the GWAS, Cov_g is the genetic covariance between traits *1* and *2*, and ℓ_j is the LD Score of SNP *j*. The slope parameter from the regression of $\hat{z}_{1j}\hat{z}_{2j}$ on $\sqrt{N_1N_2}\ell_j$ can therefore be used to estimate the genetic covariance between the two traits. From separate, univariate LD Score regressions of traits 1 and 2, it is also possible to recover estimates of the respective heritabilities of the two traits, h^2_{g1} and h^2_{g2}.

Putting these pieces together, an estimate of the genetic correlation is given by:

$$\hat{r}_g = \frac{\widehat{Cov}_g}{\sqrt{\hat{h}^2_{g1}\hat{h}^2_{g2}}}.$$

Following Finucane et al.[49], LD Scores are computed with genotypes from the European-ancestry samples in the 1000 Genomes Project using only HapMap3 SNPs. We additionally follow the common convention of restricting our analyses to SNPs with MAF > 0.01, thus ensuring that all analyses are performed using a set of SNPs that are imputed with reasonable accuracy across all cohorts that contributed to the meta-analyses.

The standard errors are estimated (by the LDSC software) using a block jackknife over SNPs. As such, they should be interpreted as the variability of the estimate holding the sample constant but drawing a new set of SNPs. This is in contrast to the conventional interpretation of standard errors, which measure the variability of the estimate holding the covariates constant but drawing new sets of individuals. Ideally we would have standard errors that represent the latter, but it is unclear how one might obtain such estimates with the available data. For this reason, we only report the block jackknife standard errors as in Bulik-Sullivan et al.[28].

D. Genetic correlations: results

The estimated genetic correlations are shown at the top of **Supplementary Table 1** and in **Figure 2a**. For SWB, we estimate a genetic correlation of $\hat{r}_g$= -0.814 with DS and $\hat{r}_g$= -0.749 with neuroticism. The genetic

correlation between DS and neuroticism is $\hat{r}_g$ = 0.750. As expected, the genetic correlations with height are modest: $\hat{r}_g$ = 0.065, -0.062, and -0.061, for SWB, DS, and neuroticism.

To provide some evidence on whether there may be shared biological pathways with a number of neuropsychiatric and physical health phenotypes, we also examined genetic overlap between these and each of SWB, DS, and neuroticism. The neuropsychiatric phenotypes are Alzheimer's disease[96], anxiety disorders[97,m], autism spectrum disorder[98], bipolar disorder[99], and schizophrenia[25]. The physical health phenotypes are body mass index (BMI)[100], coronary artery disease[101,n], ever-smoker status[102], fasting glucose[103,o], and triglycerides[104].

We had several criteria for choosing which phenotypes to include in this analysis. First, we chose from among the set of phenotypes that have well-powered GWAS summary statistics in the public domain (which we could use for the analysis). Second, we focused on phenotypes believed to be phenotypically or genetically correlated with SWB. Finally, to keep the number of genetic correlations manageable, among sets of phenotypes that are highly phenotypically or genetically correlated with each other, we included only one from each set[p].

SWB or depression have all been found to be phenotypically related to a number of mental health measures (such as survey responses to "ever diagnosed with depression/anxiety" or "currently in treatment for a mental health condition"), smoking rates, and incidence of coronary heart disease[105-106,107]. Although there are no large GWA studies on coronary *heart* disease, that condition is the result of coronary *artery* disease, which we include. Both are related to lipids in the blood, including triglycerides, which is a phenotype we also include. Moreover, we investigate schizophrenia, bipolar disorder, and triglycerides because each has been found to be genetically correlated with major depressive disorder[36,38]. Although there is a known phenotypic correlation between SWB and general immune system function[108, 109, 110, 111, 112], we excluded uncommon autoimmune diseases (such as Crohn's disease and ulcerative colitis) since we believe such diseases are unlikely to play a major role in that relationship.

The estimated genetic correlations are shown in **Supplementary Table 1** and **Figure 2b and 2c.** For ease of interpretation, given that DS and neuroticism are each negatively genetically correlated with SWB, we plot genetic correlations with SWB alongside *negative* genetic correlations for DS and neuroticism. Every statistically significant genetic correlation has a sign in the direction that would be expected based on the

---

[m] Otowa et al. used two phenotype measures in their GWAS of anxiety disorders: one based on a categorical case-control designation for having any anxiety disorder diagnosis, and another based on continuous factor scores derived by combining information across clinical phenotypes. Our analyses are based on the case-control measure.

[n] We downloaded the data on coronary artery disease / myocardial infarction from www.CARDIOGRAMPLUSC4D.ORG. Those data were contributed by the Coronary ARtery DIsease Genome wide Replication and Meta-analysis plus the Coronary Artery Disease Genetics (CARDIoGRAMplusC4D) investigators.

[o] We downloaded the data on glycaemic traits from www.magicinvestigators.org. Those data were contributed by the Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) investigators.

[p] For example, Bulik-Sullivan et al.[36] found that waist-to-hip ratio (which we exclude) is highly genetically correlated with body mass index (BMI); type-2 diabetes and glycated hemoglobin (HbA1C) (which we exclude) are highly genetically correlated with fasting glucose; and high-density lipoprotein (HDL) and low-density lipoprotein (LDL) cholesterol (which we exclude) are highly genetically correlated with triglycerides.

corresponding phenotypic correlation. In the remainder of this discussion, when reporting the genetic correlations we estimate, we put them in brackets in the order [$\hat{r}_g$ with SWB, $\hat{r}_g$ with DS, $\hat{r}_g$ with neuroticism].

Anxiety disorders show the strongest genetic overlap with all three of our phenotypes [0.729, 0.877, 0.864]. We find moderate genetic correlations between each of our three phenotypes and schizophrenia [-0.266, 0.333, 0.215] and bipolar disorder [-0.127, 0.264, 0.109]. Only SWB has a statistically significant genetic correlation with autism spectrum disorder [-0.286, 0.106, 0.155], but we cannot reject equality of the three genetic correlations at the 5% level. Alzheimer's disease is not significantly genetically correlated with any of our three phenotypes (point estimates: [-0.108, 0.025, 0.088]).

All three phenotypes have a mild, albeit statistically significant, genetic correlation with ever-smoker status [-0.164, 0.238, 0.130]. Coronary artery disease has very mild genetic correlations, with estimates significant at the 5% level for DS and neuroticism [-0.096, 0.134 and 0.125]. While we find no statistically significant evidence of genetic overlap with fasting glucose [0.037, 0.007, -0.067], there is statistically significant evidence of minor genetic correlation between triglycerides and all three of our phenotypes [-0.073, 0.115, 0.066]. BMI is significantly genetically correlated only with DS [-0.034, 0.122, -0.011].

E. Post hoc lookup of schizophrenia-associated loci in neuroticism

The three neuropsychiatric conditions with which we observe the strongest genetic correlations are schizophrenia, anxiety disorder and bipolar disorder (**Supplementary Table 1**). Of these three, schizophrenia—for which we estimate a genetic correlation with neuroticism of $r_g = 0.215$ (p = 3.86×10^{-7})—is the only condition for which a large number of genome-wide significant associations have been reported (in a recent paper by Ripke et al.[25]). Following the suggestion of a referee, we conducted a post hoc lookup of genome-wide significant associations that have been reported for schizophrenia.

We looked up these published results in our summary statistics from a GWAS of neuroticism performed in our sample of UKB respondents. We focus on the UKB sample because it does not overlap with Ripke et al.'s schizophrenia sample. We exclude summary statistics from the GPC study because we do not have we do not have access to their cohort-level results, and there is cohort overlap with the Ripke et al. cohorts. Of the 128 genome-wide significant genetic variants identified in Ripke et al.'s GWAS of schizophrenia, 106 were available in our UKB results file.

We began by examining how often the allele associated with greater schizophrenia risk was associated with greater neuroticism. We found concordant signs for 73 out of 106 matched SNPs, a sign concordance of 69%. Using a two-sided binomial test, we strongly reject the null hypothesis that the true proportion of sign alignment is 50% ($p = 1.28 \times 10^{-4}$).

Using the procedure outlined in **Supplementary Note 7.A**, we also conducted a test of enrichment of the schizophrenia-associated SNPs in the neuroticism results. We strongly reject the null hypothesis of no enrichment relative to a set of SNPs matched on allele frequency ($p = 6.50 \times 10^{-71}$). Also, 23 of the 106 matched schizophrenia SNPs are nominally significantly associated ($p < 0.05$) with neuroticism in our sample, and 19 of these 23 SNPs have concordant signs for schizophrenia and neuroticism. Four SNPs are significantly associated with neuroticism after correction for multiple testing ($p < 0.05 / 106 = 4.72 \times 10^{-4}$): rs2514218 (β = 0.0737, $p = 2.98 \times 10^{-7}$), rs11682175 (β = -0.062, $p = 8.11 \times 10^{-6}$), rs7432375 (β = 0.057, $p = 4.55 \times 10^{-5}$), and rs832187 (β = 0.051, $p = 3.23 \times 10^{-4}$). We thus conclude that SNPs that have

been shown to be associated with schizophrenia also tend to be associated with normal variation in neuroticism.

For anxiety (for which our estimated genetic correlation with neuroticism is $r_g = 0.864$, *p*-value = 1.06×10^{-7}) and bipolar disorder ($r_g = 0.109$, *p* = 0.011), no large-scale GWA studies have been performed yet, and few genome-wide significant associations have been reported. Otowa et al.[97] reported two genome-wide associations with anxiety disorders, neither of which replicate in our UKB sample (the *p*-values are 0.36 and 0.43). The most recent GWAS of bipolar disorder[113] identified genome-wide significant associations in five distinct loci (see Table 2 of that paper); for each of the five loci, we looked up the SNP with the lowest reported *p*-value. Of these five SNPs, two have the predicted sign in our neuroticism summary statistics, and only one is statistically significant but with the wrong sign (the *p*-values are 0.002, 0.053, 0.16, 0.70, and 0.93).

8. Bayesian credibility analysis

When a SNP in a GWAS is statistically significant, that provides some evidence against the hypothesis that the SNP is null. Using a Bayesian framework with a prior distribution over effect sizes that is a mixture of a normal distribution and a point mass at zero, we can discuss a broader set of questions regarding the credibility of these estimates given the results of the GWAS. The Bayesian analysis corrects for the Winners' Curse and can allow us to leverage information contained in the GWAS estimates of related traits. The specific questions we will address in this section are (1) what is the posterior probability that a SNP is non-null given its GWAS estimate, (2) what is the expected record of our quasi-replication analyses (of a different phenotype) given the GWAS estimates, (3) by how much does the credibility of a lead SNP improve or worsen when we incorporate its quasi-replication results, and (4) how well-powered are our GWA studies relative to previous GWA studies of these traits.

A. Single phenotype framework

Set-up

We begin with some prior belief of the distribution of effect sizes. Using a common assumption in modeling SNP effects[87,114], we use a mixture of a Gaussian prior with a point mass at zero. More precisely, the prior for the GWAS coefficient corresponding to any given SNP is

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

where π is our prior belief of the fraction of non-null SNPs, and τ^2 is our prior belief of the variance of the effect size of non-null SNPs. We denote $\hat{\beta}$ as the GWAS estimate of β, and note that the conditional density of the GWAS estimate given the true effect size (regardless of whether the SNP is null or non-null) is

$$f(\hat{\beta} \mid \beta) = \frac{1}{\sigma}\phi\left(\frac{\hat{\beta} - \beta}{\sigma}\right),$$

where $\phi(\cdot)$ corresponds to the standard normal density function, and σ^2 is the sampling variance of $(\hat{\beta} \mid \beta)$. We likewise note that the unconditional density of non-null SNPs is

$$f(\hat{\beta} \mid \beta \neq 0) = \frac{1}{\sqrt{\sigma^2 + \tau^2}}\phi\left(\frac{\hat{\beta} - \beta}{\sqrt{\sigma^2 + \tau^2}}\right),$$

since $(\hat{\beta} \mid \beta \neq 0)$ is the sum of two normally distributed random variables, β and the sampling error.

Posterior non-null probability

We are first interested in the probability that a SNP is non-null given its estimated effect size and given that it is significant (say, at the genome-wide threshold, 5×10^{-8}). We denote this probability by $p_{\hat{\beta}}$. Define S to be the event that $\hat{\beta}$ is significant. Since we will only consider SNPs that are estimated to be significant, conditioning on S in addition to $\hat{\beta}$ is redundant. Therefore,

$$\begin{aligned} p_{\hat{\beta}} &\equiv P(\beta \neq 0 \mid \hat{\beta}, S) \\ &= P(\beta \neq 0 \mid \hat{\beta}). \end{aligned}$$

Then, by Bayes' Rule,

$$p_{\hat{\beta}} = \frac{f(\hat{\beta} \mid \beta \neq 0)P(\beta \neq 0)}{f(\hat{\beta} \mid \beta = 0)P(\beta = 0) + f(\hat{\beta} \mid \beta \neq 0)P(\beta \neq 0)}$$
$$= \frac{\frac{1}{\sqrt{\sigma^2+\tau^2}}\phi\left(\frac{\hat{\beta}}{\sqrt{\sigma^2+\tau^2}}\right)\pi}{\frac{1}{\sigma}\phi\left(\frac{\hat{\beta}}{\sigma}\right)(1-\pi) + \frac{1}{\sqrt{\sigma^2+\tau^2}}\phi\left(\frac{\hat{\beta}}{\sqrt{\sigma^2+\tau^2}}\right)\pi}.$$

Rearranging slightly,

$$p_{\hat{\beta}} = \frac{\frac{\pi}{\sqrt{\sigma^2+\tau^2}}\phi\left(\frac{\hat{\beta}}{\sqrt{\sigma^2+\tau^2}}\right)}{\frac{1-\pi}{\sigma}\phi\left(\frac{\hat{\beta}}{\sigma}\right) + \frac{\pi}{\sqrt{\sigma^2+\tau^2}}\phi\left(\frac{\hat{\beta}}{\sqrt{\sigma^2+\tau^2}}\right)}. \quad (1)$$

We refer to this quantity as the "first-stage credibility." It is the credibility that the SNP is non-null given the GWAS estimate of the association between the SNP and the (first-stage) phenotype.

Posterior non-null distribution

To calculate the other measures of credibility, we need to know the posterior distribution of non-null SNPs. Similar to before, we calculate

$$\begin{aligned} f(\beta \mid \beta \neq 0, \hat{\beta}, S) &= f(\beta \mid \beta \neq 0, \hat{\beta}) \\ &\propto f(\hat{\beta} \mid \beta, \beta \neq 0) f(\beta \mid \beta \neq 0) \\ &= \frac{1}{\sigma}\phi\left(\frac{\hat{\beta}-\beta}{\sigma}\right)\frac{1}{\tau}\phi\left(\frac{\beta}{\tau}\right) \\ &\propto \exp\left\{-\frac{1}{2}\left[\frac{(\hat{\beta}-\beta)^2}{\sigma^2} + \frac{\beta^2}{\tau^2}\right]\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left[-2\left(\frac{\tau^2}{\sigma^2+\tau^2}\right)\hat{\beta}\beta + \beta^2\right]\left(\frac{\sigma^2+\tau^2}{\sigma^2\tau^2}\right)\right\} \\ &\propto \exp\left\{-\frac{1}{2}\frac{\left[\beta - \left(\frac{\tau^2}{\sigma^2+\tau^2}\right)\hat{\beta}\right]^2}{\left(\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}\right)}\right\}, \end{aligned}$$

where we highlight that each of the proportionalities here are proportional with respect to integrating over β. (Specifically, we define this notion of proportionality such that, if $h_1(\beta) \propto h_2(\beta)$ with respect to β for some functions h_1 and h_2, then $h_1(\beta)/h_2(\beta)$ is a constant for all β.) For ease of notation in future calculations, we define

$$b \equiv \left(\frac{\tau^2}{\sigma^2+\tau^2}\right)\hat{\beta}$$
$$s^2 \equiv \frac{\sigma^2\tau^2}{\sigma^2+\tau^2}.$$

Using this notation, the above calculation implies that

$$\left(\beta \mid \beta \neq 0, \hat{\beta}, S\right) \sim N(b, s^2). \tag{2}$$

Posterior credible interval

To calculate the posterior Bayesian credible intervals, we first note that for some constant threshold t_{lb} (that will represent the lower bound of the interval),

$$P\left(\beta \leq t_{lb} \mid \hat{\beta}\right) = \begin{cases} \left(1 - p_{\hat{\beta}}\right) + p_{\hat{\beta}} \int_{-\infty}^{t_{lb}} f\left(\beta \mid \beta \neq 0, \hat{\beta}, S\right) d\beta & \text{if } t_{lb} \geq 0 \\ p_{\hat{\beta}} \int_{-\infty}^{t_{lb}} f\left(\beta \mid \beta \neq 0, \hat{\beta}, S\right) d\beta & \text{otherwise.} \end{cases}$$

Similarly,

$$P\left(\beta \geq t_{ub} \mid \hat{\beta}\right) = \begin{cases} \left(1 - p_{\hat{\beta}}\right) + p_{\hat{\beta}} \int_{t_{ub}}^{\infty} f\left(\beta \mid \beta \neq 0, \hat{\beta}, S\right) d\beta & \text{if } t_{ub} \leq 0 \\ p_{\hat{\beta}} \int_{t_{ub}}^{\infty} f\left(\beta \mid \beta \neq 0, \hat{\beta}, S\right) d\beta & \text{otherwise.} \end{cases}$$

We then solve for t_{lb} and t_{ub} such that

$$P\left(\beta \leq t_{lb} \mid \hat{\beta}\right) = \frac{\alpha}{2}$$

and

$$P\left(\beta \geq t_{ub} \mid \hat{\beta}\right) = \frac{\alpha}{2},$$

giving us a $(1 - \alpha)\%$ credible interval of $[t_{lb}, t_{ub}]$. Solving gives:

$$t_{lb} = \begin{cases} b + s\Phi^{-1}\left[\frac{1}{p_{\hat{\beta}}}\left(\frac{\alpha}{2}\right)\right] & \text{if } p_{\hat{\beta}}\Phi\left(-\frac{b}{s}\right) > \frac{\alpha}{2} \\ 0 & \text{if } p_{\hat{\beta}}\Phi\left(-\frac{b}{s}\right) + \left(1 - p_{\hat{\beta}}\right) > \frac{\alpha}{2} \geq p_{\hat{\beta}}\Phi\left(-\frac{b}{s}\right) \\ b + s\Phi^{-1}\left[\frac{1}{p_{\hat{\beta}}}\left(\frac{\alpha}{2} - \left(1 - p_{\hat{\beta}}\right)\right)\right] & \text{otherwise,} \end{cases} \tag{3}$$

and

$$t_{ub} = \begin{cases} b + s\Phi^{-1}\left[\frac{1}{p_{\hat{\beta}}}\left(1-\frac{\alpha}{2}\right)\right] & \text{if } p_{\hat{\beta}}\Phi\left(-\frac{b}{s}\right) > 1-\frac{\alpha}{2} \\ 0 & \text{if } p_{\hat{\beta}}\Phi\left(-\frac{b}{s}\right) + \left(1-p_{\hat{\beta}}\right) > 1-\frac{\alpha}{2} \geq p_{\hat{\beta}}\Phi\left(-\frac{b}{s}\right) \\ b + s\Phi^{-1}\left(\frac{1}{p_{\hat{\beta}}}\left[\left(1-\frac{\alpha}{2}\right) - \left(1-p_{\hat{\beta}}\right)\right]\right) & \text{otherwise,} \end{cases} \quad (4)$$

where $\Phi(\cdot)$ and $\Phi^{-1}(\cdot)$ are the standard normal cumulative distribution function and its inverse, respectively.

B. Posterior distribution of effect on second-stage phenotype

In a quasi-replication analysis, such the one done in this paper, we may be interested in the expected performance of the quasi-replication given estimates for the lead SNPs identified in the first stage. To answer this question, we must generalize the model described above to account for a second phenotype and the genetic correlation between the first-stage and second-stage phenotypes.

To carry out the two tests of quasi-replication performance we report in this paper, we will need two measures of expected performance: first, the probability that the second-stage estimate has the predicted sign based on the sign of the first-stage estimate; and second, the probability that the *p*-value for the association between the SNP and the second-stage phenotype will be less than 0.05. To make the calculation of these probabilities analogous to a frequentist power analysis, we will assume that the SNP is non-null for both the first- and second-stage phenotype. The expected quasi-replication performance can then be compared to the observed sign concordance and significance.

Set-up

Consider two phenotypes, and define β_1 and β_2 to be the respective true GWAS parameters of a particular SNP for the phenotypes (i.e., the association that would be estimated in an infinite sample). As discussed in the previous paragraph, throughout this derivation, we will assume that the SNP is non-null for both phenotypes, and we will denote this event by the variable C (since these SNPs may be thought of as causal or tagging a causal SNP). Similar to the parameterization above, conditional on the SNP being non-null for both phenotypes, we denote the variance of the parameters β_1 and β_2 across SNPs by τ_1^2 and τ_2^2, respectively, and we denote the correlation of the parameters by r_g. Thus, for a non-null SNP, the covariance of β_1 and β_2 is $\tau_1\tau_2 r_g$.

Let $\hat{\beta}_1$ and $\hat{\beta}_2$ denote GWAS estimates of their corresponding parameters. Assuming that these coefficients are estimated in non-overlapping samples, the variance-covariance matrix of the estimation error will be diagonal. We denote the diagonal entries of this matrix, which signify the variance of the estimation error in the two samples, by σ_1^2 and σ_2^2. This gives us the joint prior distribution

$$\left(\begin{bmatrix}\hat{\beta}_1 \\ \hat{\beta}_2\end{bmatrix} \mid \mathrm{C}\right) \sim N\left(\begin{bmatrix}0 \\ 0\end{bmatrix}, \begin{bmatrix}\tau_1^2 & \tau_1\tau_2 r_g \\ \tau_1\tau_2 r_g & \tau_2^2\end{bmatrix} + \begin{bmatrix}\sigma_1^2 & 0 \\ 0 & \sigma_2^2\end{bmatrix}\right). \quad (5)$$

Again, note that in (5) the first matrix in the variance term represents variability in the parameter across SNPs and the second matrix represents variability due to estimation error.

Posterior non-null distribution

For this analysis, we are interested in the posterior distribution of one estimate given the estimate of the other and given that the SNP is non-null for both phenotypes. Without loss of generality, we assume that $\hat{\beta}_1$ is the observed estimate from the first stage and that we need to calculate the distribution of $\hat{\beta}_2$ given $\hat{\beta}_1$. By the properties of bivariate normal distributions and by the parameterization in (5), we know that the conditional distribution is

$$\left(\hat{\beta}_2 \mid \hat{\beta}_1, C\right) \sim N\left[\frac{\tau_1\tau_2 r_g}{\tau_1^2+\sigma_1^2}\hat{\beta}_1, \frac{(\tau_1^2+\sigma_1^2)(\tau_2^2+\sigma_2^2)-\tau_1^2\tau_2^2 r_g^2}{\tau_1^2+\sigma_1^2}\right]. \tag{6}$$

We can use (6) to calculate various measures of expected quasi-replication. For example, considering the relevant case where $r_g < 0$ and the reference allele is chosen so that $\hat{\beta}_1 < 0$, the probability that $\hat{\beta}_2$ will have the predicted (i.e., opposite) sign given $\hat{\beta}_1$ for a particular SNP, assuming that the SNP is non-null for both phenotypes, is

$$P\left[\text{sign}\left(\hat{\beta}_2\right) = \text{sign}\left(\hat{\beta}_1\right) \mid \hat{\beta}_1, C\right] = \Phi\left[\frac{\tau_1\tau_2 r_g}{(\tau_2^2+\sigma_2^2)(\tau_1^2+\sigma_1^2)-\tau_1^2\tau_2^2 r_g^2}\hat{\beta}_1\right], \tag{7}$$

where $\Phi(\cdot)$ is the standard normal cdf. Similarly, the probability that $\hat{\beta}_2$ will have a *p*-value less than α is

$$\begin{aligned} P\left(p < \alpha \mid \hat{\beta}_1, C\right) &= \Phi\left(\frac{-t_\alpha(\tau_1^2+\sigma_1^2)-\tau_1\tau_2 r_g\hat{\beta}_1}{(\tau_1^2+\sigma_1^2)(\tau_2^2+\sigma_2^2)-\tau_1^2\tau_2^2 r_g^2}\right) \\ &+ 1-\Phi\left(\frac{t_\alpha(\tau_1^2+\sigma_1^2)-\tau_1\tau_2 r_g\hat{\beta}_1}{(\tau_1^2+\sigma_1^2)(\tau_2^2+\sigma_2^2)-\tau_1^2\tau_2^2 r_g^2}\right), \end{aligned} \tag{8}$$

where t_α is the threshold corresponding to significance at the α-level. For the expected number of SNPs with matching estimated signs or with a second-stage estimate exceeding some significant threshold, we can simply add these probabilities up across the SNPs being tested. Since the SNPs are chosen to be independent, to get the variance of the measures, we sum over the variance of these binary random variables. By the properties of binary variables, this is simply $q(1-q)$, where q is the probability of the event, given above.

C. Bivariate framework

As a final set of analyses, we may be interested in the credibility of our associations, or posterior distribution of effect sizes, conditional on both the estimates for the first- and second-stage phenotypes. Having these allow us to answer questions such as how the credibility of our findings increased (or decreased) based on the analysis of an additional related phenotype. This framework will also allow us to leverage the additional information in the analysis of the second-stage phenotype to generate more precise posterior estimates of the association between the SNPs and the first-stage phenotype.

Set-up

Here, we use the same framework as in the previous subsection, where β_1 and β_2 are the population parameters for the effect sizes of some SNP with a pair of phenotypes, τ_1^2 and τ_2^2 are the respective variances of β_1 and β_2 across non-null SNPs, and r_g is the genetic correlation between the traits (i.e., the correlation between β_1 and β_2 across SNPs). Similarly, $\hat{\beta}_1$ and $\hat{\beta}_2$ are GWAS estimates of the population parameters, and σ_1^2 and σ_2^2 are their respective sampling variances. Since we assume that the traits are estimated in non-overlapping samples, the estimation error is uncorrelated. This means that the distributions of $\hat{\beta}_1$ and $\hat{\beta}_2$ when the SNP is non-null for both phenotypes is the same as (5).

We make the simplifying assumption that a SNP is null for one phenotype if and only if it is null for both phenotypes. Relaxing this assumption makes the derivation below significantly less tractable (and requires several additional parameters for the probabilities of being null for one phenotype conditional on being null for the other). For closely related phenotypes such as those studied in this paper, we believe that this assumption will hold sufficiently well to avoid introducing significant bias to our results. As evidence that this appears to be the case, we will show in the results section that the predicted replication record closely matches the observed replication record under this assumption. We caution, however, that this assumption should be relaxed when the pair of phenotypes being considered are not closely related.

We again denote the event that both SNPs are non-null by C, which occurs with probability π. Given the assumption described in the previous paragraph, we note that the complement of C—that is, the event that both SNPs are null, which we denote $\bar{C}$—occurs with probability $1-\pi$.

Posterior non-null probability

This derivation follows closely that of the single trait case. As before, we will need the distribution of the estimates when the SNP is null and when it is non-null. The case where the SNP is non-null is defined by (5). When the SNP is null, then the only source of variation in the estimates is the estimation error, giving us

$$\left(\begin{bmatrix}\hat{\beta}_1\\ \hat{\beta}_2\end{bmatrix} \mid \bar{C}\right) \sim N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\sigma_1^2 & 0\\ 0 & \sigma_2^2\end{bmatrix}\right). \tag{9}$$

We can the substitute the pdfs for these distributions into this expression, which is derived by Bayes' Rule:

$$P\left(C \mid \hat{\beta}_1, \hat{\beta}_2\right) = \frac{f\left(\hat{\beta}_1, \hat{\beta}_2 \mid C\right)\pi}{f\left(\hat{\beta}_1, \hat{\beta}_2 \mid C\right)\pi + f\left(\hat{\beta}_1, \hat{\beta}_2 \mid \bar{C}\right)(1-\pi)}. \tag{10}$$

This gives us our measure of "combined credibility." It is the credibility that the SNP is non-null given both the observed first- and second-stage estimates.

Posterior non-null distribution

We will also be interested in the posterior distributions of β_1 and β_2 given their estimates when the SNP is non-null. We refer to these distributions as the "posterior non-null distributions." Throughout this derivation, we focus on the posterior non-null distribution for β_1, although the derivation is symmetric, and thus the results can be used to find the non-null distribution of β_2 as well.

We begin by using Bayes' Rule, which implies

$$f(\beta_1 \mid \hat{\beta}_1, \hat{\beta}_2, C) = \frac{f(\hat{\beta}_2 \mid \beta_1, \hat{\beta}_1, C) f(\beta_1 \mid \hat{\beta}_1, C)}{f(\hat{\beta}_2 \mid \hat{\beta}_1, C)} \propto f(\hat{\beta}_2 \mid \beta_1, \hat{\beta}_1, C) f(\beta_1 \mid \hat{\beta}_1, C).$$

We note that, since we are conditioning on the SNP being non-null, we have derived $f(\beta_1 \mid \hat{\beta}_1, C)$ already as in (2) above,

$$(\beta_1 \mid \hat{\beta}_1, C) \sim N\left(\frac{\tau_1^2}{\tau_1^2 + \sigma_1^2}\hat{\beta}_1, \frac{\tau_1^2 \sigma_1^2}{\tau_1^2 + \sigma_1^2}\right).$$

We also note that, since these parameters are estimated in non-overlapping samples, conditioning on $\hat{\beta}_1$ adds no additional information to the posterior distribution of $\hat{\beta}_2$ after already conditioning on β_1:

$$f(\hat{\beta}_2 \mid \beta_1, \hat{\beta}_1, C) = f(\hat{\beta}_2 \mid \beta_1, C).$$

To derive this conditional distribution, we decompose β_1 and $\hat{\beta}_2$ as

$$\begin{aligned} \beta_1 &\equiv \tau_1 Z_1 \\ \hat{\beta}_2 &\equiv \tau_2 r_g Z_1 + \sqrt{\tau_2^2(1 - r_g^2) + \sigma_2^2} Z_2. \end{aligned}$$

where Z_1 and Z_2 are independent standard normal random variables. (The coefficient on Z_1 in the first equation is defined so that $\mathrm{Var}(\beta_1) = \tau_1^2$, and the coefficients in the second equation are defined so that $\mathrm{Var}(\hat{\beta}_2) = \tau_2^2 + \sigma_2^2$ and $\mathrm{Cov}(\beta_1, \hat{\beta}_2) = \tau_1 \tau_2 r_g$.) Fixing β_1 and substituting for Z_1,

$$\hat{\beta}_2 \equiv \frac{\tau_2}{\tau_1} r_g \beta_1 + \sqrt{\tau_2^2(1 - r_g^2) + \sigma_2^2} Z_2,$$

implying that

$$(\hat{\beta}_2 \mid \beta_1, C) \sim N\left(\frac{\tau_2}{\tau_1} r_g \beta_1, \tau_2^2(1 - r_g^2) + \sigma_2^2\right).$$

Using the conditional distributions defined above, we calculate

$$\begin{aligned} f(\beta_1 \mid \hat{\beta}_1, \hat{\beta}_2, C) &\propto f(\hat{\beta}_2 \mid \beta_1, C) f(\beta_1 \mid \hat{\beta}_1, C) \\ &\propto \exp\left\{-\frac{1}{2}\left[\frac{\left(\hat{\beta}_2 - \frac{\tau_2}{\tau_1} r_g \beta_1\right)^2}{\tau_2^2(1 - r_g^2) + \sigma_2^2}\right]\right\} \exp\left\{-\frac{1}{2}\left[\frac{\left(\beta_1 - \frac{\tau_1^2}{\tau_1^2 + \sigma_1^2}\hat{\beta}_1\right)^2}{\frac{\tau_1^2 \sigma_1^2}{\tau_1^2 + \sigma_1^2}}\right]\right\} \\ &= \exp\left\{-\frac{1}{2}\left[\frac{\left(\hat{\beta}_2 - \frac{\tau_2}{\tau_1} r_g \beta_1\right)^2}{\tau_2^2(1 - r_g^2) + \sigma_2^2} + \frac{\left(\beta_1 - \frac{\tau_1^2}{\tau_1^2 + \sigma_1^2}\hat{\beta}_1\right)^2}{\frac{\tau_1^2 \sigma_1^2}{\tau_1^2 + \sigma_1^2}}\right]\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left[\frac{\left(\frac{\tau_2^2}{\tau_1^2} r_g^2 \beta_1^2 - 2\frac{\tau_2}{\tau_1} r_g \beta_1 \hat{\beta}_2\right)}{\tau_2^2(1 - r_g^2) + \sigma_2^2} + \frac{\left(\beta_1^2 - 2\frac{\tau_1^2}{\tau_1^2 + \sigma_1^2}\beta_1 \hat{\beta}_1\right)}{\frac{\tau_1^2 \sigma_1^2}{\tau_1^2 + \sigma_1^2}}\right]\right\} \\ &= \exp\left\{-\frac{1}{2}\left[\frac{\beta_1^2 - 2m(\hat{\beta}_1, \hat{\beta}_2)\beta_1}{s(\hat{\beta}_1, \hat{\beta}_2)}\right]\right\} \end{aligned}$$

$$\propto \exp\left\{-\frac{1}{2}\left(\frac{\left[\beta_1 - m(\hat{\beta}_1, \hat{\beta}_2)\right]^2}{s(\hat{\beta}_1, \hat{\beta}_2)}\right)\right\},$$

where we define the functions

$$m(\hat{\beta}_1, \hat{\beta}_2) \equiv \left(\frac{\tau_1^2(\tau_2^2 + \sigma_2^2) - r_g^2\tau_1^2\tau_2^2}{(\tau_1^2 + \sigma_1^2)(\tau_2^2 + \sigma_2^2) - r_g^2\tau_1^2\tau_2^2}\right)\hat{\beta}_1 + \left(\frac{r_g\tau_1\tau_2\sigma_1^2}{(\tau_1^2 + \sigma_1^2)(\tau_2^2 + \sigma_2^2) - r_g^2\tau_1^2\tau_2^2}\right)\hat{\beta}_2$$
$$s(\hat{\beta}_1, \hat{\beta}_2) \equiv \frac{\tau_1^2\sigma_1^2(\tau_2^2 + \sigma_2^2 - r_g^2\tau_2^2)}{(\tau_1^2 + \sigma_1^2)(\tau_2^2 + \sigma_2^2) - r_g^2\tau_1^2\tau_2^2}.$$

Using this substitution, the above derivation implies that

$$(\beta_1 \mid \hat{\beta}_1, \hat{\beta}_2, C) \sim N\left[m(\hat{\beta}_1, \hat{\beta}_2), s(\hat{\beta}_1, \hat{\beta}_2)\right], \tag{11}$$

giving us the distribution of β_1 when the SNP is non-null. Using (10) and (11), we can calculate many different values of interest, such as the conditional expectation,

$$E(\beta_1 \mid \hat{\beta}_1, \hat{\beta}_2) = m(\hat{\beta}_1, \hat{\beta}_2)P(C \mid \hat{\beta}_1, \hat{\beta}_2), \tag{12}$$

or the expected R^2 of the SNP for phenotype 1,

$$\begin{aligned} E(R_1^2 \mid \hat{\beta}_1, \hat{\beta}_2) &= E(R_1^2 \mid \hat{\beta}_1, \hat{\beta}_2, C)P(C \mid \hat{\beta}_1, \hat{\beta}_2) \\ &= E(\beta_1^2 \mid \hat{\beta}_1, \hat{\beta}_2, C)\sigma_x^2 P(C \mid \hat{\beta}_1, \hat{\beta}_2) \\ &= \left[m(\hat{\beta}_1, \hat{\beta}_2)^2 + s(\hat{\beta}_1, \hat{\beta}_2)\right]\sigma_x^2 P(C \mid \hat{\beta}_1, \hat{\beta}_2), \end{aligned} \tag{12}$$

where $\sigma_x^2 = 2\,\mathrm{maf}(1 - \mathrm{maf})$ is the variance of the genotype. The final step in this derivation follows from the definition of the variance of a random variable: for a random variable ζ, rearranging $\mathrm{Var}(\zeta) \equiv \mathrm{E}[\zeta^2] - \mathrm{E}[\zeta]^2$ gives $\mathrm{E}[\zeta^2] = \mathrm{E}[\zeta]^2 + \mathrm{Var}(\zeta)$.

D. Application

Selection of prior

As in all Bayesian analyses, our results partly depend on our prior, which is parameterized by π, τ_1^2, τ_2^2, and r_g. We choose our prior empirically using a simple maximum likelihood framework.

For any SNP i, we denote the variance of its genotype by σ_{xi}^2, which, as noted above, equals $2\,\mathrm{maf}_i(1 - \mathrm{maf}_i)$. For a given phenotype, we make the standard assumption[28,48] that the variance of effect sizes across non-null SNPs that have minor allele frequency maf_i is inversely proportional to σ_{xi}^2. To be precise, we define a (phenotype-specific) constant $t^2 \equiv \tau_i^2\sigma_{xi}^2$, where τ_i^2 is the variance of effect sizes across non-null SNPs that have minor allele frequency maf_i. (This assumption implies that rare variants are more likely to have large effects.)

Note that in a GWAS conducted in a large sample of size N_i, the *z*-statistic for SNP i, z_i, has distribution

$$z_i \sim \begin{cases} N(0, 1 + N_i t^2) & \text{with probability } \pi \text{ (the SNP is non} - \text{null)} \\ N(0, 1) & \text{with probability } (1 - \pi) \text{ (the SNP is null).} \end{cases}$$

Thus, we can construct a log-likelihood function from the *z*-statistics of the measured SNPs in the GWAS:

$$LL(z_i;\pi,t^2)=\sum_i \log\left[(1-\pi)\frac{1}{\sqrt{1+N_i t^2}}\phi\left(\frac{z_i}{\sqrt{1+N_i t^2}}\right)+\pi\phi(z_i)\right].$$

By maximizing this function over π and t^2, we can obtain estimates of these parameters. As a robustness analysis, we can also fix π at a hypothesized value and maximize over only t^2.

Note that we do not take into account here that the *z*-scores are correlated across SNPs that are in linkage disequilibrium. The maximum-likelihood procedure will nonetheless generate unbiased point estimates of the parameters. The usual maximum-likelihood standard errors will be incorrect, but since we only use the point estimates, we make no effort to account for the correlation across SNPs.

Under the assumption that a SNP is null for one phenotype if and only if it is null for the other phenotypes, we restrict the model so that π is equal for each phenotype and estimate the parameters pooling the data from all three GWAS. The resulting estimates are:

$$\pi = 0.3157$$
$$t^2_{SWB} = 2.609\times10^{-6}$$
$$t^2_{dep} = 2.950\times10^{-6}$$
$$t^2_{neur} = 6.268\times10^{-6}.$$

For the prior parameter r_g, we use our empirically estimated genetic correlations reported in **Supplementary Table 1**.

First-stage credibility of GWAS results

Using equation (1) above and the prior parameters estimated in the previous subsection, we calculate the first-stage credibility of the GWAS results; that is, we calculate the posterior probability that each lead SNP is non-null given the observed GWAS estimate for that SNP. These results are reported in **Supplementary Table 14**. As can be seen, the first-stage credibility is above 99% for all but the two DS SNPs, which have credibility 98.1% and 96.6%.

We may worry to what degree these credibility results are driven by our estimate of the prior probability that the SNP is non-null. As a robustness analysis, we also evaluated the first-stage credibility over a range of values for $\pi \in [0.01, 0.99]$. **Supplementary Figures 12a**, **12c**, and **12e** contain plots of the posterior probability that the SNP is null over these values for the three SWB lead-SNPs. As can be seen, except local to the prior that all SNPs are null, the *maximum* posterior null probability is around 1%, and over a wide range of priors it is much less than that.

Panels B, D, and F of that same figure also show the posterior mean and posterior credible intervals, evaluated according to (2), (3), and (4). We see that the posterior credible intervals do not overlap zero for all of the priors tested (again with the exception of the priors local to $\pi = 0$), even though the posterior mean is about one third the size of the estimated coefficient in each case. This implies that while the effect sizes of the three SWB lead-SNPs are probably inflated by the Winner's Curse, the associations themselves are very unlikely to be chance findings.

We conduct the same robustness checks for the DS and neuroticism lead-SNPs and find similar results: in the range $\pi \in [0.01, 0.99]$, the lowest credibility estimate for DS is 95.3%, and the lowest for neuroticism is 98.6%.

We note that, as can be seen in the figure, the posterior null probability is not monotonic in π. That is because there are two opposing factors influencing this probability: first, as we increase the prior non-null probability, this has a direct effect of increasing the posterior non-null probability. Second, increasing the prior non-null probability decreases the prior non-null variance (in order to match the empirical variance), reducing the amount of Bayesian shrinkage of the estimate and thereby reducing the posterior null probability. As a result, the posterior null probability may increase or decrease in π, depending on which of the above two effects dominate.

The posterior credible intervals, however, appear to have a monotonic relationship over most of the range of priors considered. This suggests that, except for extreme priors that assume that over 99% of SNPs are null, the second effect described above has the dominant impact on the overall shape of the posterior distribution.

Expected quasi-replication record

In **Supplementary Note 8B**, we derive a formula for the expected replication record. Specifically, we can calculate the probability that the estimated effect of a SNP on the second-stage phenotype will have the same sign as the estimated effect on the first-stage phenotype. This probability is given by equation (7). We similarly derive the probability that the second-stage estimate will have a *p*-value smaller than a certain threshold in equation (8). In our case, we focus on the *p*-value threshold of 0.05. Plugging our estimates and prior parameters into these equations and summing across our lead SNPs will give us the expected number of SNPs meeting each criterion.

Using this method, we calculate that we should expect to see the predicted sign of the second-stage GWAS estimates, given the first-stage estimate, in 16.7 cases (with a standard deviation of 2.0 cases). This is consistent with the 16 cases observed in our quasi-replication. The expected number of SNPs in which the second-stage estimate has a *p*-value smaller than 0.05 is 6.9 cases (with a standard deviation of 2.1). We observe that 7 SNPs replicate at this level—again, quite close to the expected number.

The similarity of the observed quasi-replication record to the record that would be expected if our lead SNPs were non-null increases our confidence that the associations are not chance findings. It additionally suggests that our Bayesian model is a reasonable approximation, strengthening our confidence in the results of analyses we conduct with our Bayesian framework.

Combined credibility results

Finally, as described at the beginning of this section, we can leverage the additional information in the GWAS of the second-stage phenotype to update the credibility from the GWAS of the first-stage phenotype. (In some cases, the lead SNP itself is not available in the second-stage data. In these cases, we conduct the credibility analysis using the SNP in highest linkage disequilibrium with the lead SNP that is available in both sets of results.)

The results of this analysis are found in **Supplementary Table 14**. As can be seen, although the credibility of these SNPs from just the first-stage GWAS is very high, adding information from the results of the second-

stage GWAS increases the credibility in 13 out of 19 cases. The few cases where the credibility falls are cases where the second-stage estimate is very small in magnitude or has the opposite sign than what would be predicted from the first-stage estimate. Despite these cases, the credibility for all SNPs remains above 98%.

Supplementary Table 33 reports the average degree of inflation of the GWAS estimates due to the Winner's Curse, as measured by a comparison of the GWAS estimates and their respective posterior estimates. We begin by taking the ratio of these estimates for each SNP. We would like to then take the geometric mean of these ratios, but in some cases this ratio is negative, so the geometric mean is not defined or not sensible. We therefore proceed in two different ways: first, we take the absolute value of each of the ratios before calculating the geometric mean. Second, we calculate the geometric mean among only the set of SNPs for which the ratio is positive. We report the average inflation of the GWAS estimates using both approaches.

For both approaches, to estimate confidence intervals, we simulate GWAS and posterior estimates. Simulated GWAS estimates are drawn from a normal distribution with mean equal to the GWAS estimate and standard deviation equal to the estimated standard error. Simulated posterior estimates are drawn from the posterior distribution for non-null SNPs. (We ignore the small probability that the true effect on the second-stage phenotype is null because the geometric mean of the ratios would be undefined in that case.) We conduct one million replications of this simulation. We report the 2.5^{th} and 97.5^{th} percentiles of the simulations as the bounds of the 95% confidence interval.

For the first-stage estimates, the two approaches give the same answer for the average inflation (because the GWAS estimate and the posterior mean necessarily have the same sign): 2.276. That is, the GWAS estimates appear to be approximately twice as large as the true effect sizes.

For the second-stage estimates, both approaches yield a point estimate that suggests that the GWAS estimates are roughly half the size of the expected true parameter. This relationship, however, is very imprecise, with a 95% confidence interval spanning from 0.523 to 1.370. This imprecision is largely driven by two outlier SNPs (rs3756290 and rs6808710), which have negligibly small second-stage GWAS estimates. When these SNPs are omitted, the average inflation rises to 0.846 with a confidence interval of [0.543, 1.427]. All these 95% confidence intervals for the average inflation of our second-stage estimates include 1. This is consistent with the second-stage being less vulnerable to bias due to the winner's curse, a property that holds as well for traditional replication analyses.

9. Biological annotation

A. Functional partitioning of heritability

Background and methods

First, we discuss the results of our analyses of functional genomic regions enriched for GWAS signals. These regions include, for example, evolutionarily conserved regions or regions epigenetically regulated to be accessible in specific tissues (brain, immune system, etc.). Using LD Score regression[28,49], we aim to determine whether specific types of genomic regions are enriched in their association with SWB, DS, or neuroticism. Enrichment of specific annotations may shed light on the biological processes underlying the phenotypes analyzed here.

Stratified LD Score regression is based on the relationship

$$E[\chi_j^2] = N\sum_c^C \tau_c \ell(j,c) + Na + 1,$$

where $\chi_j^2 = N\hat{\beta}_j^2$ is the GWAS chi-square statistic for SNP *j*, *N* is the sample size, *c* indexes the functional categories (which do not have to be disjoint), $\ell(j,c)$ is the stratified LD Score of SNP *j* with respect to functional category c, τ_c is the average contribution to heritability of a SNP due to its membership in category c, and a is a term that measures the contribution of confounding biases such as cryptic relatedness and population stratification.

Finucane et al.[49] present derivations of this equation, and they show how estimates of τ_c that result from estimating the implied regression can be used to obtain estimates of the heritability ascribable to the various functional categories. Enrichment is then calculated for each functional category as the fraction of the total heritability captured by the category divided by the fraction of SNPs in that category.

To partition the SNP-based heritability of SWB, DS, and neuroticism using the results of our GWAS meta-analyses, we followed exactly the procedure described by Finucane et al.[49]. We used the stratified LD Scores calculated from the European-ancestry samples in the 1000 Genomes Project (1000G), but in the regressions themselves took forward only the chi-square statistics of the ~1 million HapMap3 SNPs with minor allele frequency (MAF) > 0.05 (because the LD Scores of SNPs with low MAFs can introduce a great deal of sampling variation).

We first estimated the stratified LD Score regression for the "baseline" model, in which the functional categories consist of one category consisting of all SNPs, 24 categories corresponding to 24 main annotations of interest, categories corresponding to 500-bp windows around regions belonging to each of these 24 annotations, and categories corresponding to 100-bp windows around ChIP-seq peaks (regions that are DNase hypersensitive or associated with histones bearing the marks H3K4me1, H3K4me3, or H3K9ac). There were thus 53 predictor variables in total.

"Histone marks" are posttranslational modifications of histones that alter their interaction with the DNA wound around them. The SNPs bearing the annotations referring to histone marks in the baseline model just described, however, are a union of SNPs located in regions associated with the defining mark in any cell type whatsoever. To gain tissue-level resolution, we followed the analysis of Finucane et al.[49] and grouped 220 distinct types of histone marks—defined by both mark and cell type—into 10 broad tissue types (ADRENAL/PANCREAS, CENTRAL NERVOUS SYSTEM, CARDIOVASCULAR, CONNECTIVE/BONE, GASTROINTESTINAL, IMMUNE/HEMATOPOIETIC, KIDNEY, LIVER, SKELETAL MUSCLE, and OTHER). We then added each of these 10 tissue annotations to the baseline model, one at a time, and assessed the magnitude and statistical significance of the enrichment thus observed. To benchmark these results, we downloaded the summary statistics of a recent GWAS meta-analyses of height[95] (http://www.broadinstitute.org/collaboration/giant/index.php) and applied the tissue-level analysis to these phenotypes. The sample sizes employed in these three meta-analyses are similar to our own and therefore enable an informative comparison.

Results: The "baseline" model

The results for the baseline model for SWB, DS, and neuroticism are shown in **Supplementary Tables 18-20**). To correct for multiple hypothesis testing within our main phenotype (SWB), we examined whether

the enrichment estimates were significant after a Bonferroni correction for 62 two-sided tests (i.e., 52 annotations in the baseline model[q] and 10 tissue types). The resulting significance threshold is $p < 0.05/62 = 8.1\times10^{-4}$. It can be seen from the last column of the table for SWB that 8 baseline annotations met this threshold for SWB, and four baseline annotations were significant for each of DS and neuroticism. Regions that are evolutionarily conserved in mammals (labeled CONSERVED in **Supplementary Table 18**) exhibited the strongest enrichment for all three phenotypes (SWB ∼14-fold, DS ∼12-fold, and neuroticism ∼12-fold). Evolutionarily conserved regions of the genome accumulate base-pair substitutions differentiating distinct species more slowly than predicted by a model of selective neutrality, which implies that mutations in such regions tend to have phenotypic effects that are visible to natural selection. The functional category comprising 500bp windows around conserved regions (CONSERVED + 500bp) was significantly enriched for all three phenotypes (albeit with less extreme enrichment). Coding SNPs (labeled CODING in **Supplementary Tables 18-20**), which tend to be highly conserved, also exhibited enrichment for SWB (∼7-fold) and DS (∼7-fold), but these results were not significant after Bonferroni correction.

The functional category showing the next most statistically significant enrichment for SWB, after evolutionarily conserved regions, was the histone mark H3K4me3 (∼3.6-fold, $p = 2.22\times10^{-5}$). This histone modification is associated with increased expression of nearby genes. Indeed, a common theme of nominally significant annotations was residence upstream of protein-coding genes and likely regulation of their expression (DNASE I HYPERSENSITIVE, FETAL DNASE I HYPERSENSITIVE, TRANSCRIPTION START SITE, ENHANCER). Several other histone marks also showed significant enrichment for SWB but not DS or neuroticism. These included monomethylation of histone 3 lysine 4 (H3K4me1) peaks, which are associated with enhancer regions, and 500bp windows around acetylation of histone 3 lysine 9 (H3K9ac) marks. The acetylation of lysine may facilitate gene expression by reducing the electrical attraction between DNA and the histone residue.

The other baseline results for DS are also broadly similar to those for SWB: annotations referring to the evolutionarily conserved regions, to the histone marks H3K9ac and H3K4me3, and to the coding SNPs are enriched. These same annotations are also enriched for neuroticism but less so.

Results: tissue types

The results of the tissue-level analysis are reported in Panel C of **Supplementary Table 24**. It is the enrichment of ADRENAL/PANCREAS that is strongest when the phenotype is SWB (∼3.7-fold). The enrichment statistics are potentially misleading, however, because of possible confounding. For example, many SNPs are associated with histone marks observed in multiple tissues. It is thus of interest to examine the τ_c of each tissue, which are the coefficients from the stratified LD score regression and are the expected increase in the phenotypic variance accounted for by a SNP due to the SNP's being in a given tissue category, controlling for the annotations in the baseline model. (The τ_c for a given tissue type and phenotype also corresponds to the effect of a one-unit increase in a SNP's stratified tissue-specific LD score on the expected square of the SNP's GWAS estimate from the phenotype's GWAS, where the SNP has been standardized and controlling for the annotations in the baseline model.) SNPs that bear a tissue annotation with a large and positive τ_c will tend to account for a larger share of a phenotype's heritability. **Figure 4a**

[q] The baseline model include 53 predictor variables: one for each of the 52 annotations and one for the set of all SNPs. We adjust the significance threshold for the 52 annotations, but not for the set of all SNPs.

shows the τ_c's divided by the LD score estimates of the heritability of the phenotypes for which they were estimated[r,s], and the τ_c's are reported in **Supplementary Table 21**. (We normalize the τ_c's in **Figure 4a** by dividing them by the LD Score heritability[t] of the phenotypes for which they were estimated to increase comparability across phenotypes.) Now we see that it is CENTRAL NERVOUS SYSTEM that has the largest coefficient (16×10^{-9}), although this is not significantly different from that of the runner-up ADRENAL/PANCREAS (10×10^{-9}). The CENTRAL NERVOUS SYSTEM annotation is also the most enriched by this measure when the phenotype is neuroticism. In fact, here the τ_c coefficient of CENTRAL NERVOUS SYSTEM (29×10^{-9}) is the only one attaining a positive value. In this respect neuroticism is quite unlike height, where CENTRAL NERVOUS SYSTEM is the only tissue-level annotation with a *negative* coefficient. For DS, CENTRAL NERVOUS SYSTEM is also significantly enriched and has the second largest τ_c coefficient; ADRENAL/PANCREAS and KIDNEY are also enriched and have large and positive τ_c coefficients.

Discussion

The baseline results for SWB, DS, and neuroticism are broadly similar to those reported earlier for other phenotypes. For instance, Okbay et al.[22] also find that the annotation for evolutionarily conserved regions is the most highly enriched for educational attainment (and the most significantly so), and that annotations associated with coding SNPs and with the histone marks H3K9ac and H3K4me3 are enriched. Finucane et al.[49] take the average enrichment across nine different phenotypes and find that conserved regions are the most enriched, followed by coding SNPs; they also find that annotations associated with the histone marks H3K9ac and H3K4me3 are enriched.

Significant enrichment of CENTRAL NERVOUS SYSTEM for SWB, DS, and neuroticism is clearly in line with previous enrichment results from educational attainment, schizophrenia, and bipolar disorder[49]. Enrichment of the CENTRAL NERVOUS SYSTEM indicates that the signal captured in these GWAS indeed is stronger in regions in the genome with likely effects in the brain. On the other hand, enrichment of the ADRENAL/PANCREAS for SWB and DS specifically is convergent with previous research on the role of hypothalamic-pituitary-adrenal (HPA)-axis and autonomic dysregulation in depressed patients[115]. Moreover, we note that ADRENAL/PANCREAS is not significant in schizophrenia nor bipolar disorder[49], although both disorders have also been associated with HPA-axis functioning[116–118]. Thus, enrichment of the ADRENAL/PANCREAS appears to be a feature that might distinguish DS and SWB from genetically similar psychiatric disorders.

The specific involvement of HPA- and/or autonomic dysregulation in depression might be the reason behind the enrichment of the ADRENAL/PANCREAS. However, we note the HPA-axis is a complex pathway that

---

[r] Based on the LD Score framework, $h^2_{LDSC,y} = \sum_c M_c\, \tau_{c,y}$, where *y* denotes the phenotype and M_c is the number of SNPs with annotation *c* among the SNPs used to calculate the LD Scores. Thus, normalizing the τ_c's by $h^2_{LDSC,y}$ is equivalent to normalizing the τ_c's by a weighted sum of the τ_c's, where the weights are given by the number of SNPs with the different annotations.

[s] The confidence intervals were obtained by the delta method, assuming zero covariance between the τ_c's and the phenotypes' LD Score heritabilities.

[t] Each phenotype's LD Score heritability was obtained from the phenotype's baseline stratified LD Score regression.

has intricate yet widespread effects on the brain and endocrine and immune system; its precise involvement in depression is not clearly understood. Moreover, evidence for dysregulation of the autonomic nervous system in depressed patients is inconsistent[115].

In brief, HPA-dysregulation is thought to influence depression through maladaptive initiation and termination of the human response to stress[115,119]. This notion is supported by the consistent association between depression and elevated blood serum levels of cortisol[120]. Cortisol is produced in the adrenal cortex, and is the end product of the HPA-axis. Its function is to mobilize glucose stores into the blood, suppress the immune system, and provide negative feedback to the brain.

Two other hormones produced in the adrenal gland, and potentially involved in depression, are epinephrine and norepinephrine. As opposed to cortisol, they are produced in the adrenal medulla, and their release from the adrenals is the result of direct autonomic nervous system stimulation. This direct neural activation results in a fast release of (nor)epinephrine into the blood (in contrast, cortisol release from the adrenals is dependent on activation by circulatory adrenocorticotropic hormone (ACTH), a much slower system). In this way, epinephrine and norepinephrine help trigger the *acute* stress system, which is characterized by the "fight-or-flight" response and involves increases in heart rate, blood pressure, sweating, and so on. Dysregulation of the autonomic nervous system has also been linked to depression, albeit inconsistently[115].

The mechanism that accounts for the enrichment of the KIDNEY for DS is unclear. One intriguing but speculative possibility involves the hormone vasopressin. There is clear evidence for increased hypertension in depressed patients[121], and a potential culprit for the link between hypertension, kidney function, and depression is the hormone vasopressin. The hormone vasopressin is involved in the human stress response[119] and has two main functions in the human body: it primarily regulates water reabsorption by the kidney, and it increases peripheral vascular constriction, which increases arterial pressure. We emphasize, however, that this potential mechanism is purely speculative: the enrichment found for KIDNEY may also be an artifact related to the physical vicinity of the kidney to the adrenal glands (which are positioned, as their name describes, on top of the kidneys).

In the next section, we describe follow-up analyses that we conducted to gain biological insight into the 20 SNPs in **Table 1**. In short, we ascertained whether any genome-wide significant SNPs (or variants in strong LD with those SNPs) fall into one of the following three classes: (1) nonsynonymous SNPs that alter the composition of the protein encoded by a gene, (2) eQTLs that are associated with the abundance of mRNA transcript in whole blood, and (3) genome-wide significant SNPs associated with other phenotypes in previously conducted GWAS. In addition, we obtained a list of all genes overlapping the loci defined by the 20 SNPs listed in **Table 1**, and queried the Gene Network database (GO, KEGG, and Reactome) with these in order to obtain insight into their functions.

B. LD with non-synonymous variants

Generating LD partners

In order to obtain a uniform list of "LD partners" for the 20 SNPs in **Table 1**, we used the 1000 Genomes Phase I CEU reference panel and performed the "ld"-command in PLINK[50]. This list was used for the nonsynonymous SNP look up described below. A SNP was considered an LD partner if it is in strong LD ($R^2 \geq 0.6$) with the queried SNP and within 250kb of it.

Nonsynonymous variants

We used the tool HaploReg[122] to identify nonsynonymous variants in strong LD with any of the 20 queried SNPs and their LD partners. Here, we report that one SWB SNP (rs2075677), two neuroticism SNPs (rs193236081 and rs10838738), and one neuroticism and DS SNP (rs6904596) are in strong LD with nonsynonymous variants (**Supplementary Table 23**).

Our most notable finding here is neuroticism SNP rs193236081, which tags 11 missense variants in near-perfect LD (R^2 = 0.97) residing in *MAPT*, *STH*, or *KANSL1*. These genes lie in the large chromosome 17 inversion we detected (see **Supplementary Note 5.B**). A missense variant is a type of nonsynonymous mutation that results in the production of a different amino acid. According to the Gene Network co-expression database (described below in **Supplementary Note 9.E**), *MAPT* and *STH* are heavily expressed in many regions of the brain. While *STH*'s specific function is unclear, *MAPT* is significantly involved in a myriad of neuronal processes (e.g., axonogenesis). In fact, *MAPT* is the gene that encodes tau, a protein that is important in the stabilization of microtubules in neurons. Improper functioning of tau may be an important pathological mechanism for disorders such as Alzheimer's and several forms of parkinsonism (which are collectively known as "tauopathies")[123]. In previously conducted GWAS in the literature, *MAPT* has been associated with neurodegenerative disorders, as well as intracranial volume in healthy individuals (**Supplementary Table 22**). Finally, we report that *KANSL1* (known by its synonym *KIAA1267* in the Gene Network database) is highly expressed in the cerebellum, and is predicted to be involved in gene transcription (e.g., CHROMATIN and HISTONE MODIFICATION).

Neuroticism and DS SNP rs6904596 tags two missense variants in *OR2B2* (R^2 = 0.9) and one in *HIST1H2BL* (R^2 = 0.84). *OR2B2* encodes an olfactory receptor that is involved in smell and taste, while *HIST1H2BL* encodes a histone protein. Both genes are located in the HLA-region on chromosome 6.

Neuroticism SNP rs10838738 is in perfect LD (R^2 = 1) with a missense SNP residing in *MTCH2*. This gene is highly expressed in neural stem cells and astrocytes, and is predicted to be involved in energy metabolism in the mitochondria. rs10838738 itself has significantly been associated with BMI in previous GWAS (see **Supplementary Table 22**).

Lastly, we report that SWB SNP rs2075677 tags two missense variants that reside in *ZNFX1* (R^2 = 0.84) and *DDX27* (R^2 = 0.86), respectively. Their functional relevance for SWB is not directly apparent. *DDX27* is predicted to be involved in RNA processing, and is highly expressed in specific brain regions, for instance the visual cortex. *ZNFX1* is predicted to be involved in interferon signaling and production, and is mainly expressed in immune cells.

C. eQTL analyses

Blood cis-eQTL lookup

The goal of the eQTL lookup performed here is to explore the extent to which the SNPs associated with our phenotypes of interest influence the expression level of certain genes, and thus the quantity of protein produced. This section also aims to elucidate which specific gene transcripts are associated with our query SNPs and thus are potentially involved in the biology of SWB, DS, and neuroticism. It is important to realize that gene expression, and its genomic regulation, can differ across different tissues. Here we test for associations between SNPs and gene expression in whole blood, which means genes expressed in white

blood cells (cells involved in the human immune system). The immune system might very well be involved in SWB; however, gene expression in the brain may be more likely to be involved in the biological processes underlying SWB, DS, and neuroticism. Fortunately gene expression across tissues is known to substantially overlap[124]. Whole blood is often considered an excellent tissue for the purpose of studying the genetics of gene expression, as obtaining whole blood is substantially less invasive then obtaining other tissues such as brain tissue.

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. The Affymetrix platform measures RNA using 423,201 probes that can be grouped into 44,241 probesets targeting 18,238 genes. In *cis*-eQTL analysis, probeset-SNP associations are tested. Genotypes were imputed using the 1000 Genomes phase 1 version 3 (CEU) reference panel. We deemed a SNP a potential *cis*-eQTL when the distance between the SNP and the midpoint of the probeset was smaller than 1Mb. The true maximum distance between a *cis*-eQTL and its regulatory target is a matter of arbitrary definition[125], and we elected to relax the required proximity. The full RNA and DNA extraction procedure and *cis*-eQTL analysis procedure are described elsewhere[126]. Permutation tests were performed to control for the presence of related individuals in the eQTL discovery sample. FDR-corrected *p*-values corrected for the total number of probeset-SNP tests were performed across all 44,241 probesets.

In total 7 out of 19 query SNPs were found to be significantly associated with the expression of 19 genes in whole blood (see **Supplementary Table 24**). Some interesting findings are summarized here. Note that the eQTL results of the SNPs located in and around the inversion on chromosome 8 can be found in **Supplementary Note 5.A.**

We find that SWB SNP rs2075677 affects the expression levels of *ZNFX1* and *DDX27* in the blood. As noted above, this SNP also tags missense variants in these genes, and has been implicated in height. SWB SNP rs3756290 is an eQTL for *ACSL6* and *IRF1. ACSL6* encodes an enzyme that synthesizes acyl-CoA, and is highly expressed in many brain areas. This enzyme is essential for fatty acid metabolism. A noteworthy observation is that cholesterol is an essential component of myelin, a structural component of nerves that is crucial for signal conduction[127]. Moreover, cholesterol stabilizes serotonin receptors on the membrane, and appears to be important for normal serotonin receptor function[128–131]. Notably, serotonin dysregulation is hypothesized to play a role in depression and anxiety, and many first-line antidepressants may exert their therapeutic effect by increasing serotonin concentrations in the synaptic cleft (although the exact therapeutic mechanism is still under heavy debate)[132].

IRF1, on the other hand, is mainly expressed in the immune system, and is known to play a role in a wide range of immune signaling pathways (in particular interferon signaling), as well as regulation of cell proliferation and tumor necrosis[133]. Interferons are pro-inflammatory cytokines that are naturally produced by the body, although they are also used as medication in the treatment of hepatitis and various cancers. They are of limited clinical utility because they commonly induce symptoms of clinical depression in patients (as well as in healthy individuals in experimental settings)[134]. Concurrent use of anti-depressants lowers the chances of developing depression[135]. However, naturally-circulating interferon levels do not appear to correlate with depression-status[136].

Neuroticism SNP rs139237746 affects the expression of *ADM* (“adrenomedullin”), a gene that derives its name from its initial discovery in epinephrine-secreting tumors and encodes a vasodilatory hormone. ADM

is also a neuropeptide that has been described to be critical for neuroprotection following brain injury[137]. It is ubiquitously expressed in the body, and helps cells to respond to hypoxic injury induced by oxidative stress. Central nervous system-knockout of this gene in mice increases anxious behavior, changes the structure of neuronal microtubules, and reduces the adaptive response to hypoxia in the brain[138]. rs139237746 is also an eQTL for *SBF2,* a gene that encodes an enzyme that regulates the activity of myotubularin proteins; a myotubularin-encoding gene was also implicated in our eQTL analysis of the inversion on chromosome 8 (**Supplementary Note 5.A**). Neuroticism SNP rs10838738 is an eQTL for *MTCH2,* and also tags a missense variant in this gene. Moreover, this SNP also affects the expression of *MYBPC3*, a gene important for cardiac muscle function. However, it is also highly expressed in several brain regions that are important for motor function (i.e., the basal ganglia and motor cortex), although its function there is unknown. Lastly, neuroticism SNP rs193236081 is an eQTL for both *MAPT* and *KANSL1*, and also tags missense variants in these genes.

Tissue-specific cis-eQTL lookup

We supplemented the analyses described in the previous section with eQTL lookups of the 20 SNPs (including the chromosome 8 inversion-tagging SNP) in the Genotype-Tissue Expression Portal (www.GTExportal.org)[34,139]. For a given SNP, the portal provides results for tests of association between the SNP and gene expression in various tissues of genes whose transcription start site is within one Mb of the SNP. 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, nerve (tibial), thyroid, adrenal gland and pituitary. (We included pituitary tissue and thyroid tissue because these feature prominently in many biological-psychiatric theories of depression and related disorders[140,141] and pituitary gland tissue because of the central role the pituitary plays in the HPA axis[142].) The sample sizes vary by tissue type. The brain-tissue analyses were based on samples in the range 70 to 103, whereas the analyses of the three non-brain tissues were based on samples in the range 87 to 278.

Panel C of **Supplementary Table 24** reports the results from the portal for the 20 SNPs for the trait-relevant tissues (the *p*-values for the test of associations whose results are returned by the portal are all smaller than 5×10^{-5}). Despite the small samples, five of the queried SNPs—rs10838738, rs12938775, rs193236081, rs6904596, rs2572431—are significantly associated with gene expression levels in the trait-relevant tissues.

Interestingly, we observe the strongest results for the two SNPs tagging the inversions on chromosomes 8 (rs2572431) and 17 (rs193236081). The chromosome 8 tagging SNP is an eQTL for the expression of several genes in thyroid and tibial nerve tissue. The chromosome 17 tagging SNP is associated with the expression of 17 distinct genes in thyroid tissue, 8 distinct genes in pituitary tissue and is also strongly associated with the expression of a large number of genes in most of brain tissues queried. This lookup confirms the SNPs identified in our analyses are associated with gene expression in brain and nerve tissues as well as in the adrenal gland, pituitary and thyroid. The genes identified here could inform future efforts to determine the functional effects of specific genes.

D. GWAS catalog lookup

We scanned the NHGRI GWAS catalog[143] (http://www.genome.gov/gwastudies, accessed 10 September 2015) for any overlap between previous GWAS findings and the 20 SNPs in **Table 1** (or SNPs in close LD with these SNPs). Since the GWAS catalog does not always include the most recent GWAS results available, we also scanned the most recent GWAS on height[95], body mass index (BMI)[100], waist-to-hip ratio (WHR) adjusted for BMI[144], schizophrenia[25], Parkinson's disease[145], Alzheimer's disease[96], and brain volume[146] for any overlap with our phenotypes.

Before describing the results, we note that any overlap at a locus for two phenotypes does not imply sharing of important underlying genetic or biological etiological mechanisms. For instance, it is possible that a SNP associated with two phenotypes tags distinct causal genes affecting each of the different phenotypes. Moreover, the issue of pleiotropy should also be kept in mind. As Wagner and Zang[147] point out, the human phenotypic space is very high dimensional, while the human genome contains finite numbers of common variants and protein-coding genes.

But even when two phenotypes do share a causal site, this does not mean that the multiple phenotypes share underlying etiologies at large. Recent empirical work has corroborated this notion[36]. One striking example here was the near-zero genetic correlation between the autoimmune diseases rheumatoid arthritis (RA) and inflammatory bowel disease (IBD), which are known to share risk loci. The reason was the lack of an overall directional trend: whereas some RA risk alleles were also risk alleles for IBD, other RA risk alleles were protective for IBD, resulting in a near-zero correlation at the genome-wide level.

The results of the GWAS catalog lookup are reported in **Supplementary Table 22**.

First, we note that SWB SNP rs2075677 is in high (R^2 = 0.83) or perfect LD (R^2 = 1) with two SNPs that were highly associated with height in the most recent height GWAS[94,95]. These SNPs lie in the intron of *STAU1*.

In addition, four neuroticism SNPs (or their LD partners) surfaced in the GWAS catalog: rs193236081 overlaps with four variants ($R^2 \geq 0.94$), all located in (the vicinity of) *MAPT* or *KANSL1*. They are associated with Parkinson's disease (two variants), intracranial volume, and progressive supranuclear palsy (a neurodegenerative disorder that is frequently misdiagnosed as Parkinson's or Alzheimer's due to similarities in symptomology[148]).

Interestingly, neuroticism SNP rs4938021 is in high LD (R^2 = 0.87) with a top hit in the most recent schizophrenia GWAS[25]. The SNP in question is located 50kb upstream from *DRD2*, a gene that codes for a type of dopamine receptor. *DRD2* is one of only two genes residing in the locus tagged by rs4938021, making it a plausible biological candidate (the other gene is micro-RNA gene *MIR4301*). Dopamine receptor D2 is in fact a known target of all effective antipsychotic drugs used in the treatment of schizophrenia[149], but also plays an important role in locomotion. Generally, dopamine D2 receptors are known to play a major role in the mesolimbic reward pathway, which is where their link to neuroticism might lie. Importantly, Ripke et al.'s[25] finding highlighted that the small effect sizes of common SNPs identified using the experimental design of GWAS do not necessarily preclude biological relevance.

DS and neuroticism SNP rs6904596 has also been highly associated with schizophrenia and bipolar disorder in several GWAS[25,150–152]. In fact, this SNP lies in the HLA-region: the most strongly associated locus for schizophrenia.

E. Using co-expression to predict gene function

We used a recently developed co-expression database called Gene Network [61] (http://www.genenetworki.nl:8080/GeneNetwork/mgi.html, accessed 15 October 2015) to gain insight into the functions of the genes that are in LD with our lead SNPs. The database can assign functions to a gene even if that gene does not bear explicit annotations referring to those functions, by establishing "guilt-by-association" in terms of co-expression with genes that are members of a functionally defined set taken from standard databases such as Gene Ontology (GO).

In brief, Gene Network uses information from a total of 77,840 Affymetrix human, mouse, or rat microarrays from the Gene Expression Omnibus, which measure expression levels of 19,997 genes in total. Gene co-expression levels were subjected to principal component analysis, which yielded a total of 2,206 reliable "transcriptional components" (TCs). Each TC captures a shared pattern of gene expression across experiments, indicating shared biology of the genes that load highly on it. For each TC and predefined gene set, the difference between the mean TC loading and the mean loading of all genes *not* present in the gene set was subjected to a *t*-test. Roughly speaking, this tests whether the pattern of co-expression captured by the TC discriminates members of the gene set from all other genes. Finally, each row of the resultant 14,461×2,206 matrix of *t*-statistics (each element corresponding to a gene set and TC) was correlated with the TC loadings of the individual genes. This establishes whether the expression pattern of an individual gene is aligned with the overall expression pattern of the gene set. The thresholds for declaring a correlation between a gene's TC loadings and a gene set's *t*-statistics to be statistically significant were chosen to satisfy FDR < 0.05. In depth-documentation of the method can be found in Fehrmann et al. (2015)[61] and Pers et al. (2015)[153].

Now, we describe how we queried the Gene Network database. First, we ascertained the list of genes that are in LD with the 20 SNPs ($R^2 \geq 0.5$). To this end, we used the latest version of SNPsnap[154] (http://www.broadinstitute.org/mpg/snpsnap/, accessed 12 October 2015). SNPsnap uses LD information from the 1000 Genomes phase III CEU reference panel and GENCODE genes downloaded via Ensembl GRCh37 Biomart. This results in a database of 57,734 genes (of which 20,314 are protein-coding). Subsequently, we queried the Gene Network co-expression database with the Ensembl Gene IDs that SNPsnap provides. This ensures that all gene information was captured, even if any gene was entered with an unusual synonym. All genes in the loci tagged by the queried SNPs can be found in **Supplementary Table 34**; in total, 269 genes were present, of which 105 genes were present in Gene Network. We note that the majority of genes not present in the database (N = 142) were non-protein coding (e.g., pseudogenes or lincRNA genes).

We recorded all statistically significant results listed under Gene Ontology[155] (http://amigo.geneontology.org/amigo) biological process, cellular compartment, and molecular function. We also recorded the results listed under the Reactome[156] (http://www.reactome.org/) and Kyoto Encyclopedia of Genes and Genomes[157] (KEGG, http://www.genome.jp/kegg/) pathways. We then recorded all tissues, organs, and cell types where the area under the receiver operating characteristic curve (AUC) with respect to the discriminating power of measured gene expression exceeded 0.80. Gene Network derives the AUC from the difference between the samples of the focal tissue and all other tissues in the distribution of the query gene's expression level, as determined by text-mining the descriptions provided by experimenters who uploaded expression data to the Gene Expression Omnibus (GEO). Note

that the tissue/cell type labels taken from the Medical Subject Headings (MeSH) database can refer to different levels of a hierarchy and are therefore not necessarily mutually exclusive.

We do note that the method has two potential inherent drawbacks. First, only genes with reliable gene expression data (i.e., which have survived quality control) are included in the database. For this reason, *FTO* (an important obesity gene) is for instance not included in the database. In our case, too, not all genes in the loci tagged by our SNPs were present in the database. Therefore, we cannot exclude the possibility that this method misses genes that might be important for our phenotypes, but that have not yielded reliable expression data. Second, the method does not assign function to genes that have unique patterns of gene expression, but that may well play a role in known biological processes. In our case, another substantial drawback is the relative lack of power caused by the modest number of queried SNPs, which prevents the use of DEPICT[153] (a software tool based on the Gene Network correlations between genes and gene sets) to test the formal significance of any putative enrichment. Furthermore, our wide locus definition ($R^2 \geq 0.5$) is likely to include many irrelevant genes. Still, we report the top 10 most frequently occurring terms in our Gene Network lookups, because it is likely that at least some of these terms reflect genuine enrichment.

The results can be found in **Supplementary Table 35**, which lists the ten most frequently occurring search results yielded by each data source and their respective counts. In the results, we find that gene functions and cell components pertaining to gene transcription rank particularly high. As an illustration, the top occurring terms are CHROMATIN ORGANIZATION in GO biological process, CHROMATIN in GO cellular component, ATP-DEPENDENT HELICASE ACTIVITY in GO molecular function, and TRANSCRIPTION in Reactome. These findings are consistent with the view that common genetic variants involved in complex traits are most likely to exert their effects through modification of gene expression. Lastly, we do note that the genes in the loci tagged by our SNPs are mainly specifically expressed in subcortical brain areas.

F. Summary overview of results

We used stratified LD score regression to ascertain the proportion of heritability attributable to various functional SNP categories. The most strongly enriched functional categories for SWB, DS, and neuroticism were evolutionarily conserved regions, the histone marks H3K4me3 and H3K9ac, and coding SNPs. In these respects our three primary phenotypes are similar to those examined by Finucane et al.[49]. The findings of the tissue-type enrichment analyses were perhaps more surprising. Here, we found significant enrichment of not only the CENTRAL NERVOUS SYSTEM (as expected), but also of the ADRENAL/PANCREAS for our SWB and DS phenotypes (for neuroticism, the only significant enriched tissue was CENTRAL NERVOUS SYSTEM). This potentially strengthens the evidence for HPA-axis involvement in SWB and mental health.

Our lookup exercises concentrated on specific SNPs and nearby genes. We found that many of our identified SNPs are associated with the expression of nearby genes or other phenotypes that have been studied in GWAS. It is worth reemphasizing, however, that the overlaps with other phenotypes in the GWAS catalog must be interpreted cautiously, especially in the absence of at least a moderate genetic correlation between the two phenotypes. For instance, in the case of the association between the SWB SNP rs2075677 and height, it may be tempting to take this overlap as evidence of height affecting SWB. Whereas mechanisms such as greater physical height leading to increased social standing might account for such a causal relationship, the reverse causal chain seems inherently implausible. But we must also consider other possibilities, such as the possibility that some of the overlap is simply an artefact of gene co-

localization. Further understanding of all overlap results will require follow-up analyses, including fine-mapping of the relevant loci.

Generally, our most notable finding is the association of two large, well-known inversions on chromosome 8 and chromosome 17 with neuroticism and SWB. Our eQTL analysis of the SNPs located in and around the inversion's breakpoints revealed involvement of several well-annotated genes. One candidate was *FDFT1,* a gene involved in cholesterol synthesis that is ubiquitously expressed in the brain and possibly crucial for the development of the central nervous system. Again, this could hint at a potential role for brain cholesterol in SWB and mental health. Another candidate was *MSRA,* a gene highly expressed in the cerebellum, temporal lobe, motor neurons, and visual cortex. This gene appears to have important effects on protein repair following oxidative stress, and is associated with lifespan in experimental studies in mouse and *Drosophila*. Another candidate gene arising out of several analyses was *MAPT*, a gene involved in neurodegenerative disorders. *MAPT* lies in the large chromosome 17 inversion polymorphism we identified. We note that the inversion also contains corticotropin-releasing hormone receptor 1 (*CRCH1*; a candidate gene for depression), although this gene was not implicated in our functional annotation analyses.

Finally, we note that the protein-coding genes located in the loci tagged by the SNPs identified in our analyses are mainly predicted to be involved in modification of gene expression. Moreover, genes implicated in our lookup analyses were disproportionately expressed in subcortical brain areas, and some are well known for encoding neuronal parts (e.g., *MAPT*, *DRD2*). Other implicated genes were involved in cholesterol synthesis in the brain (*ACSL6, FDFT1*) or protection from cell damage due to oxidative stress in the brain (*MSRA, ADM*). However, some implicated genes also appeared to have distinct immune functions. This does not mean that our phenotypes show significant immune function involvement overall; for instance, there was no evidence for enrichment of associated SNPs for immune tissues and cell types in our LD Score partitioning analyses. Further increases in the GWAS sample sizes for SWB and related phenotypes will improve the resolution of stratified LD Score regression and render feasible other sophisticated enrichment analyses. Finally, we emphasize that it is not possible to draw any conclusions from our results regarding which genes are causally involved in SWB, neuroticism, or depression. For that purpose, future fine-mapping studies are needed to explore the implicated loci (and their role in well-being) in greater depth.

10. Lookup of Top SNPs in Companion Study of Depression

A. Background

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. The phenotype in this companion study is a binary indicator equal to 1 if the respondent had experienced depression at least once. This phenotype was measured from survey questions administered to 23andMe customers and had a prevalence of about 25%. For full details on association models, quality-control filters, and the ascertainment of depression status, we refer to the companion study[158].

We obtained association results from the investigators of the companion study for the 54 DS-associated SNPs in **Supplementary Table 15**, and 117 out of the 118 neuroticism-associated SNPs in **Supplementary Table 15** (one neuroticism SNP, rs117893837, was not available in the summary statistics from the companion study). The standard errors in the association statistics have been adjusted for inflation using the square root of the estimated LD-score intercept[28] ($\sqrt{1.059}$).

Of the 54 DS-associated SNPs, in the 23andMe sample, 40 have the expected sign and ten are associated with DS at 1% level (always with a sign in the anticipated direction). Of the 117 neuroticism-associated SNPs available in the 23andMe sample, 85 have the expected sign, and 16 are significant at the 1% level (again, always with the anticipated signs).

B. Meta-Analysis of 54 DS-associated SNPs

We also meta-analyzed the results of our study and the results of the 23andMe cohort for the 54 available SNPs reaching $P < 10^{-5}$ in our analysis of DS. In this meta-analysis, we weight both cohorts by their effective sample size (assuming a prevalence of 25% in the companion study). The results are shown in **Supplementary Table 15**. Five of the 54 SNPs reach genome-wide significance in the weighted meta-analysis.

11. Additional information

A. Author contributions

Meike Bartels, Daniel Benjamin, David Cesarini, Jan-Emmanuel De Neve, Philipp Koellinger, and Robert Krueger designed and oversaw the study.

The lead analysts responsible for quality control and meta-analyses were Aysu Okbay and Bart Baselmans. Mark Alan Fontana ran the GWAS in UKB for SWB, DS, and neuroticism.

Jonathan Beauchamp and Patrick Turley applied methods to test for population stratification, developed the analysis of the tradeoff between increasing sample size and maintaining phenotype homogeneity and conducted the Bayesian credibility analyses.

Polygenic-prediction analyses were conducted by Aysu Okbay and Bart Baselmans.

The genetic correlation analyses were conducted by Mark Alan Fontana, Patrick Turley, and Jonathan Beauchamp. Richard Karlsson Linnér and Aysu Okbay conducted the proxy-phenotype and cross-phenotype enrichment analyses.

The bioinformatics analyses were conducted by Jonathan Beauchamp (LD Score partitioning, inversion replication), Mark Alan Fontana (LD Score partitioning), Fleur Meddens (Gene Network, HaploReg, and GWAS catalog look-ups), Michel Nivard (eQTL lookups, LD Score partitioning, detection, calling and replicating the inversion), and Tõnu Esko (eQTL lookups, calling and replicating the inversion). Tõnu Esko, James Lee, Fleur Meddens, and Michel Nivard led the writing of the results of the bioinformatics analyses.

Richard Karlsson Linnér and Fleur Meddens prepared the majority of the figures and tables.

Harm-Jan Westra and Juan Ramon Gonzalez provided data and developed the analytical software used to detect the inversion polymorphism and performed eQTL-mapping to understand the functional effects of the inversion polymorphism.

The Advisory Board of the SSGAC (Dalton Conley, George Davey Smith, Tõnu Esko, Albert Hofman, Robert Krueger, David Laibson, Sarah Medland, and Michelle Meyer) supported the project in a variety of important ways. Samantha Cherney provided organizational support for the project and legal counsel.

All authors contributed to and critically reviewed the manuscript. Daniel Benjamin, Meike Bartels, Jonathan Beauchamp, David Cesarini, Jan-Emmanuel De Neve, Michel Nivard, Philipp Koellinger, Aysu Okbay, and Patrick Turley made especially major contributions to the writing and editing.

Cohort	Author	Study Design & Managmt.	Data Collection	Genotyping	Genotype Prep.	Phenotype Prep.	Data Analysis	Wrote MS
1958BC	Alana Cavadino				X	X	X	
1958BC	Christine Power	X	X			X		
1958BC	Elina Hyppönen	X	X			X		
23andMe, Inc.	Nicholas A. Furlotte						X	
23andMe, Inc.	David A. Hinds	X						
AEGS	Gerard Pasterkamp	X						
AEGS	Saskia Haitjema		X		X	X	X	
AEGS	Sander van der Laan		X	X	X		X	
AEGS	Hester den Ruijter	X						
AGES	Albert V. Smith				X	X	X	
AGES	Lenore J. Launer	X				X		
AGES	Tamara B. Harris	X				X		
AGES	Vilmundur Gudnason	X				X		

ALSPAC	Nicholas J. Timpson	X	X	X	X	X	X	
ALSPAC	Evie Stergiakouli			X	X	X	X	
ALSPAC	George Davey Smith	X	X					
ARIC	Jian Yang						X	
ARIC	Andrew Bakshi						X	
ARIC	Anna Vinkhuyzen						X	
ARIC	Cornelius A. Rietveld						X	
ASPS	Marisa Loitfelder		X			X	X	
ASPS	Katja E. Petrovic		X			X	X	
ASPS	Helena Schmidt			X	X			
ASPS	Reinhold Schmidt	X	X					
BASE II	Tian Liu				X		X	
BASE II	Peter Eibich	X	X			X	X	
BASE II	Gert Wagner	X	X			X	X	
BASE II	Lars Bertram	X	X	X	X		X	

BLSA	Luigi Ferrucci		X					
BLSA	Toshiko Tanaka				X		X	
BLSA	Antonio Terracciano		X			X	X	
BLSA	Angelina Sutin					X	X	
CoLaus	Rico Rueedi				X	X	X	
CoLaus	Pedro Marques-Vidal	X				X		
CoLaus	Peter Vollenweider	X	X			X		
CROATIA	Ivana Kolcic		X			X		
CROATIA	Igor Rudan	X	X					
CROATIA	Ozren Polasek	X	X			X		
CROATIA	Caroline Hayward	X		X	X		X	
deCODE	Gudmar Thorleifsson						X	
deCODE	Gyda Bjornsdottir					X		
deCODE	Unnur Thorsteinsdottir	X	X	X	X			
deCODE	Kari Stefansson	X	X					

DHS	Juergen Wellmann				X	X	X	
DHS	Klaus Berger	X	X			X	X	
EGCUT	Evelin Mihailov				X		X	
EGCUT	Anu Realo					X	X	
EGCUT	Andres Metspalu	X	X					
EGCUT	Tõnu Esko	X		X	X			X
ELSA	Meena Kumari	X	X	X				
ELSA	Andrew Steptoe	X	X					
ELSA	Victoria Garfield						X	
ELSA	Ghazaleh Fatemifar				X			
ELSA	Delilah Zabaneh				X			
ERF	Sven J. van der Lee					X	X	
ERF	Najaf Amin				X	X		
ERF	Cornelia M. van Duijn		X	X				
ERF	Sara M. Willems					X	X	

FTC	Jadwiga Buchwald						X	
FTC	Juho Wedenoja						X	
FTC	Anu Loukola			X	X			
FTC	Jaakko Kaprio	X	X			X		
GERA	Joseph Pickrell						X	
GERA	Jimmy Liu						X	
GOYA	Tarunveer S. Ahluwalia					X	X	
GOYA	Lavinia Paternoster			X		X		
GOYA	Torben Hansen	X	X					
GOYA	Thorkild I.A. Sørensen	X	X	X		X		
HBCS	Jari Lahti		X			X	X	
HBCS	Katri Raikkonen	X	X					
HBCS	Eero Kajantie	X	X			X		
HBCS	Johan G. Eriksson	X	X	X				
HCS	Christopher Oldmeadow						X	

HCS	Elizabeth G. Holliday				X		X	
HCS	Rodney J. Scott	X	X	X				
HCS	John R. Attia	X	X					
HNRS	Börge Schmidt				X	X	X	
HNRS	Andreas J. Forstner			X				
HNRS	Marie Henrike Geisel				X	X	X	
HNRS	Karl-Heinz Jöckel	X	X					
HRS	Wei Zhao				X		X	
HRS	Jennifer A. Smith				X	X	X	
HRS	Jessica D. Faul	X	X	X		X		
HRS	Sharon LR. Kardia			X	X			
HRS	David R. Weir	X	X	X				
KORA	Karl-Heinz Ladwig	X	X			X		
KORA	Rajesh Rawal					X		
KORA	Rebecca Emeny	X				X		

KORA	Christian Gieger	X	X	X	X			
LBC	Alison Pattie		X			X		
LBC	Gail Davies			X	X	X	X	
LBC	Simon Cox		X			X		
LBC	Ian J. Deary	X	X	X				
LifeLines	LifeLines cohort study		X					
LifeLines	Lude Franke		X	X	X			
LifeLines	Harold Snieder		X					
LifeLines	Ute Bultmann	X				X		
LifeLines	Behrooz Z. Alizadeh	X			X		X	
MCTFR	James J. Lee							X
MCTFR	Michael B. Miller			X	X		X	
MCTFR	Matt McGue	X	X	X	X	X		
MCTFR	William G. Iacono	X	X	X	X			
MCTFR	Jaime Derringer					X	X	

MCTFR	Lindsay Matteson					X	X	
MESA	Stephen Rich	X	X	X	X			
MESA	Alexis Frazier-Wood				X	X		
MESA	Kent Taylor				X			
MESA	Jennifer A. Smith					X		
MILLS	Melinda Mills		X					
NEO	Dennis O. Mook-Kanamori			X	X	X	X	
NEO	Ruifang Li-Gao			X	X		X	
NEO	Renée de Mutsert	X	X			X		
NEO	Frits R. Rosendaal	X	X					
NFBC1966	Marika A. Kaakinen					X	X	
NFBC1966	Ville Karhunen					X	X	
NFBC1966	Marjo-Riitta Järvelin	X	X					
NFBC1966	Rauli Svento	X						
NHS / HPFS	Laura Kubzansky	X	X			X		

NHS / HPFS	Shun-Chiao Chang	X	X			X	X	
NHS / HPFS	Peter Kraft	X	X	X	X			
NTR	Meike Bartels	X	X			X	X	X
NTR	Bart Baselmans						X	
NTR	Eco de Geus		X					
NTR	Jouke-Jan Hottenga			X	X			
NTR	Dorret I. Boomsma	X	X					
NTR	Michel Nivard						X	
NTR	Camelia Minica				X		X	
QIMR	Grant W. Montgomery	X		X				
QIMR	Penelope Lind				X	X	X	
QIMR	Andrew C. Heath	X	X	X				
QIMR	Sarah E. Medland		X	X	X			
QIMR	Nicholas G. Martin	X	X	X				
QIMR	Miriam A. Mosing					X		

RS	Cornelius A. Rietveld				X	X	X	X
RS	Nese Direk		X				X	
RS	Patrick J.F. Groenen	X						
RS	Albert Hofman	X	X					
RS	A. Roy Thurik	X						
RS	Henning Tiemeier	X	X					
RS	André G. Uitterlinden	X	X	X	X			
MAP ROS	Lei Yu		X			X	X	
MAP ROS	Patricia A. Boyle		X			X		
MAP ROS	Philip L. DeJager		X		X			
MAP ROS	David A. Bennett	X	X					
SardiNIA	Yong Qian						X	
SardiNIA	Jun Ding						X	
SardiNIA	Antonio Terracciano	X	X			X	X	
SardiNIA	Francesco Cucca	X	X	X	X			

SardiNIA	David Schlessinger	X	X	X				
STR	Robert Karlsson		X	X	X			
STR	Nancy L. Pedersen	X	X			X		
STR	Patrik K.E. Magnusson	X	X			X		
STR	Magnus Johannesson	X	X			X		
STR	David Cesarini	X	X	X	X	X	X	X
STR	Cornelius A. Rietveld				X	X	X	
TEDS	Claire Haworth	X	X			X	X	
TEDS	Oliver Davis			X	X	X	X	
TEDS	Maciej Trzaskowski						X	
TEDS	Robert Plomin	X	X					
TRAILS	Albertine Oldehinkel	X	X			X		
TRAILS	Peter Van der Most				X		X	
TRAILS	Ilja Nolte			X	X		X	
TRAILS	Catharina Hartman	X	X	X				

TwinsUK	Lydia Quaye					X	X	
TwinsUK	Juliette Harris		X					
TwinsUK	Tim D. Spector	X						
YFS	Olli Raitakari	X	X	X		X		
YFS	Laura Pulkki-Raback	X	X			X		
YFS	Terho Lehtimäki	X	X	X		X		
YFS	Liisa Keltikangas-Järvinen	X						

B. Additional acknowledgments

1958BC (1958 Birth Cohort) – DNA collection was funded by MRC grant G0000934 and cell-line creation by Wellcome Trust grant 068545/Z/02. This research used resources provided by the Type 1 Diabetes Genetics, a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases, National Human Genome Research Institute, National Institute of Child Health and Human Development, and Juvenile Diabetes Research Foundation International (JDRF) and supported by U01 DK062418. This study makes use of data generated by the Wellcome Trust Case-Control Consortium. A full list of investigators who contributed to generation of the data is available from the Wellcome Trust Case-Control Consortium website. Funding for the project was provided by the Wellcome Trust under the award 076113. Great Ormond Street Hospital/University College London, Institute of Child Health receives a proportion of funding from the Department of Health's National Institute for Health Research (NIHR) ('Biomedical Research Centres' funding). The work was supported by the Department of Health Policy Research Programme through the Public Health Research Consortium (PHRC). The views expressed in the publication are those of the authors and not necessarily those of the Department of Health. Information about the wider programme of the PHRC is available from http://phrc.lshtm.ac.uk. Written consent was obtained from participants for the use of information in medical studies. The 45-year biomedical survey and genetic studies were approved by the South-East Multi-Centre Research Ethics Committee (ref: 01/1/44) and the joint UCL/UCLH Committees on the Ethics of Human Research (Ref: 08/H0714/40). The individual-level genotype and phenotype data for WTCCC-58BC and DIL (T1DGC) can be applied for through the data access committee at http://www2.le.ac.uk/projects/birthcohort/1958bc/available-resources.

23andMe, Inc. – 23andMe research participants provided informed consent to take part in this research under a protocol approved by the AAHRPP-accredited institutional review board, Ethical and Independent Review Services. We would like to thank the research participants and employees of 23andMe for making this work possible. This work was supported by the National Human Genome Research Institute of the National Institutes of Health (grant number R44HG006981).

AEGS (Athero-Express Genomics Study) – Genotyping of Athero-Express Genomics Study 1 was funded by Cavadis B.V. Sander W. van der Laan is funded through grants from the Interuniversity Cardiology Institute of the Netherlands (ICIN, 09.001) and CVON (GENIUS, CVON2011-19). Saskia Haitjema and Sander W. van der Laan are funded through the FP7 EU project CVgenes@target (HEALTH-F2-2013-601456). Claudia Tersteeg, Krista den Ouden, Mirjam B. Smeets, and Loes B. Collé are graciously acknowledged for their work on the DNA

extraction. Astrid E.M.W. Willems, Evelyn Velema, Kristy M. J. Vons, Sara Bregman, Timo R. ten Brinke, Sara van Laar, Louise M. Catanzariti, Joyce E.P. Vrijenhoek, Sander M. van de Weg, Arjan H. Schoneveld, Petra H. Homoed-van der Kraak, and Aryan Vink are graciously acknowledged for their past and continuing work on the Athero-Express Biobank Study. We would also like to thank all the (former) employees involved in the Athero-Express Biobank Study of the Departments of Surgery of the St. Antonius Hospital Nieuwegein and University Medical Center Utrecht for their continuing work. Jessica van Setten is graciously acknowledged for her help in the quality assurance and quality control of the genotype data. We graciously thank the team of Golden Helix Inc. especially Autumn Laughbaum, Bryce Christensen, Christophe Lambert, Greta Linse Peterson, and Gabe Rudy for their continuing support in data analysis. Lastly, we would like to thank all participants of the Athero-Express Biobank Study; without you these kinds of studies would not be possible. For further information please go to www.atheroexpress.nl.

AGES (Age, Gene/Environment Susceptibility-Reykjavik Study) – AGES is funded by NIH contract N01-AG-12100, the NIA Intramural Research Program, Hjartavernd (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament). Genotyping was conducted at the NIA IRP Laboratory of Neurogenetics. Researchers interested in using the AGES data must obtain approval from the AGES study group. Researchers using the data are required to follow the terms of a research agreement between them and the AGES investigators. In accordance with Icelandic law, individual level data cannot be released to external investigators, only summary GWAS results. Investigators interested in collaboration can work on individual data at the Icelandic Heart Association site. For further information contact Prof. V. Gudnason (v.gudnason@hjarta.is).

ALSPAC (Avon Longitudinal Study of Parents and Children) – The MRC IEU is supported by the Medical Research Council and the University of Bristol (grant code MC_UU_12013/1-9). The authors are extremely grateful to all the families who took part in the ALSPAC study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses. The UK Medical Research Council and the Wellcome Trust (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. ALSPAC GWAS data was generated by Sample Logistics and Genotyping Facilities at the Wellcome Trust Sanger Institute and LabCorp (Laboratory Corporation of America) supported by 23andMe. ALSPAC summary data will be published on the data repository at www.data.bris.ac.uk. Please note that the study website contains details of all the data that is available through a fully searchable data dictionary. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees.

ARIC (Atherosclerosis Risk in Communities Study) – ARIC data was obtained by way of dbGaP. ARIC specific acknowledgments and more information for researchers interested in obtaining data are available at https://www2.cscc.unc.edu/aric/

ASPS (Austrian Stroke Prevention Study) – We thank Ing. Johann Semmler and Irmgard Pölzl for creating the DNA bank and for supervising the quality management of the biobanking and DNA analyses. The ASPS is funded by the Austrian Science Fond (FWF) grant number P20545-P05 and P13180. The data file of ASPS regarding all analyses of this manuscript can be provided to external researchers upon request.

BASE-II (The Berlin Aging Study II) – This work is supported by the German Federal Ministry of education and Research (BMBF [grants #16SV5536K, #16SV5537, #16SV5538, and #16SV5837; previously #01UW0808]), the Max Planck Society, the Leibniz Society, and a Gottfried Wilhelm Leibniz Award of the German Research Foundation to Co-PI Ulman Lindenberger. External scientists can apply to the Steering Committee of BASE-II for data access. All existing variables are documented in a handbook. Contact: Katrin Schaar, scientific coordinator, schaar@mpib-berlin.mpg.de.

BLSA (Baltimore Longitudinal Study of Aging) – We acknowledge support from the Intramural Research Program of the NIH, National Institute on Aging. More information for researchers interested in BLSA is available at www.blsa.nih.gov.

CoLaus (Cohorte Lausannoise) – The CoLaus study was and is supported by research grants from GlaxoSmithKline, the Faculty of Biology and Medicine of Lausanne, and the Swiss National Science Foundation (grants 33CSCO-122661, 33CS30-139468 and 33CS30-148401). Researchers must obtain approval from the Steering Committee of the CoLaus Study and from the Institutional Ethics Committee of the University in Lausanne, Switzerland. Researchers using the data are required to follow the terms of an Cooperation Agreement containing a number of clauses designed to ensure protection of privacy and compliance with relevant laws. For further information go to www.colaus.ch. All data requests should be sent to the PI of the CoLaus study.

CROATIA (Croatia Korcula – Split – Vis studies) – The Croatian cohorts were supported by the "Croatian Science Foundation, grant 8875". The CROATIA-Vis study was funded by grants from the Medical Research Council (UK) and Republic of Ministry of Science, Education and Sports research grants to I.R. (108-1080315-0302). We would like to acknowledge the staff of several institutions in Croatia that supported the field work, including but not limited to The University of Split and Zagreb Medical Schools, the Institute for Anthropological Research in Zagreb and Croatian Institute for Public Health. The SNP genotyping for the CROATIA-Vis cohort was performed in the core genotyping laboratory of the Wellcome Trust Clinical Research Facility at

the Western General Hospital, Edinburgh, Scotland. We would like to acknowledge the staff of several institutions in Croatia that supported the field work, including but not limited to The University of Split and Zagreb Medical Schools, the Institute for Anthropological Research in Zagreb and Croatian Institute for Public Health.

The CROATIA-Split study was funded by grants from the Medical Research Council (UK), European Commission Framework 6 project EUROSPAN (Contract No. LSHG-CT-2006-018947) and Republic of Croatia Ministry of Science, Education and Sports research grants to I.R. (108-1080315-0302). We would like to acknowledge the staff of several institutions in Croatia that supported the field work, including but not limited to The University of Split and Zagreb Medical Schools and the Croatian Institute for Public Health. The SNP genotyping for the CROATIA-Split cohort was performed by AROS Applied Biotechnology, Aarhus, Denmark.

The CROATIA-Korcula study was funded by grants from the Medical Research Council (UK), European Commission Framework 6 project EUROSPAN (Contract No. LSHG-CT-2006-018947), European Commission Framework 7 project BBMRI-LPC (Contract No. 313010), Croatian Science Council (Grant no. 8875), and Republic of Croatia Ministry of Science, Education and Sports research grants to I.R. (108-1080315-0302). We would like to acknowledge the invaluable contributions of the recruitment team in Korcula, the administrative teams in Croatia and Edinburgh and the people of Korcula. The SNP genotyping for the CROATIA-Korcula cohort was performed in Helmholtz Zentrum München, Neuherberg, Germany. External researchers who wish to obtain access to CROATIA-Korcula's data may contact Ozren Polasek, ozren.polasek@mefst.hr.

deCODE (deCODE Genetics) – All deCODE collaborators in this study are employees of deCODE Genetics/Amgen, Inc. External researchers who wish to obtain access to data or EA2 results may contact Gudmar Thorleifsson gudmar.thorleifsson@decode.is.

DHS (Dortmund Health Study) – DHS (Dortmund Health Study) - The collection of sociodemographic and clinical data in the Dortmund Health Study was supported by the German Migraine & Headache Society (DMKG) and by unrestricted grants of equal share from Almirall, Astra Zeneca, Berlin Chemie, Boehringer, Boots Health Care, Glaxo-Smith-Kline, Janssen Cilag, McNeil Pharma, MSD Sharp & Dohme and Pfizer to the University of Muenster. Blood collection in the Dortmund Health Study was done through funds from the Institute of Epidemiology and Social Medicine University of Muenster. Genotyping for the Human Omni Chip was supported by the German Ministry of Education and Research (BMBF, grant no. 01ER0816). Researchers interested in using DHS data are required to sign and follow the terms of a Cooperation Agreement that includes a number of clauses designed to ensure protection of privacy and compliance with relevant laws. For further information, contact Klaus Berger (bergerk@uni-muenster.de).

EGCUT (Estonian Genome Center) – EGCUT received targeted financing from Estonian Research Council grant IUT20-60, Center of Excellence in Genomics (EXCEGEN) and University of Tartu (SP1GVARENG). We acknowledge EGCUT technical personnel, especially Mr V. Soo and S. Smit. Data analyzes were carried out in part in the High Performance Computing Center of University of Tartu. For more information, please contact Tonu Esko (tesko@broadinstitute.org).

ELSA (English Longitudinal Studies of Ageing) – Samples from the ELSA DNA Repository (EDNAR), received support under a grant (AG1764406S1) awarded by the National Institute on Ageing (NIA). Genotyping was supported under grants from the Economic and Social Research Council (ES/K005774/1) and NIA (AG017644-13S1). ELSA was developed by a team of researchers based at University College London, the Institute of Fiscal Studies and the National Centre for Social Research. The data were collected by the National Centre for Social Research. ELSA data are made available through the ESDS website (http://www.elsa-project.ac.uk/availableData).

ERF (Erasmus Rucphen Family Study) – Erasmus Rucphen family study: The ERF study as a part of EUROSPAN (European Special Populations Research Network) was supported by European Commission FP6 STRP grant number 018947 (LSHG-CT-2006-01947) and also received funding from the European Community's Seventh Framework Programme (FP7/2007-2013)/grant agreement HEALTH-F4-2007-201413 by the European Commission under the programme "Quality of Life and Management of the Living Resources" of 5th Framework Programme (no. QLG2-CT-2002-01254). High-throughput analysis of the ERF data was supported by a joint grant from the Netherlands Organization for Scientific Research and the Russian Foundation for Basic Research (NWO-RFBR 047.017.043). Exome sequencing analysis in ERF was supported by the ZonMw grant (project 91111025). We are grateful to all study participants and their relatives, general practitioners and neurologists for their contributions and to P. Veraart for her help in genealogy, J. Vergeer for the supervision of the laboratory work and P. Snijders for his help in data collection. Najaf Amin is supported by the Netherlands Brain Foundation (project number F2013(1)-28). Femke M.S. de Vrij was supported by the Netherlands Brain Foundation (grant number FS2012(1)-39). The ERF study genome-wide array data and phenotype data (age and gender) is archived in European Genome-Phenome Database (EGA). Researchers who wish to use other phenotypic data of the Erasmus Rucphen Family Study must seek approval from the management team of the Erasmus Rucphen Family study. They are advised to contact the study PI, professor Cornelia van Duijn (c.vanduijn@erasmusmc.nl).

FTC (Finnish Twin Registry) – We warmly thank the participating twin pairs and their family members for their contribution. Anja Häppölä and Kauko Heikkilä are acknowledged for their valuable contribution in recruitment, data collection, and data management. Phenotyping and genotyping of the Finnish twin cohorts has been supported by the Academy of Finland Center

of Excellence in Complex Disease Genetics (grants 213506, 129680), the Academy of Finland (grants 100499, 205585, 118555, 141054, 265240, 263278 and 264146 to J. Kaprio), National Institute of Alcohol Abuse and Alcoholism (grants AA-12502, AA-00145, and AA-09203 to R. J. Rose and AA15416 and K02AA018755 to D. M. Dick), Sigrid Juselius Foundation (to J. Kaprio), Global Research Award for Nicotine Dependence, Pfizer Inc. (to J. Kaprio), and the Wellcome Trust Sanger Institute, UK. Antti-Pekka Sarin and Samuli Ripatti are acknowledged for genotype data quality controls and imputation. GWAS analyses were run at the ELIXIR Finland node hosted at CSC – IT Center for Science for ICT resources. Researchers interested in using FTC data must obtain approval from the Data Access Committee of the Institute for Molecular Medicine Finland FIMM (fimm-dac@helsinki.fi). Note that anonymized individual level data can only be released after study has been approved by Research Ethics Committee of University of Helsinki and must be carried out in collaboration with FTC investigators. To ensure protection of privacy and compliance with national data protection legislation, a data use/transfer agreement is needed, the content and specific clauses of which will depend on the nature of the requested data. For further information please contact Jaakko Kaprio (jaakko.kaprio@helsinki.fi).

GERA (Resource for Genetic Epidemiology Research on Adult Health and Aging) - The Resource for Genetic Epidemiology Research on Aging (GERA) Cohort was created by a RC2 "Grand Opportunity" grant that was awarded to the Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH) and the UCSF Institute for Human Genetics (AG036607; Schaefer/Risch, PIs). Access to the GERA cohort is available from dbGaP accession number phs000674.v1.p1. More information is available here: http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000674.v1.p1

GOYA (Genetics of Overweight Young Adults) - This study was conducted as part of the activities of the Gene-diet Interactions in Obesity project (GENDINOB, www.gendinob.dk) and the MRC centre for Causal Analyses in Translational Epidemiology (MRC CAiTE). We thank all the participants of the study. TSA was also funded by the GENDINOB project and acknowledges the same. External researchers who wish to obtain access to GOYA (male) data or EA2 results may contact Tarunveer Singh Ahluwalia (tarun.ahluwalia@dbac.dk or veertarun@gmail.com).

HBCS (Helsinki Birth Cohort Study) – We thank all study participants as well as everybody involved in the Helsinki Birth Cohort Study. Helsinki Birth Cohort Study has been supported by grants from the Academy of Finland, the Finnish Diabetes Research Society, Folkhälsan Research Foundation, Novo Nordisk Foundation, Finska Läkaresällskapet, Signe and Ane Gyllenberg Foundation, University of Helsinki, Ministry of Education, Ahokas Foundation, Emil Aaltonen Foundation. Researchers interested in using HBCS data must obtain approval from the Steering Committee of the Helsinki Birth Cohort Study. Researchers using the data are required to follow the terms in a number of clauses designed to ensure protection of privacy and

compliance with relevant Finnish laws. For further information, contact Johan Eriksson (johan.eriksson@helsinki.fi).

HCS (Hunter Community Study) – The authors would like to thank the men and women participating in the HCS as well as all the staff, investigators and collaborators who have supported or been involved in the project to date. The University of Newcastle provided $300 000 from its Strategic Initiatives Fund, and $600 000 from the Gladys M Brawn Senior Research Fellowship scheme; the Vincent Fairfax Family Foundation, a private philanthropic trust, provided $195 000; The Hunter Medical Research Institute provided media support during the initial recruitment of participants; and Dr Anne Crotty, Prof. Rodney Scott and Associate Prof. Levi provided financial support towards freezing costs for the long-term storage of participant blood samples. External researchers may request data access from Chris Oldmeadow, Elizabeth Holliday or John Attia by email.

HNRS (Heinz Nixdorf Recall Study) – The Heinz Nixdorf Recall Study thank the Heinz Nixdorf Foundation (Germany), the German Federal Ministry of Research and Education and projects SI 236/8-1 and SI 236/9-1 from the German Research Council for the generous support of this study. We acknowledge the support of the Sarstedt AG & Co. (Nümbrecht, Germany) for laboratory equipment. The genotyping was partially supported by the German Federal Ministry of Education and Research (BMBF) through the Integrated Network IntegraMent (Integrated Understanding of Causes and Mechanisms in Mental Disorders), under the auspices of the e:Med Programme (grant 01ZX1314A). We are indebted to all study participants and to the dedicated personnel of the study centre of the Heinz Nixdorf Recall study. Advisory Board: Meinertz T, Hamburg, Germany (Chair); Bode C, Freiburg, Germany; de Feyter PJ, Rotterdam, Netherlands; Güntert B, Hall i.T., Austria; Gutzwiller F, Bern, Switzerland; Heinen H, Bonn, Germany; Hess O, Bern, Switzerland; Klein B, Essen, Germany; Löwel H, Neuherberg, Germany; Reiser M, Munich, Germany; Schwaiger M, Munich, Germany; Steinmüller C, Bonn, Germany; Theorell T, Stockholm, Sweden; Willich SN, Berlin, Germany. External researchers can send their inquiry to boerge.schmidt@uk-essen.de to get access to the study results.

HRS (Health and Retirement Study) – HRS is supported by the National Institute on Aging (NIA U01AG009740). The genotyping was funded separately by the National Institute on Aging (RC2 AG036495, RC4 AG039029). Our genotyping was conducted by the NIH Center for Inherited Disease Research (CIDR) at Johns Hopkins University. Genotyping quality control and final preparation of the data were performed by the Genetics Coordinating Center at the University of Washington. Genotype data can be accessed via the database of Genotypes and Phenotypes (dbGaP, http://www.ncbi.nlm.nih.gov/gap, accession number phs000428.v1.p1). Researchers who wish to link genetic data with other HRS measures that are not in dbGaP, such as

educational attainment, must apply for access from HRS. See the HRS website (http://hrsonline.isr.umich.edu/gwas) for details.

The KORA research platform (KORA, Cooperative Research in the Region of Augsburg) – KORA was initiated and financed by the Helmholtz Zentrum München - German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ. External researchers can apply for KORA data via our KORA.PASST project application self-service tool at https://helmholtz-muenchen.managed-otrs.com/otrs/customer.pl.

LBC (Lothian Birth Cohort) –We thank the cohort participants and team members who contributed to these studies. Phenotype collection in the Lothian Birth Cohort 1921 was supported by the UK's Biotechnology and Biological Sciences Research Council (BBSRC), The Royal Society, and The Chief Scientist Office of the Scottish Government. Phenotype collection in the Lothian Birth Cohort 1936 was supported by Age UK (The Disconnected Mind project). Genotyping of the cohorts was funded by the BBSRC. The work was undertaken by The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative (MR/K026992/1). Funding from the BBSRC and Medical Research Council (MRC) is gratefully acknowledged. The raw data collected in the course of our research with human participants is available upon request (please contact, by e-mail, Professor Ian Deary, at the University of Edinburgh, and ask for a 'Lothian Birth Cohort Data Request Form'). The process is facilitated by a full-time Lothian Birth Cohort database manager. Such proposals, when approved, are conducted in collaboration with appropriate members of the Lothian Birth Cohort study team.

LifeLines (LifeLines) – Expanded Banner or Group Author: Behrooz Z Alizadeh (1), Rudolf A de Boer (2), H Marike Boezen (1), Marcel Bruinenberg (3), Lude Franke (4), Pim van der Harst (2), Hans L Hillege (1,2), Melanie M van der Klauw (5), Gerjan Navis (6), Johan Ormel (7), Dirkje S Postma (8), Judith GM Rosmalen (7), Joris P Slaets (9), Harold Snieder (1), Ronald P Stolk (1), Bruce HR Wolffenbuttel (5), Cisca Wijmenga (4).

(1) Department of Epidemiology, University of Groningen, University Medical Center Groningen, The Netherlands

(2) Department of Cardiology, University of Groningen, University Medical Center Groningen, The Netherlands

(3) LifeLines Cohort Study, University of Groningen, University Medical Center Groningen, The Netherlands

(4) Department of Genetics, University of Groningen, University Medical Center Groningen, The Netherlands

(5) Department of Endocrinology, University of Groningen, University Medical Center Groningen, The Netherlands

(6) Department of Internal Medicine, Division of Nephrology, University of Groningen, University Medical Center Groningen, The Netherlands

(7) Interdisciplinary Center of Psychopathology of Emotion Regulation (ICPE), Department of Psychiatry, University of Groningen, University Medical Center Groningen, The Netherlands

(8) Department of Pulmonology, University of Groningen, University Medical Center Groningen, The Netherlands

(9) University Center for Geriatric Medicine, University of Groningen, University Medical Center Groningen, The Netherlands

The LifeLines Cohort Study, and generation and management of GWAS genotype data for the LifeLines Cohort Study is supported by the Netherlands Organization of Scientific Research NWO (grant 175.010.2007.006), the Economic Structure Enhancing Fund (FES) of the Dutch government, the Ministry of Economic Affairs, the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the Northern Netherlands Collaboration of Provinces (SNN), the Province of Groningen, University Medical Center Groningen, the University of Groningen, Dutch Kidney Foundation and Dutch Diabetes Research Foundation. We thank Behrooz Z. Alizadeh, Annemieke Boesjes, Marcel Bruinenberg, Noortje Festen, Pim van der Harst, Ilja Nolte, Lude Franke, Mitra Valimohammadi for their help in creating the GWAS database, and Rob Bieringa, Joost Keers, René Oostergo, Rosalie Visser, Judith Vonk for their work related to data-collection and validation. The authors are grateful to the study participants, the staff from the LifeLines Cohort Study and the contributing research centers delivering data to LifeLines and the participating general practitioners and pharmacists. All data and samples collected by LifeLines are available to scientific researchers worldwide. It is also possible to prospectively collect additional data and samples in a selected group of LifeLines participants in an add-in study. Researchers can apply for data, samples or an add-on study by filling in the application form for research and submitting the completed form through our data catalogue, together with a selection of the requested data. Please contact LifeLines Research office by emailing to LLscience@umcg.nl or calling to +31 50 3615803; when you may need more specific information.

MCTFR (The Minnesota Center for Twin and Family Research) – MCTFR was supported in part by USPHS Grants from the National Institute on Alcohol Abuse and Alcoholism (AA09367 and AA11886), the National Institute on Drug Abuse (DA05147, DA13240, and DA024417), and the

National Institute on Mental Health (MH066140). William Iacono was supported in part by a grant from the National Institute on Drug Abuse (DA 036216). GWAS and phenotypic data for MCTFR subjects who provided consent to place their data in a public repository are deposited into the database of Genotypes and Phenotypes (dbGaP, www.ncbi.nlm.nih.gov/gap) under phs000620. For further information, please contact Matt McGue (mcgue001@umn.edu).

MESA (Multi-Ethnic Study of Atherosclerosis) – The authors thank the participants of the MESA study, the Coordinating Center, MESA investigators, and study staff for their valuable contributions. A full list of participating MESA investigators and institutions can be found at http://www.mesa-nhlbi.org. MESA and the MESA SHARe project are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-001079, UL1-TR-000040, and DK063491. Funding for SHARe genotyping was provided by NHLBI Contract N02-HL-64278. Genotyping was performed at Affymetrix (Santa Clara, California, USA) and the Broad Institute of Harvard and MIT (Boston, Massachusetts, USA) using the Affymetrix Genome-Wide Human SNP Array 6.0. This study was funded by an American Heart Association grant to Frazier-Wood (grant no: 14BGIA18740011), who is also supported by USDA/ARS (CRIS 309-5-001-058). This work is a publication of the USDA/ARS Children's Nutrition Research Center, Department of Pediatrics, Baylor College of Medicine, Houston, Texas and funded in part with federal funds from the USDA/ARS under Cooperative Agreement No. 58-6250-0-008. The contents of this publication do not necessarily reflect the views or policies of the USDA, nor does mention of trade names, commercial products, or organizations imply endorsement from the US government. More information on MESA is available here: http://www.mesa-nhlbi.org

NEO (The Netherlands Epidemiology of Obesity Study) - The authors of the NEO study thank all individuals who participated in the Netherlands Epidemiology in Obesity study, all participating general practitioners for inviting eligible participants and all research nurses for collection of the data. We thank the NEO study group, Petra Noordijk, Pat van Beelen and Ingeborg de Jonge for the coordination, lab and data management of the NEO study. The genotyping in the NEO study was supported by the Centre National de Génotypage (Paris, France), headed by Jean-Francois Deleuze. The NEO study is supported by the participating Departments, the Division and the Board of Directors of the Leiden University Medical Center, and by the Leiden University, Research Profile Area Vascular and Regenerative Medicine. Dennis Mook-Kanamori is supported by Dutch Science Organization (ZonMW-VENI Grant 916.14.023). More information on NEO is available here: http://www.einthovenlaboratory.com/onderzoeken/neo-study/

NFBC1966 (Northern Finland Birth Cohorts (1966 Cohort)) – We thank Professor Paula Rantakallio (launch of NFBC1966 and initial data collection), Ms Sarianna Vaara (data collection), Ms Tuula Ylitalo (administration), Mr Markku Koiranen (data management), Ms Outi Tornwall and Ms Minttu Jussila (DNA biobanking). This work was supported by the Academy of Finland [project grants 104781, 120315, 129418, Center of Excellence in Complex Disease Genetics and Public Health Challenges Research Program (SALVE)], University Hospital Oulu, Biocenter, University of Oulu, Finland (75617), the European Commission [EURO-BLCS, Framework 5 award QLG1-CT-2000-01643], The National Heart, Lung and Blood Institute [5R01HL087679-02] through the SNP Typing for Association with Multiple Phenotypes from Existing Epidemiologic Data (STAMPEED) program [1RL1MH083268-01], The National Institute of Health/The National Institute of Mental Health [5R01MH63706:02], European Network of Genomic and Genetic Epidemiology (ENGAGE) project and grant agreement [HEALTH-F4-2007-201413], and the Medical Research Council, UK [G0500539, G0600705, PrevMetSyn/ Public Health Challenges Research Program (SALVE)]. Researchers interested in using NFBC1966 data must obtain approval from the Ethical Committee of Northern Ostrobothnia Hospital District and from the Data and Publication Committee of the Northern Finland Birth Cohorts. Researchers using the data are required to follow The Declaration of Helsinki and rules of practice containing a number of clauses designed to ensure protection of privacy and compliance with relevant laws. For further information, contact Marjo-Riitta Jarvelin (m.jarvelin@imperial.ac.uk).

NHS / HPFS (Nurses Health Study / Health Professional Follow-up Study) – The Nurses' Health Study is supported by research grants UM1CA186107, R01CA49449, U01HG004399, R01CA40356, U01CA98233, and 5P01DK070756.

The Health Professionals Follow-up Study is supported by research grants UM1CA167552, R01HL35464, 5P01DK070756, and U01HG004399.

The authors thank the participants and staff of the Nurses' Health Study (NHS) and the Health Professionals Follow-up Study (HPFS) for their valuable contributions, and the Channing Division of Network Medicine of Brigham and Women's Hospital and Harvard Medical School, Boston, for its supervision of the NHS and HPFS. For more information and possible collaboration please see http://www.channing.harvard.edu/nhs/ and https://www.hsph.harvard.edu/hpfs/

NTR (Netherlands Twin Register) - We would like to thank all the twins and family members for their participation. This work was supported by the Netherlands Organization for Scientific Research (NWO: MagW/ZonMW grants 904-61-090, 985-10-002,904-61-193,480-04-004, 400-05-717, NOW-bilareral agreement 463-06-001, NWO-VENI 451-04-034, Addiction-31160008 Middelgroot-911-09-032, Spinozapremie 56-464-14192), Biobanking and Biomolecular Resources Research Infrastructure (BBMRI –NL, 184.021.007), the VU University's Institute for

Health and Care Research (EMGO+) and Neuroscience Campus Amsterdam (NCA), the European Science Council (ERC Advanced, 230374), the Avera Institute for Human Genetics, Sioux Falls, South Dakota (USA) and the National Institutes of Health (NIH, R01D0042157-01A). Part of the genotyping was funded by the Genetic Association Information Network (GAIN) of the Foundation for the US National Institutes of Health (NIMH, MH081802) and by the Grand Opportunity grants 1RC2MH089951-01 and 1RC2 MH089995-01 from the NIMH. Part of the analyses were carried out on the Genetic Cluster Computer (http://www.geneticcluster.org), which is financially supported by the Netherlands Scientific Organization (NWO 480-05-003), the Dutch Brain Foundation, and the department of Psychology and Education of the VU University Amsterdam. M.Bartels is/was financially supported by a senior fellowship of the (EMGO+) Institute for Health and Care and a VU University Research Chair position. For access to NTR results, please contact Meike Bartels at m.bartels@vu.nl.

QIMR (Queensland Institute of Medical Research) – Funding was provided by the Australian National Health and Medical Research Council (241944, 339462, 389927, 389875, 389891, 389892, 389938, 442915, 442981, 496739, 552485, 552498), the Australian Research Council (A7960034, A79906588, A79801419, DP0770096, DP0212016, DP0343921), the FP-5 GenomEUtwin Project (QLG2-CT-2002-01254), and the U.S. National Institutes of Health (NIH grants AA07535, AA10248, AA13320, AA13321, AA13326, AA14041, DA12854, MH66206). A portion of the genotyping on which the QIMR study was based (Illumina 370K scans) was carried out at the Center for Inherited Disease Research, Baltimore (CIDR), through an access award to the authors' late colleague Dr. Richard Todd (Psychiatry, Washington University School of Medicine, St Louis). Imputation was carried out on the Genetic Cluster Computer, which is financially supported by the Netherlands Scientific Organization (NWO 480-05-003). S.E.M., is supported by the Australian Research Council (ARC) Fellowship Scheme. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Researchers interested in using QIMR data can contact Nick Martin (Nick.Martin@qimrberghofer.edu.au).

RS (Rotterdam Study) – The generation and management of GWAS genotype data for the Rotterdam Study is supported by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012). This study is funded by the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) project nr. 050-060-810. We thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Lizbeth Herrera and Marjolein Peters for their help in creating the GWAS database, and Karol Estrada and Maksim V. Struchalin for their support in creation and analysis of imputed data. The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE),

the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists. Researchers who wish to use data of the Rotterdam Study must obtain approval from the Rotterdam Study Management Team. They are advised to contact the PI of the Rotterdam Study, Dr Albert Hofman (a.hofman@erasmusmc.nl).

Rush MAP (Memory and Aging Project) – The data used in this analysis is supported by NIA Grants R01AG15819, R01AG17917, R01AG33678, the Illinois Department of Public Health and the Translational Genomics Research Institute. We thank the study participants of the Rush Memory and Aging Project and the staff of the Rush Alzheimer's Disease Center. To obtain data from the Rush Alzheimer's Disease Center (RADC), please submit a request through the RADC research website: https://www.radc.rush.edu/res/ext/home.htm.

SardiNIa (SardinNIA Study of Aging) - The SardiNIA study thanks the many individuals who generously participated in this study, the Mayors and citizens of the Sardinian towns involved, the head of the Public Health Unit ASL4, and the province of Ogliastra for their volunteerism and cooperation. In addition, we are grateful to the Mayor and the administration in Lanusei for providing and furnishing the clinic site. This work was supported by the Intramural Research Program of the National Institute on Aging (NIA), National Institutes of Health (NIH), with contracts NO1-AG-1–2109 and HHSN271201100005C. For more information, contact Francesca Cucca (francesco.cucca@irgb.cnr.it) or David Schlessinger (SchlessingerD@grc.nia.nih.gov).

STR (Swedish Twin Registry) – The Jan Wallander and Tom Hedelius Foundation (P2012-0002:1), the Ragnar Söderberg Foundation (E9/11), The Swedish Research Council (421-2013-1061), the Ministry for Higher Education, The Swedish Research Council (M-2205-1112), GenomEUtwin (EU/QLRT-2001-01254; QLG2-CT-2002-01254), NIH DK U01-066134, The Swedish Foundation for Strategic Research (SSF). Researchers interested in using STR data must obtain approval from the Swedish Ethical Review Board and from the Steering Committee of the Swedish Twin Registry. Researchers using the data are required to follow the terms of an Assistance Agreement containing a number of clauses designed to ensure protection of privacy and compliance with relevant laws. For Further information, contact Patrik Magnusson (Patrik.magnusson@ki.se).

TEDS (Twins Early Development Study) – We gratefully acknowledge the ongoing contribution of the participants in the Twins Early Development Study (TEDS) and their families. TEDS is supported by a program grant to Robert Plomin from the UK Medical Research Council (G0901245; and previously G0500079). Funding for the genotyping for this study was provided by the Wellcome Trust Case Control Consortium 2 project (085475/B/08/Z and 085475/Z/08/Z). Robert Plomin is supported by a Medical Research Council Research Professorship award

(G19/2) and a European Research Council Advanced Investigator award (295366). Additional support from the US National Institutes of Health [HD044454; HD059215] is acknowledged. Maciej Trzaskowski is supported by British Academy Post-doctoral Fellowship [pf140018]. For questions about accessing these data, please contact teds-project@kcl.ac.uk.

TRAILS (TRacking Adolescents' Individual Lives Survey) – TRAILS is a prospective cohort study of Dutch adolescents with bi- or triennial measurements from age 11 to up until adulthood (for a cohort profile see Oldehinkel et al., 2015). Five assessment waves have been completed to date, which ran from March 2001 to July 2002 (T1), September 2003 to December 2004 (T2), September 2005 to August 2007 (T3), October 2008 to September 2010 (T4), and January 2012 to December 2013 (T5). Data for the present study were collected during T4. At T1, 2230 (pre)adolescents were enrolled in the study (response rate 76%, mean age 11.1, SD 0.6, 51% girls (De Winter et al., 2005), of whom 84% (N = 1816, mean age 16.3, SD 0.7, 52% girls) participated at T4. Participating centers of TRAILS include the University Medical Center and University of Groningen, the Erasmus University Medical Center Rotterdam, the University of Utrecht, the Radboud Medical Center Nijmegen, and the Parnassia Bavo group, all in the Netherlands. TRAILS has been financially supported by various grants from the Netherlands Organization for Scientific Research NWO (Medical Research Council program grant GB-MW 940-38-011; ZonMW Brainpower grant 100-001-004; ZonMw Risk Behavior and Dependence grants 60-60600-97-118; ZonMw Culture and Health grant 261-98-710; Social Sciences Council medium-sized investment grants GB-MaGW 480-01-006 and GB-MaGW 480-07-001; Social Sciences Council project grants GB-MaGW 452-04-314 and GB-MaGW 452-06-004; NWO large-sized investment grant 175.010.2003.005; NWO Longitudinal Survey and Panel Funding 481-08-013 and 481-11-001), the Dutch Ministry of Justice (WODC), the European Science Foundation (EuroSTRESS project FP-006), Biobanking and Biomolecular Resources Research Infrastructure BBMRI-NL (CP 32), and the participating universities. Statistical analyses were carried out on the Genetic Cluster Computer (http://www.geneticcluster.org), which is financially supported by the Netherlands Scientific Organization (NWO 480-05-003) along with a supplement from the Dutch Brain Foundation. More information on TRAILS is available at: http://www.trails.nl/en/

TwinsUK (St. Thomas' UK Adult Twin Registry) – The study was funded by the Wellcome Trust; European Community's Seventh Framework Programme (FP7/2007-2013). The study also receives support from the National Institute for Health Research (NIHR) BioResource Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London. SNP Genotyping was performed by The Wellcome Trust Sanger Institute and National Eye Institute via NIH/CIDR. We would like to thank our funders, twin volunteers and TwinsUK team.
To obtain access to TwinsUK data please follow our data access policy (http://www.twinsuk.ac.uk/data-access/submission-procedure/).

YFS (The Cardiovascular Risk in Young Finns Study) – The Young Finns Study has been financially supported by the Academy of Finland: grants 286284 (T.L.), 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), 41071 (Skidi), 265869 (MIND; L.K.-J.), and 258711 (L.K.-J.); the Signe and Ane Gyllenberg Foundation (L.P.-R.); the Social Insurance Institution of Finland; Kuopio, Tampere and Turku University Hospital Medical Funds (grant X51001 for T.L.); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation of Cardiovascular Research (T.L.); Finnish Cultural Foundation; Tampere Tuberculosis Foundation (T.L.); Emil Aaltonen Foundation (T.L.); and Yrjö Jahnsson Foundation (T.L.). We gratefully acknowledge the THL DNA laboratory for its skillful work to produce the DNA samples used in this study, and Irina Lisinen for the expert technical assistance in the statistical analyses. For further information contact Olli Raitakari (olli.raitakari@utu.fi).

Individual acknowledgments – Philipp Koellinger acknowledges financial support from the European Research Council (consolidator grant 647648). Niels Rietveld acknowledges support from The Netherlands Organization for Scientific Research (Veni 016.165.004). Michel Nivard is supported by Royal Netherlands Academy of Science Professor Award (PAH/6635) to Dorret I. Boomsma. Joseph K. Pickrell and Jimmy Z. Liu were partially supported by the National Institute of Mental Health (R01 MH106842 to JKP).

12. References

1. Bartels, M. Genetics of wellbeing and its components satisfaction with life, happiness, and quality of life: a review and meta-analysis of heritability studies. *Behav. Genet.* **45,** 137–156 (2015).
2. Rietveld, C. A. *et al.* Molecular genetics and subjective well-being. *Proc. Natl. Acad. Sci.* **110,** 9692–9697 (2013).
3. Bartels, M. & Baselmans, B. in *Genetics of Psychological Well-Being* (ed. Pluess, M.) 163–176 (Oxford University Press, 2015).
4. De Neve, J.-E., Fowler, J. H., Frey, B. S. & Christakis, N. A. Genes, economics, and happiness. *J. Neurosci. Psychol. Econ.* **5,** 193–211 (2012).
5. Chen, H. *et al.* The MAOA gene predicts happiness in women. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **40,** 122–125 (2012).
6. Bartels, M. *et al.* Heritability and genome-wide linkage scan of subjective happiness. *Twin Res. Hum. Genet.* **13,** 135–142 (2010).
7. Kahneman, D. & Deaton, A. High income improves evaluation of life but not emotional well-being. *Proc. Natl. Acad. Sci. U. S. A.* **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. Busseri, M. A. Toward a resolution of the tripartite structure of subjective well-being. *J. Pers.* **83,** 413–428 (2015).
11. 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).
12. Diener, E., Emmons, R. A., Larsen, R. J. & Griffin, S. The satisfaction with life scale. *J. Pers. Assess.* **49,** 71–75 (1985).
13. Yesavage, J. *et al.* Development and validation of a geriatric depression screening scale: a preliminary report. *J. Psychiatr. Res.* **17,** 37–49 (1983).
14. Lyubomirsky, S. & Lepper, H. S. A measure of subjective happiness: preliminary reliability and construct validation. *Soc. Indic. Res.* **46,** 137–155 (1999).
15. Tellegen, A. & Waller, N. in *The SAGE Handbook of Personality Theory and Assessment* (eds. Boyle, G. J., Matthews, G. & H., S. D.) 261–272 (SAGE Publications, 2008).
16. Zigmond, S. & Snaith, R. P. The hospital anxiety and depression scale (HADS). *Acta Psychiatr. Scand.* **67,** 361–370 (1983).
17. Diener, E. *et al.* New measures of well-being: flourishing and positive and negative feelings. *Soc. Indic. Res.* **97,** 143–156 (2010).

18. Watson, D., Clark, L. A. & Tellegen, A. Development and validation of brief measures of positive and negative affect: the PANAS scales. *J. Pers. Soc. Psychol.* **54,** 1063–1070 (1988).

19. Radloff, L. S. The CES-D scale: a self-report depression scale for research in the general population. *Appl. Psychol. Meas.* **1,** 385–401 (1977).

20. Consortium, T. I. H. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449,** 851–861 (2007).

21. Winkler, T. W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9,** 1192–1212 (2014).

22. Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* (2016).

23. Li, Y., Willer, C. J., Ding, J., Scheet, P. & Abecasis, G. R. MaCH: Using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* **34,** 816–834 (2010).

24. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39,** 906–913 (2007).

25. Ripke, S. *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511,** 421–427 (2014).

26. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4,** 7 (2015).

27. Abecasis, G. R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491,** 56–65 (2012).

28. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47,** 291–295 (2015).

29. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26,** 2190–2191 (2010).

30. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55,** 997–1004 (1999).

31. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* **19,** 807–812 (2015).

32. Price, A. L. *et al.* Long-range LD can confound genome scans in admixed populations. *Am. J. Hum. Genet.* **83,** 132–135; author reply 135–139 (2008).

33. Mercer, T. R., Dinger, M. E. & Mattick, J. S. Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.* **10,** 155–9 (2009).

34. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45,** 580–5 (2013).

35. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42,** 565–569 (2010).

36. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47,** 1236–1241 (2015).

37. de Moor, M. H. M. *et al.* Meta-analysis of genome-wide association studies for neuroticism, and the polygenic association with Major Depressive Disorder. *JAMA Psychiatry* **72,** 642–650 (2015).

38. Ripke, S. *et al.* A mega-analysis of genome-wide association studies for major depressive disorder. *Mol. Psychiatry* **18,** 497–511 (2013).

39. dbGaP. Resource for Genetic Epidemiology Research on Adult Health and Aging (GERA). (2015). at <http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000674.v1.p1>

40. Smith, D. J. *et al.* Prevalence and characteristics of probable major depression and bipolar disorder within UK biobank: cross-sectional study of 172,751 participants. *PLoS One* **8,** e75362 (2013).

41. International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM). *2015* at <http://www.cdc.gov/nchs/icd/icd9cm.htm>

42. dbGaP. Phenotype Submission Based on ICD-9-CM as of March 15, 2013. (2013). at <http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/GetPdf.cgi?id=phd004308>

43. Eysenck, H. J. & Eysenck, S. B. G. *Manual of the Eysenck Personality Questionnaire*. (Hodder and Stroughton, 1975).

44. Marchini, J. *et al.* Genotype imputation and genetic association studies of UK Biobank. (2015). at <http://www.ukbiobank.ac.uk/wp-content/uploads/2014/04/imputation_documentation_May2015.pdf>

45. Sudmant, P. H. *et al.* An integrated map of structural variation in 2,504 human genomes. *Nature* **526,** 75–81 (2015).

46. Cai, N. *et al.* Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature* **523,** 588–591 (2015).

47. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44,** 369–375 (2012).

48. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88,** 76–82 (2011).

49. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47,** 1228–1235 (2015).

50. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).

51. Steinberg, K. M. *et al.* Structural diversity and African origin of the 17q21.31

inversion polymorphism. *Nat. Genet.* **44,** 872–80 (2012).

52. Stefansson, H. *et al.* A common inversion under selection in Europeans. *Nat. Genet.* **37,** 129–137 (2005).

53. Tian, C. *et al.* Analysis and application of European genetic substructure using 300 K SNP information. *PLoS Genet.* **4,** e4 (2008).

54. Ma, J. & Amos, C. I. Investigation of inversion polymorphisms in the human genome using principal components analysis. *PLoS One* **7,** e40224 (2012).

55. Abraham, G. & Inouye, M. Fast principal component analysis of large-scale genome-wide data. *PLoS One* **9,** e93766 (2014).

56. Benaglia, T., Chauveau, D., Hunter, D. R. & Young, D. mixtools : An R Package for Analyzing Finite Mixture Models. *J. Stat. Softw.* **32,** 1–29 (2009).

57. 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).

58. Abdellaoui, A. & Al., E. Population structure, migration, and diversifying selection in the Netherlands. *Eur. J. Hum. Genet.* **21,** 1277–1285 (2013).

59. Fehrmann, R. S. N. *et al.* Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. *PLoS Genet.* **7,** e1002197 (2011).

60. Westra, H.-J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45,** 1238–1243 (2013).

61. Fehrmann, R. S. N. *et al.* Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat. Genet.* **47,** 115–125 (2015).

62. Okazaki, H. *et al.* Increased cholesterol biosynthesis and hypercholesterolemia in mice overexpressing squalene synthase in the liver. *J. Lipid Res.* **47,** 1950–1958 (2006).

63. Borowiec, M. *et al.* Mutations at the BLK locus linked to maturity onset diabetes of the young and beta-cell dysfunction. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 14460–14465 (2009).

64. Hom, G. *et al.* Association of Systemic Lupus Erythematosus with C8orf13–BLK and ITGAM–ITGAX. *N. Engl. J. Med.* **358,** 900–909 (2008).

65. Lee, Y. H., Bae, S.-C., Choi, S. J., Ji, J. D. & Song, G. G. Genome-wide pathway analysis of genome-wide association studies on systemic lupus erythematosus and rheumatoid arthritis. *Mol. Biol. Rep.* **39,** 10627–10635 (2012).

66. Onouchi, Y. *et al.* A genome-wide association study identifies three new risk loci for Kawasaki disease. *Nat. Genet.* **44,** 517–521 (2012).

67. Moskovitz, J. *et al.* Methionine sulfoxide reductase (MsrA) is a regulator of antioxidant defense and lifespan in mammals. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 12920–12925 (2001).

68. Ruan, H. *et al.* High-quality life extension by the enzyme peptide methionine sulfoxide reductase. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 2748–2753 (2002).

69. Sakabe, T. *et al.* Identification of a novel gene, MASL1, within an amplicon at 8p23.1 detected in malignant fibrous histiocytomas by comparative genomic hybridization. *Cancer Res.* **59,** 511–515 (1999).

70. Koch, E., Ristroph, M. & Kirkpatrick, M. Long range linkage disequilibrium across the human genome. *PLoS One* **8,** (2013).

71. Sonnega, A. *et al.* Cohort Profile: the Health and Retirement Study (HRS). *Int. J. Epidemiol.* **43,** 576–585 (2014).

72. 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).

73. 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).

74. Costa, P. T. & Widiger, T. A. The 5-factor model of personality and its relevance to personality-disorders. *J. Pers. Disord.* **6,** (1992).

75. Achenbach, T. M. & Rescorla, L. A. *Manual for the ASEBA School - Age Forms & Profiles*. (University of Vermont, 2001).

76. Brim, O. G. *et al.* National Survey of Midlife Development in the United States (MIDUS), 1995-1996. (2003). doi:10.3886/ICPSR02760

77. Bradburn, N. M. *The Structure of Psychological Well-Being*. *Chicago: Aldine.* (1969). doi:10.1037/0022-3514.69.4.719

78. Kessler, R. C. *et al.* Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. *Archives of General Psychiatry* **51,** 8–19 (1994).

79. Taylor, J. A. A personality scale of manifest anxiety. *J. Abnorm. Psychol.* **48,** 285–290 (1953).

80. Macmillan, A. M. The health opinion survey: technique for estimating prevalence of psychoneurotic and related types of disorders in communities. *Psychol. Rep.* **3,** 325–339 (1957).

81. Fazio, A. *A Concurrent Validational Study of the NCHS General Well-Being Schedule.* (National Center for Health Statistics, 1977).

82. Watson, D. & Clark, L. A. *The PANAS-X manual for the positive and negative affect schedule - expanded form*. *Iowa Research Online* (1994).

83. Carstensen, L. L., Pasupathi, M., Mayr, U. & Nesselroade, J. R. Emotional experience in everyday life across the adult life span. *J. Pers. Soc. Psychol.* **79,** 644–655 (2000).

84. Ong, A. D., Edwards, L. M. & Bergeman, C. S. Hope as a source of resilience in later adulthood. *Pers. Individ. Dif.* **41,** 1263–1273 (2006).

85. Lachman, M. E. & Weaver, S. L. The Midlife Development Inventory (MIDI) personality scales: Scale construction and scoring. *Waltham, MA Brand. Univ.* 1–9 (1997).

86. Nelson, C. B., Kessler, R. C. & Mroczek, D. Scoring the World Health Organization's Composite International Diagnostic Interview Short Form (CIDI-SF; v1.0 NOV98). (1998). doi:10.1002/mpr.47

87. Vilhjálmsson, B. J. *et al.* Modeling linkage disequilibrium increases accuracy of polygenicrisk scores. *Am. J. Hum. Genet.* **97,** 576–592 (2015).

88. Purcell, S. M. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460,** 748–752 (2009).

89. Kendler, K. S. & Myers, J. The genetic and environmental relationship between major depression and the five-factor model of personality. *Psychol. Med.* **40,** 801 (2009).

90. 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).

91. 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).

92. Rietveld, C. A. *et al.* Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 13790–13794 (2014).

93. Nachar, N. The Mann-Whitney U: A test for assessing whether two independent samples come from the same distribution. *Quant. Methods Psychol.* **4,** 13–20 (2008).

94. Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467,** 832–838 (2010).

95. 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).

96. Alpérovitch, A. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45,** 1452–1458 (2013).

97. Otowa, T. *et al.* Meta-analysis of genome-wide association studies of anxiety disorders. *Mol. Psychiatry* (2016). doi:10.1038/mp.2015.197

98. Robinson, E. *et al.* Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *bioRxiv* (2015). doi:10.1101/027771

99. Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat. Genet.* **43,** 977–83 (2011).

100. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518,** 197–206 (2015).

101. Nikpay, M. *et al.* A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47,** 1121–30 (2015).

102. The Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat. Genet.* **42,** 441–7 (2010).

103. Dupuis, J. J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.* **42,** 105–16 (2010).

104. Willer, C. J. *et al.* Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45,** 1274–83 (2013).

105. Flèche, S. & Layard, R. *Do more of those in misery suffer from poverty, unemployment or mental illness?* (2015).

106. Glassman, A. J. *et al.* Smoking, smoking cessation, and major depression. *J. Am. Med. Assoc.* **264,** 1546–1549 (1990).

107. Davidson, K. W., Mostofsky, E. & Whang, W. Don't worry, be happy: positive affect and reduced 10-year incident coronary heart disease: The Canadian Nova Scotia Health Survey. *Eur. Heart J.* **31,** 1065–1070 (2010).

108. Stone, A. A., Cox, D. S., Valdimarsdottir, H., Jandorf, L. & Neale, J. M. Evidence that secretory IgA antibody is associated with daily mood. *J. Pers. Soc. Psychol.* **52,** 988–993 (1987).

109. Stone, A. A., Neale, J. M., Cox, D. S., Napoli, A. & Valdimarsdottir, H. Kennedy-Moore, E. Daily events are associated with secretory immune response to an oral antigen in men. *Heal. Psychol.* **13,** 440–446 (1994).

110. Evans, P. & Edgerton, N. Life events and mood as predictors of the common cold. *Br. J. Med. Psychol.* **64,** 35–44 (1991).

111. Stone, A. A., Reed, B. R. & Neale, J. M. Changes in daily event frequency precede episodes of physical symptomatology. *J. Human Stress* **13,** 70–74 (1987).

112. Cohen, S. *et al.* State and trait negative affect as predictors of objective and subjective symptoms of respiratory viral infections. *J. Pers. Soc. Psychol.* **68,** 159–169 (1995).

113. Mühleisen, T. W. *et al.* Genome-wide association study reveals two new risk loci for bipolar disorder. *Nat. Commun.* **5,** 3339 (2014).

114. 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).

115. Penninx, B. W. J. H., Milaneschi, Y., Lamers, F. & Vogelzangs, N. Understanding the somatic consequences of depression: biological mechanisms and the role of depression symptom profile. *BMC Med.* **11,** 129 (2013).

116. Daban, C., Vieta, E., Mackin, P. & Young, A. H. Hypothalamic-pituitary-adrenal axis and bipolar disorder. *Psychiatr. Clin. North Am.* **28,** 469–480 (2005).

117. Bradley, A. J. & Dinan, T. G. A systematic review of hypothalamic-pituitary-adrenal axis function in schizophrenia: implications for mortality. *J. Psychopharmacol.* **24,**

91–118 (2010).

118. Walker, E., Mittal, V. & Tessner, K. Stress and the hypothalamic pituitary adrenal axis in the developmental course of schizophrenia. *Annu. Rev. Clin. Psychol.* **4,** 189–216 (2008).

119. de Kloet, E. R., Joëls, M. & Holsboer, F. Stress and the brain: from adaptation to disease. *Nat. Rev. Neurosci.* **6,** 463–475 (2005).

120. 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).

121. Meng, L., Chen, D., Yang, Y., Zheng, Y. & Hui, R. Depression increases the risk of hypertension incidence: a meta-analysis of prospective cohort studies. *J. Hypertens.* **30,** 842–851 (2012).

122. 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,** D930–4 (2012).

123. Rademakers, R., Cruts, M. & van Broeckhoven, C. The role of tau (MAPT) in frontotemporal dementia and related tauopathies. *Hum. Mutat.* **24,** 277–295 (2004).

124. Sullivan, P. F., Fan, C. & Perou, C. M. Evaluating the comparability of gene expression in blood and brain. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **141B,** 261–268 (2006).

125. Albert, F. W. & Kruglyak, L. The role of regulatory variation in complex traits and disease. *Nat. Rev. Genet.* **16,** 197–212 (2015).

126. Jansen, R. *et al.* Gene expression in major depressive disorder. *Mol. Psychiatry* (2015).

127. Orth, M. & Bellosta, S. Cholesterol: its regulation and role in central nervous system disorders. *Cholesterol* **2012,** 292598 (2012).

128. Shrivastava, S., Pucadyil, T. J., Paila, Y. D., Ganguly, S. & Chattopadhyay, A. Chronic cholesterol depletion using statin impairs the function and dynamics of human serotonin(1A) receptors. *Biochemistry* **49,** 5426–5435 (2010).

129. Saxena, R. & Chattopadhyay, A. Membrane cholesterol stabilizes the human serotonin(1A) receptor. *Biochim. Biophys. Acta* **1818,** 2936–2942 (2012).

130. Kalipatnapu, S. & Chattopadhyay, A. Membrane organization and function of the serotonin(1A) receptor. *Cell. Mol. Neurobiol.* **27,** 1097–1116 (2007).

131. Saxena, R. & Chattopadhyay, A. Membrane organization and dynamics of the serotonin1A receptor in live cells. *J. Neurochem.* **116,** 726–733 (2011).

132. Andrews, P. W., Bharwani, A., Lee, K. R., Fox, M. & Thomson, J. A. Is serotonin an upper or a downer? The evolution of the serotonergic system and its role in depression and the antidepressant response. *Neurosci. Biobehav. Rev.* **51,** 164–188 (2015).

133. Nascimento, F. R. F., Gomes, E. A., Russo, M. & Lepique, A. P. Interferon regulatory factor (IRF)-1 is a master regulator of the cross talk between macrophages and L929 fibrosarcoma cells for nitric oxide dependent tumoricidal activity. *PLoS One* **10,** e0117782 (2015).

134. Loftis, J. M. & Hauser, P. The phenomenology and treatment of interferon-induced depression. *J. Affect. Disord.* **82,** 175–190 (2004).

135. Hou, X.-J., Xu, J.-H., Wang, J. & Yu, Y.-Y. Can antidepressants prevent pegylated interferon-α/ribavirin-associated depression in patients with chronic hepatitis C: meta-analysis of randomized, double-blind, placebo-controlled trials? *PLoS One* **8,** e76799 (2013).

136. Dowlati, Y. *et al.* A meta-analysis of cytokines in major depression. *Biol. Psychiatry* **67,** 446–57 (2010).

137. Cheyuo, C., Yang, W.-L. & Wang, P. The critical role of adrenomedullin and its binding protein, AMBP-1, in neuroprotection. *Biol. Chem.* **393,** 429–39 (2012).

138. Fernández, A. P., Serrano, J., Tessarollo, L., Cuttitta, F. & Martínez, A. Lack of adrenomedullin in the mouse brain results in behavioral changes, anxiety, and lower survival under stress conditions. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 12581–12586 (2008).

139. Ardlie, K. G. *et al.* The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348,** 648–660 (2015).

140. Altshuler, L. L. *et al.* Does thyroid supplementation accelerate tricyclic antidepressant response? A review and meta-analysis of the literature. *Am. J. Psychiatry* **158,** 1617–1622 (2001).

141. Hendrick V, Altshuler L, W. P. Psychoneuroendocrinology of mood disorders. The hypothalamic-pituitary-thyroid axis. *Psychitric Clin N Am* **21,** 277–292 (1998).

142. Penninx, B. W. J. H. *et al.* Late-life depressive symptoms are associated with both hyperactivity and hypoactivity of the hypothalamo-pituitary-adrenal axis. *Am. J. Geriatr. Psychiatry* **15,** 522–529 (2007).

143. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42,** D1001–1006 (2014).

144. Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518,** 187–196 (2015).

145. Nalls, M. A. *et al.* Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat. Genet.* **46,** 989–93 (2014).

146. Hibar, D. P. *et al.* Common genetic variants influence human subcortical brain structures. *Nature* **520,** 224–229 (2015).

147. Wagner, G. P. & Zhang, J. The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. *Nat. Rev. Genet.* **12,** 204–213 (2011).

148. Litvan, I. *et al.* Which clinical features differentiate progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome) from related disorders? A clinicopathological study. *Brain* **120,** 65–74 (1997).

149. Seeman, P. Dopamine D2 receptors as treatment targets in schizophrenia. *Clin. Schizophr. Relat. Psychoses* **4,** 56–73 (2010).

150. Bergen, S. E. *et al.* Genome-wide association study in a Swedish population yields support for greater CNV and MHC involvement in schizophrenia compared with bipolar disorder. *Mol. Psychiatry* **17,** 880–886 (2012).

151. Sleiman, P. *et al.* GWAS meta analysis identifies TSNARE1 as a novel Schizophrenia / Bipolar susceptibility locus. *Sci. Rep.* **3,** 3075 (2013).

152. Ruderfer, D. M. *et al.* Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. *Mol. Psychiatry* **19,** 1017–1024 (2013).

153. Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6,** 5890 (2015).

154. Pers, T. H., Timshel, P. & Hirschhorn, J. N. SNPsnap: a Web-based tool for identification and annotation of matched SNPs. *Bioinformatics* **31,** 418–420 (2015).

155. Consortium, T. G. O. Gene ontology: Tool for the identification of biology. *Nat. Genet.* **25,** 25–29 (2000).

156. Croft, D. *et al.* Reactome: A database of reactions, pathways and biological processes. *Nucleic Acids Res.* **39,** 691–697 (2011).

157. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **40,** 109–114 (2012).

158. Hyde, C. L. et al. Common genetic variants associated with major depressive disorder among individuals of European descent. *Nat. Genet.*