

Serveur Académique Lausannois SERVAL serval.unil.ch

Author Manuscript

Faculty of Biology and Medicine Publication

This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Published in final edited form as:

Title: Genome-wide association study identifies 74 loci associated with educational attainment.

Authors: Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, Turley P, Chen GB, Emilsson V, Meddens SF, Oskarsson S, Pickrell JK, Thom K, Timshel P, de Vlaming R, Abdellaoui A, Ahluwalia TS, Bacelis J, Baumbach C, Bjornsdottir G, Brandsma JH, Pina Concas M, Derringer J, Furlotte NA, Galesloot TE, Girotto G, Gupta R, Hall LM, Harris SE, Hofer E, Horikoshi M, Huffman JE, Kaasik K, Kalafati IP, Karlsson R, Kong A, Lahti J, van der Lee SJ, deLeeuw C, Lind PA, Lindgren KO, Liu T, Mangino M, Marten J, Mihailov E, Miller MB, van der Most PJ, Oldmeadow C, Payton A, Pervjakova N, Peyrot WJ, Qian Y, Raitakari O, Rueedi R, Salvi E, Schmidt B, Schraut KE, Shi J, Smith AV, Poot RA, St Pourcain B, Teumer A, Thorleifsson G, Verweij N, Vuckovic D, Wellmann J, Westra HJ, Yang J, Zhao W, Zhu Z, Alizadeh BZ, Amin N, Bakshi A, Baumeister SE, Biino G, Bønnelykke K, Boyle PA, Campbell H,

In the absence of a copyright statement, users should assume that standard copyright protection applies, unless the article contains an explicit statement to the contrary. In case of doubt, contact the journal publisher to verify the copyright status of an article.



Published in final edited form as:

Nature. ; 533(7604): 539–542. doi:10.1038/nature17671.

Genome-wide association study identifies 74 loci associated with educational attainment

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

Summary

Educational attainment (EA) is strongly influenced by social and other environmental factors, but genetic factors are also estimated to account for at least 20% of the variation across individuals¹. We report the results of a genome-wide association study (GWAS) for EA that extends our earlier discovery sample^{1,2} of 101,069 individuals to 293,723 individuals, and a replication in an independent sample of 111,349 individuals from the UK Biobank. We now identify 74 genome-wide significant loci associated with number of years of schooling completed. Single-nucleotide polymorphisms (SNPs) associated with educational attainment are disproportionately found in genomic regions regulating gene expression in the fetal brain. Candidate genes are preferentially expressed in neural tissue, especially during the prenatal period, and enriched for biological pathways involved in neural development. Our findings demonstrate that, even for a behavioral phenotype that is mostly environmentally determined, a well-powered GWAS identifies replicable associated genetic variants that suggest biologically relevant pathways. Because EA is measured in large numbers of individuals, it will continue to be useful as a proxy phenotype in efforts to characterize the genetic influences of related phenotypes, including cognition and neuropsychiatric disease.

We study educational attainment (EA), which is measured in all main analyses as the number of years of schooling completed (*EduYears*, $N = 293,723$, mean = 14.33, SD = 3.61; Supplementary Information sections 1.1-1.2). All genome-wide association studies (GWAS) were performed at the cohort level in samples restricted to individuals of European descent

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

*These authors contributed equally.

#Designed and oversaw the study.

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Author contributions Study Design and Management: D.J.B., D.C., T.E., M.J., P.D.K. and P.M.V. Quality Control and Meta Analysis: A.O., G.B.C., T.E., M.A.F., C.A.R. and T.H.P. Stratification: P.T., J.P.B., C.A.R. and J.Y. Genetic Overlap: J.P.B., M.A.F., P.T. Biological Annotation: J.J.L., T.E., J.H.B., J.K.B., J.P.B., L.F., V.E., L.F., G.A.M., M.A.F., S.F.W.M., G.A.M., T.H.P., J.K.P., P.Timshel, R.A.P., R.d.V. and H.J.W. Prediction and Mediation: J.P.B., M.A.F. and J.Y. G×E: D.Conley, S.F.L., K.O.L., S.O. and K.T. Replication in UKB: M.A.F. and C.A.R. SSGAC Advisory Board: D.Conley, T.E., A.H., R.F.K., D.I.L., S.E.M., M.N.M., G.D.S. and P.M.V. All authors contributed to and critically reviewed the manuscript. Authors not listed above contributed to the recruitment, genotyping, or data processing for the contributing components of the meta-analysis. For a full list of author contributions, see Supplementary Information section 8.

Author Information Results can be downloaded from the SSGAC website (<http://ssgac.org/Data.php>). Data for our analyses come from many studies and organizations, some of which are subject to a MTA, and are listed in the Supplementary Information. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests.

Correspondence and requests for materials should be addressed to D.J.B. (daniel.benjamin@gmail.com), D.C. (dac12@nyu.edu), P.D.K. (p.d.koellinger@vu.nl), or P.M.V. (peter.visscher@uq.edu.au).

whose EA was assessed at or above age 30. A uniform set of quality-control (QC) procedures was applied to the cohort-level summary statistics. In our GWAS meta-analysis of ~9.3M SNPs from the 1000 Genomes Project, we used sample-size weighting and applied a single round of genomic control at the cohort level.

Our meta-analysis identified 74 approximately independent genome-wide significant loci. For each locus, we define the “lead SNP” as the SNP in the genomic region that has the smallest P -value (Supplementary Information section 1.6.1). Fig. 1 shows a Manhattan plot with the lead SNPs highlighted. This includes the three SNPs that reached genome-wide significance in the discovery stage of our previous GWAS meta-analysis of EA¹. The quantile-quantile (Q-Q) plot of the meta-analysis (Extended Data Fig. 1) exhibits inflation ($\lambda_{GC} = 1.28$), as expected under polygenicity³.

Extended Data Fig. 2 shows the estimated effect sizes of the lead SNPs. The estimates range from 0.014 to 0.048 standard deviations per allele (2.7 to 9.0 weeks of schooling), with incremental R^2 in the range 0.01% to 0.035%.

To quantify the amount of population stratification in the GWAS estimates that remains even after the stringent controls used by the cohorts (Supplementary Information section 1.4), we used LD Score regression⁴. The regression results indicate that ~8% of the observed inflation in the mean χ^2 is due to bias rather than polygenic signal (Extended Data Fig. 3a), suggesting that stratification effects are small in magnitude. We also found evidence for polygenic association signal in several within-family analyses, although these are not powered for individual SNP association testing (Supplementary Information section 2 and Extended Data Fig. 3b).

To further test the robustness of our findings, we examined the within-sample and out-of-sample replicability of SNPs reaching genome-wide significance (Supplementary Information sections 1.7-1.8). We found that SNPs identified in the previous EA meta-analysis replicated in the new cohorts included here, and conversely, that SNPs reaching genome-wide significance in the new cohorts replicated in the old cohorts. For the out-of-sample replication analyses of our 74 lead SNPs, we used the interim release of the U.K. Biobank⁵ (UKB) ($N = 111,349$). As shown in Extended Data Fig. 4, 72 out of the 74 lead SNPs have a consistent sign ($P = 1.47 \times 10^{-19}$), 52 are significant at the 5% level ($P = 2.68 \times 10^{-50}$), and 7 reach genome-wide significance in the U.K. Biobank dataset ($P = 1.41 \times 10^{-42}$). For comparison, the corresponding expected numbers, assuming each SNP's true effect size is its estimated effect adjusted for the winner's curse, are 71.4, 40.3, and 0.6. (Supplementary Information section 1.8.2). We also find out-of-sample replicability of our overall GWAS results: the genetic correlation between *Edu Years* in our meta-analysis sample and in the UKB data is 0.95 (s.e. = 0.021; Supplementary Table 1.14).

It is known that EA, cognitive performance, and many neuropsychiatric phenotypes are phenotypically correlated, and several studies of twins find that the phenotypic correlations partly reflect genetic overlap⁶⁻⁸ (Supplementary Information section 3.3.4). Here, we investigate genetic correlation using our GWAS results for *Edu Years* and published GWAS results for 14 other phenotypes, using bivariate Linkage-Disequilibrium (LD) Score

regression⁹ (Supplementary Information section 3). First, we estimated genetic correlations with *EduYears*. As shown in Fig. 2, based on overall summary statistics for associated variants, we find genetic covariance between increased EA and increased cognitive performance ($P = 9.9 \times 10^{-50}$), increased intracranial volume ($P = 1.2 \times 10^{-6}$), increased risk of bipolar disorder ($P = 7 \times 10^{-13}$), decreased risk of Alzheimer's ($P = 4 \times 10^{-4}$), and lower neuroticism ($P = 2.8 \times 10^{-8}$). We also found positive, statistically significant, but very small, genetic correlations with height ($P = 5.2 \times 10^{-15}$) and risk of schizophrenia ($P = 3.2 \times 10^{-4}$).

Second, we examined whether our 74 lead SNPs are jointly associated with each phenotype (Extended Data Fig. 5 and Supplementary Information section 3.3.1). We reject the null hypothesis of no enrichment at $P < 0.05$ for 10 of the 14 phenotypes (all the exceptions are subcortical brain structures).

Third, for each phenotype, we tested (in the published GWAS results) each of our 74 lead SNPs or proxy for association at a significance threshold of $0.05/74$. We found a total of 25 SNPs meeting this threshold for any of these phenotypes, but only one reaching genome-wide significance. While these results provide suggestive evidence that some of these SNPs may be associated with other phenotypes, further testing of these associations in independent cohorts is required (Supplementary Tables 3.2-3.4, Extended Data Fig. 6).

To consider potential biological pathways, we first tested whether SNPs in particular regions of the genome are implicated by our GWAS results. Unlike what has been found for other phenotypes, SNPs in regions that are DNase I hypersensitive in the fetal brain are more likely to be associated with *EduYears* by a factor of ~ 5 (95% confidence interval 2.89–7.07; Extended Data Fig. 7). Moreover, the 15% of SNPs residing in regions associated with histones marked in the central nervous system (CNS) explain 44% of the heritable variation (Extended Data Fig. 8a and Supplementary Table 4.4.2). This enrichment factor of ~ 3 for CNS ($P = 2.48 \times 10^{-16}$) is greater than that of any of the other nine tissue categories in this analysis.

Given that our findings disproportionately implicate SNPs in regions regulating brain-specific gene expression, we examined whether genes located near *EduYears*-associated SNPs show elevated expression in neural tissue. We tested this hypothesis using data on mRNA transcript levels in the 37 adult tissues assayed by the Genotype-Tissue Expression Project (GTEx)¹⁰. Remarkably, the 13 GTEx tissues that are components of the CNS—and only those 13 tissues—show significantly elevated expression levels of genes near *EduYears*-associated SNPs (FDR < 0.05 ; Extended Data Fig. 8b and Supplementary Table 4.5.2).

To investigate possible functions of the candidate genes from the GWAS associated loci, we examined the extent of their overlap with groups of genes (“gene sets”) whose products are known or predicted to participate in a common biological process¹¹. We found 283 gene sets significantly enriched by the candidate genes identified in our GWAS (FDR < 0.05 ; Supplementary Table 4.5.1). To facilitate interpretation, we used a standard procedure¹¹ to group the 283 gene sets into “clusters” defined by degree of gene overlap. The resulting 34 clusters, shown in Fig. 3, paint a coherent picture, with many clusters corresponding to

stages of neural development: the proliferation of neural progenitor cells and their specialization (the *cluster npBAF complex*), the migration of new neurons to the different layers of the cortex (*forebrain development, abnormal cerebral cortex morphology*), the projection of axons from neurons to their signaling targets (*axonogenesis, signaling by Robo receptor*), the sprouting of dendrites and their spines (*dendrite, dendritic spine organization*), and neuronal signaling and synaptic plasticity throughout the lifespan (*voltage-gated calcium channel complex, synapse part, synapse organization*).

Many of our results implicate candidate genes and biological pathways that are active during distinct stages of prenatal brain development. To directly examine how the expression levels of candidate genes identified in our GWAS vary over the course of development, we used gene expression data from the BrainSpan Developmental Transcriptome¹². As shown in Extended Data Fig. 9, these candidate genes exhibit above-baseline expression in the brain throughout life but especially higher expression levels in the brain during prenatal development (1.36 times higher prenatally than postnatally, $P = 6.02 \times 10^{-8}$).

A summary overview of some promising candidate genes for follow-up work is provided in Table 1.

We constructed polygenic scores¹³ to assess the joint predictive power afforded by the GWAS results (Supplementary Information section 5.2). Across our two holdout samples, the mean predictive power of a polygenic score constructed from all measured SNPs is 3.2% ($P = 1.18 \times 10^{-39}$; Supplementary Table 5.2 and Supplementary Information section 5).

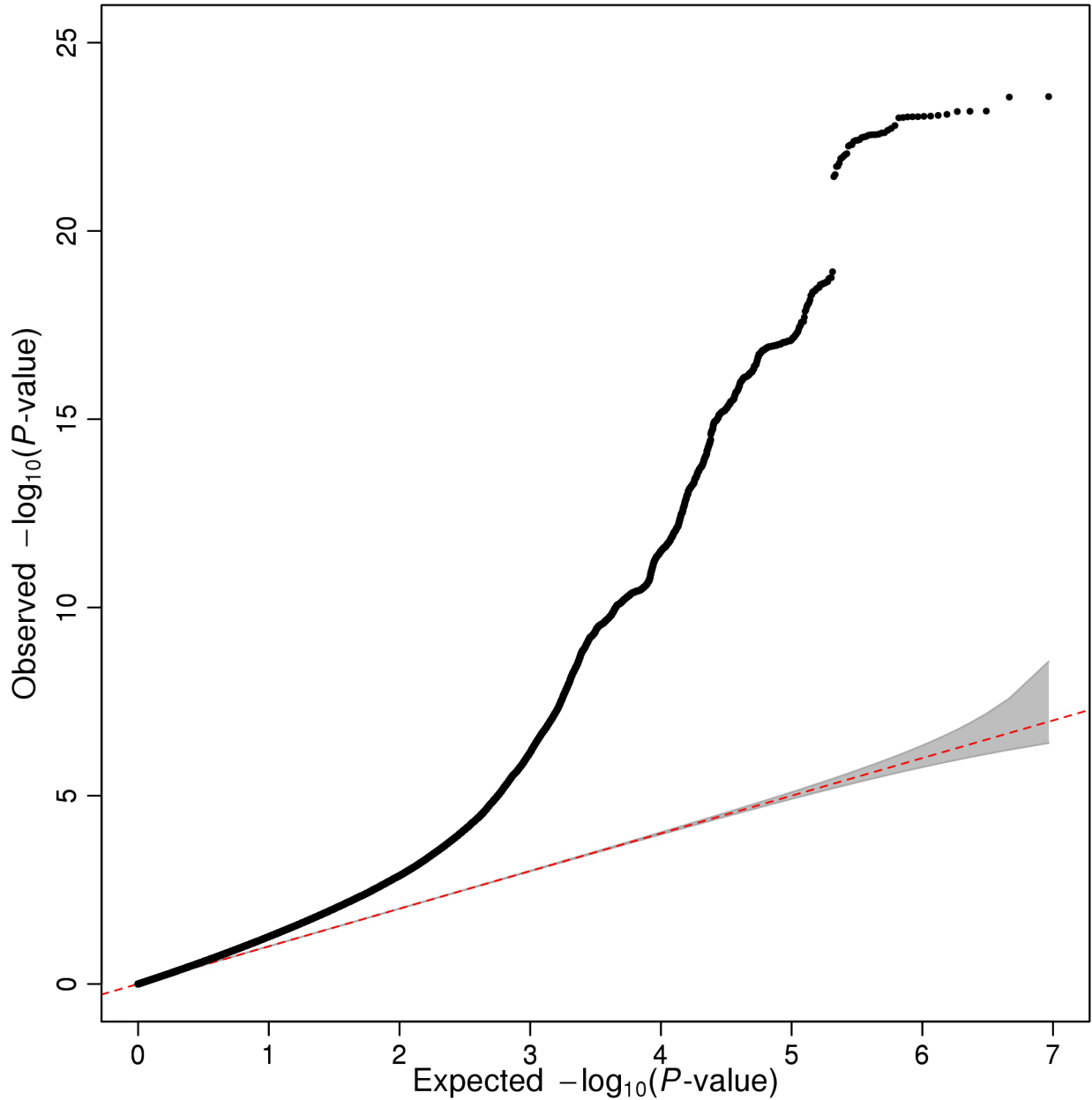
Studies of genetic analyses of behavioral phenotypes have been prone to misinterpretation, such as characterizing identified associated variants as “genes for education.” Such characterization is not correct for many reasons: EA is primarily determined by environmental factors, the explanatory power of the individual SNPs is small, the candidate genes may not be causal, and the genetic associations with EA are mediated by multiple intermediate phenotypes¹⁴. To illustrate this last point, we studied mediation of the association between the all-SNPs polygenic score and *EduYears* in two of our cohorts. We found that cognitive performance can statistically account for 23-42% of the association ($P < 0.001$) and the personality trait “openness to experience” for approximately 7% ($P < 0.001$; Supplementary Information section 6).

It would also be a mistake to infer from our findings that the genetic effects operate independently of environmental factors. Indeed, a recent meta-analysis of twin studies found that genetic influences on EA are heterogeneous across countries and birth cohorts¹⁵. We conducted exploratory analyses in the Swedish Twin Registry to illustrate how environmental factors may amplify or dampen the impact of genetic influences (Supplementary Information section 7). We found that the predictive power of the all-SNPs polygenic score is heterogeneous by birth cohort, with smaller explanatory power in younger cohorts (Extended Data Fig. 10; see also Supplementary Information section 7.4 for discussion of the contrast between these results and findings from a seminal twin study that estimated EA heritability by birth cohort¹⁶).

Methods

All methods are described in the Supplementary Information.

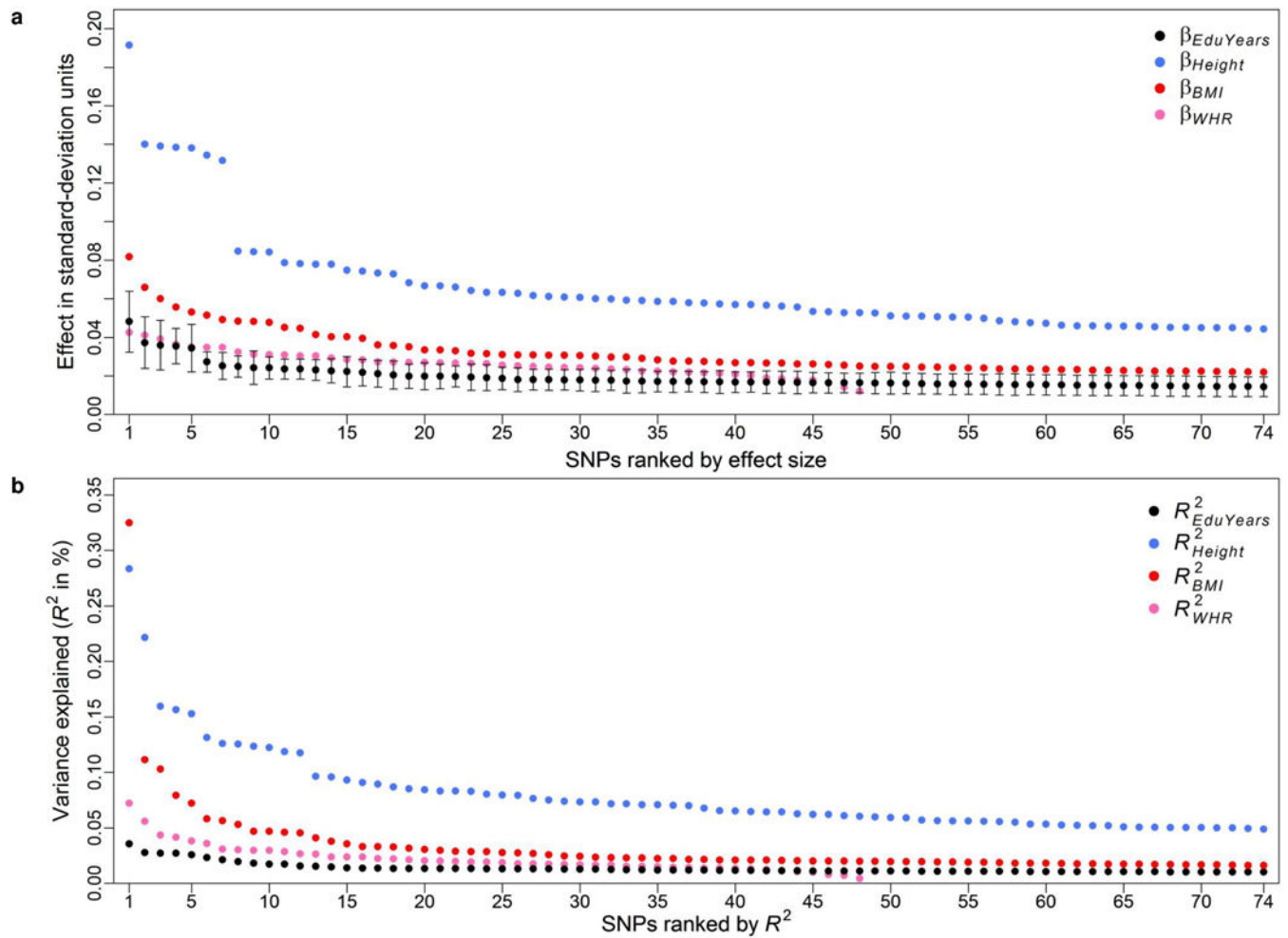
Extended Data



Extended Data Figure 1. Quantile-quantile plot of the genome-wide association meta-analysis of 64 *EduYears* results files

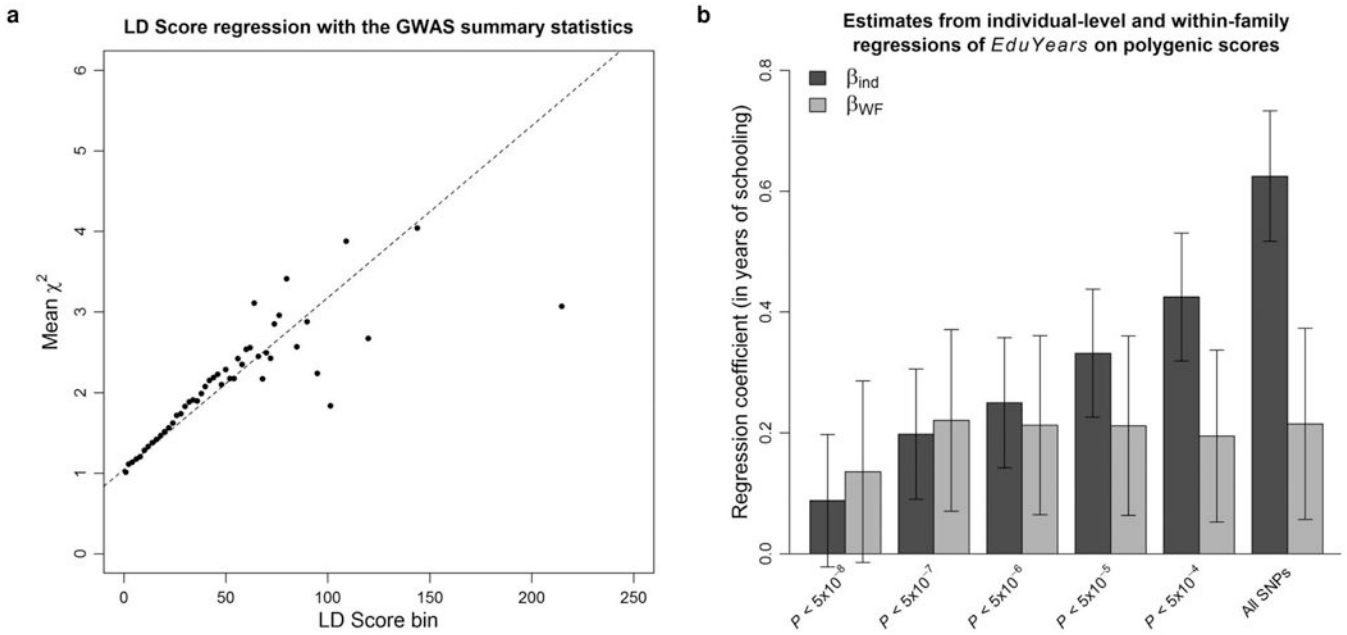
Observed and expected P -values are on a $-\log_{10}$ scale. The grey region depicts the 95% confidence interval under the null hypothesis of a uniform P -value distribution. The observed λ_{GC} is 1.28. (As reported in Supplementary Information section 1.5.4, the

unweighted mean λ_{GC} is 1.02, the unweighted median is 1.01, and the range across cohorts is 0.95–1.15.)



Extended Data Figure 2. The distribution of effect sizes of the 74 lead SNPs

a, SNPs ordered by absolute value of the standardized effect of one more copy of the education-increasing allele, with 95% confidence intervals. **b**, SNPs ordered by R^2 . Effects on *Edu Years* are benchmarked against the top 74 genome-wide significant hits identified in the largest GWAS conducted to date of height and body mass index (BMI), and the 48 associations reported for waist-to-hip ratio adjusted for BMI (WHR). These results are based on the GIANT consortium's publicly available results for pooled analyses restricted to European-ancestry individuals: https://www.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium.



Extended Data Figure 3. Assessing the extent to which population stratification affects the estimates from the GWAS

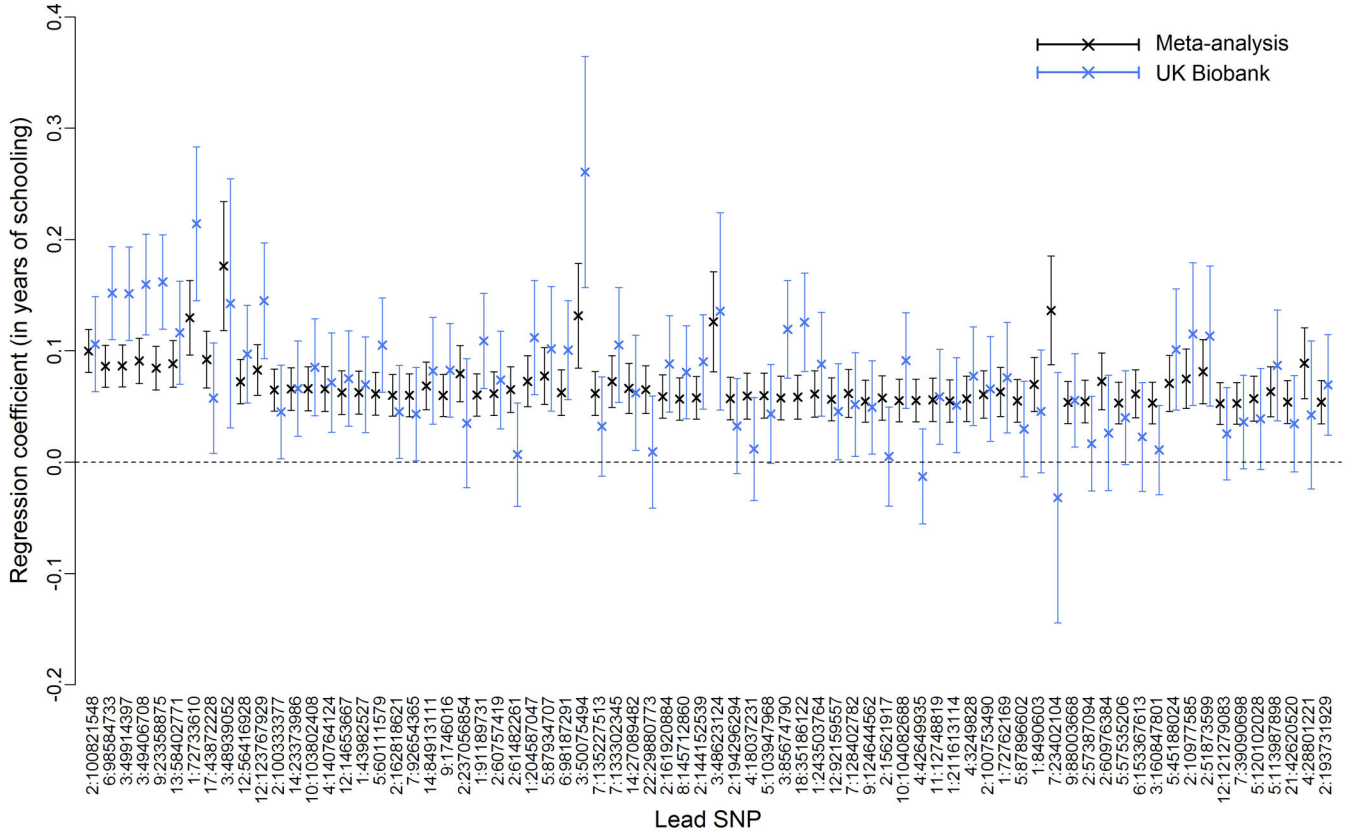
a, LD Score regression plot with the summary statistics from the GWAS. Each point represents an LD Score quantile for a chromosome (the x and y coordinates of the point are the mean LD Score and the mean χ^2 statistic of variants in that quantile). The facts that the intercept is close to one and that the χ^2 statistics increase linearly with the LD Scores suggest that the bulk of the inflation in the χ^2 statistics is due to true polygenic signal and not to population stratification. **b**, Estimates and 95% confidence intervals from individual-level and WF regressions of *EduYears* on polygenic scores, for scores constructed with sets of SNPs meeting different P -value thresholds. In addition to the analyses shown here, we conduct a sign concordance test, and we decompose the variance of the polygenic score. Overall, these analyses suggest that population stratification is unlikely to be a major concern for our 74 lead SNPs. See Supplementary Information section 3 for additional details.

Author Manuscript

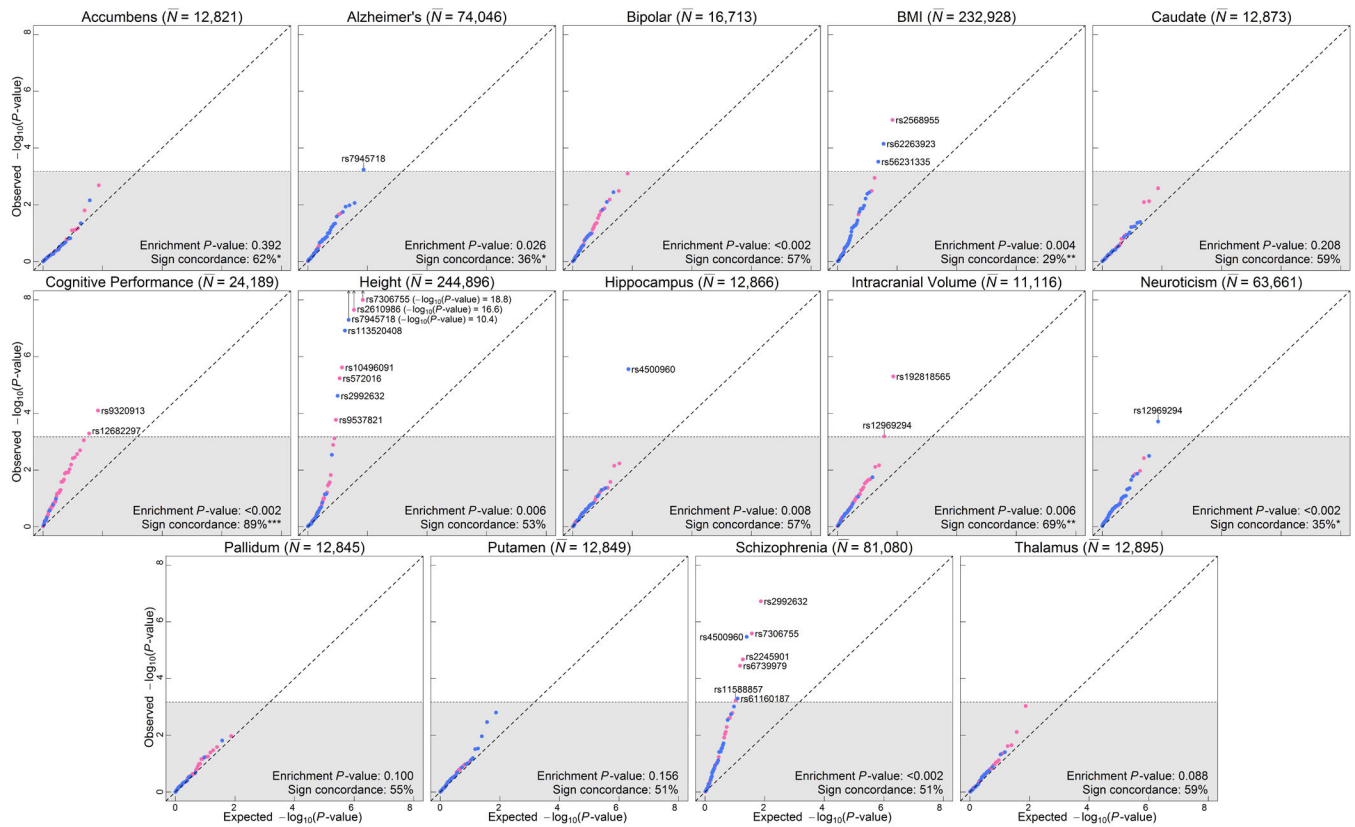
Author Manuscript

Author Manuscript

Author Manuscript

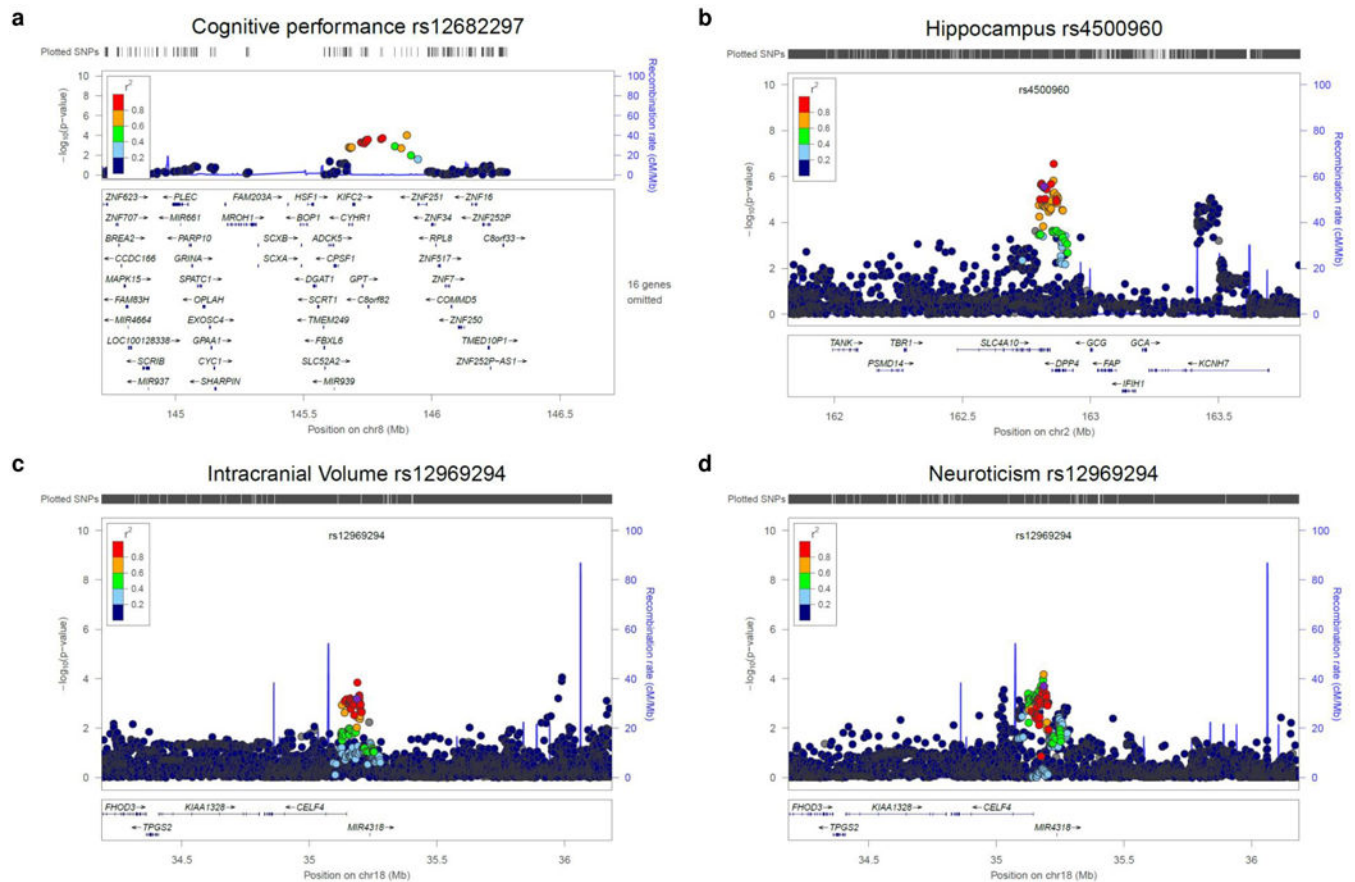


Extended Data Figure 4. Replication of 74 lead SNPs in the UK Biobank data
 Estimated effect sizes (in years of schooling) and 95% confidence intervals of the 74 lead SNPs in the meta-analysis sample ($N = 293,723$) and the UK Biobank replication sample ($N = 111,349$). The reference allele is the allele associated with higher values of *Edu Years* in the meta-analysis sample. SNPs are in descending order of R^2 in the meta-analysis sample. Of the 74 lead SNPs, 72 have the anticipated sign in the replication sample, 52 replicate at the 0.05 significance level, and 7 replicate at the 5×10^{-8} significance level.

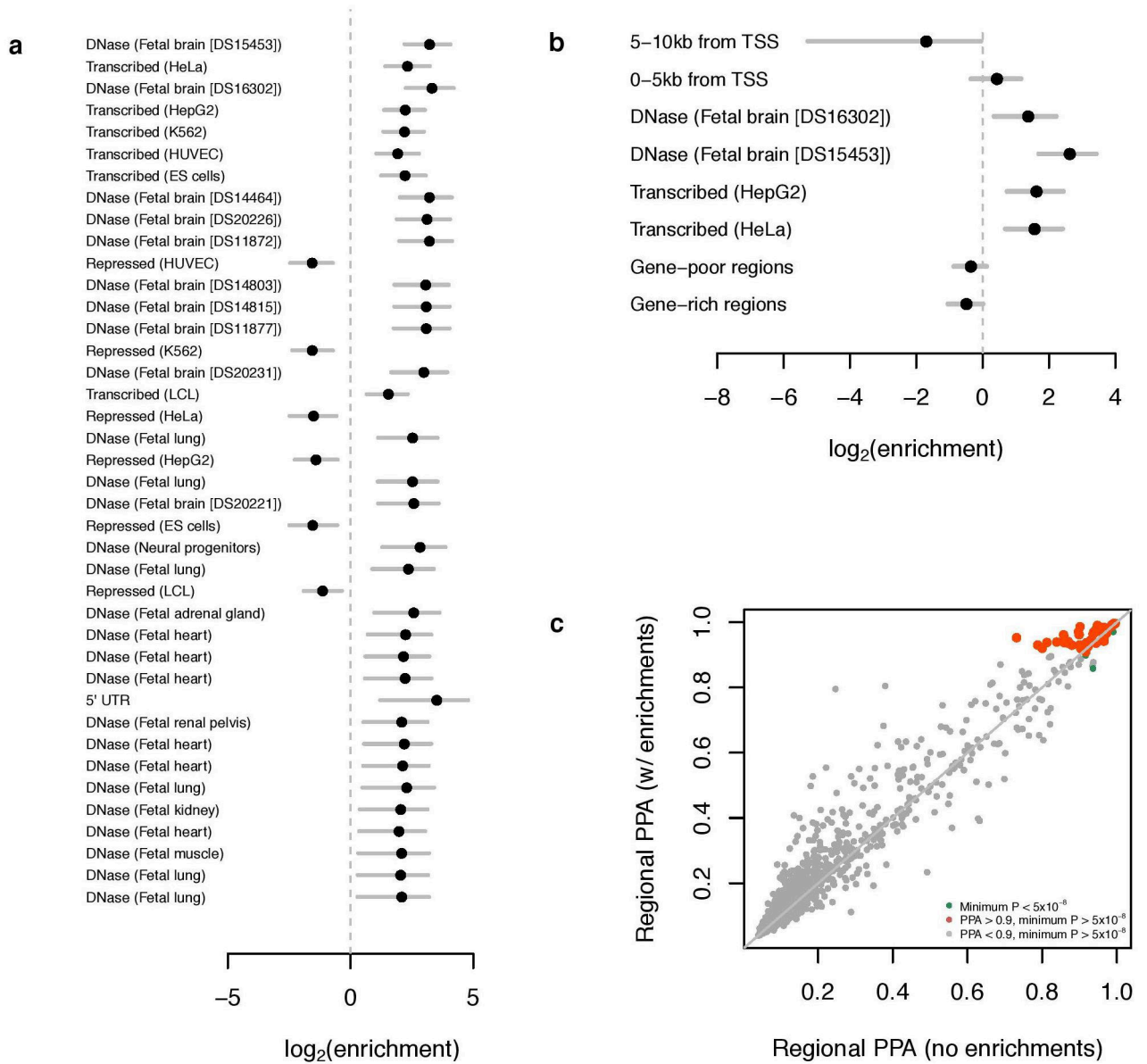


Extended Data Figure 5. Q-Q plots for the 74 lead *EduYears* SNPs (or LD proxies) in published GWAS of other phenotypes

SNPs with concordant effects on both phenotypes are pink, and SNPs with discordant effects are blue. SNPs outside the gray area pass Bonferroni-corrected significance thresholds that correct for the total number of SNPs we tested ($P < 0.05/74 = 6.8 \times 10^{-4}$) and are labeled with their rs numbers. Observed and expected P -values are on a $-\log_{10}$ scale. For the sign concordance test: * $P < 0.05$, ** $P < 0.01$, and *** $P < 0.001$.



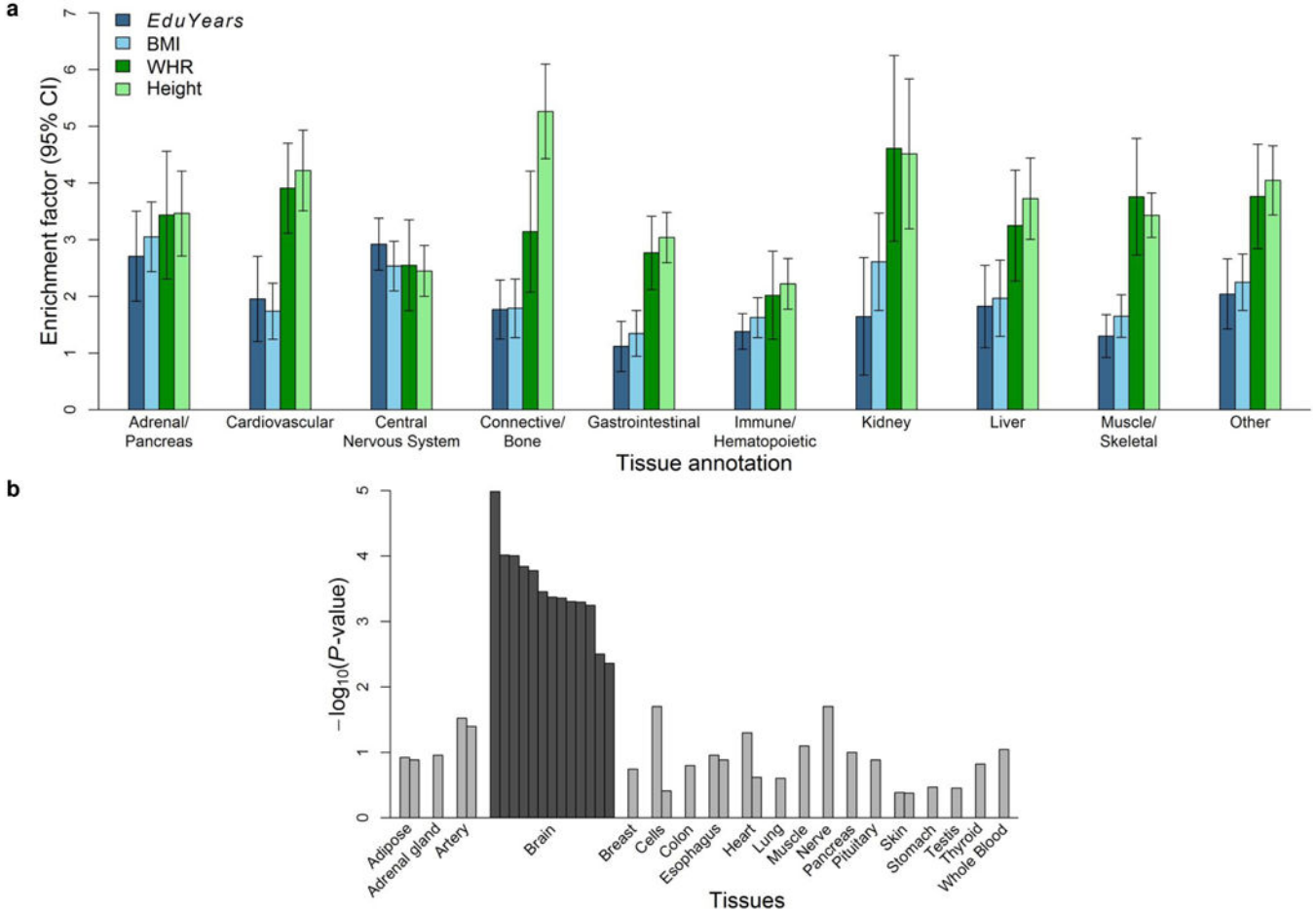
Extended Data Figure 6. Regional association plots for four of the ten prioritized SNPs for MHBA phenotypes identified using *EduYears* as a proxy phenotype **a**, cognitive performance; **b**, hippocampus; **c**, intracranial volume; **d**, neuroticism. The four were selected because very few genome-wide significant SNPs have been previously reported for these traits. Data sources and methods are described in Supplementary Information section 3. The R^2 values are from the hg19 / 1000 Genomes Nov 2014 EUR references samples. The figures were created with LocusZoom (<http://csg.sph.umich.edu/locuszoom/>). Mb, megabases.



Extended Data Figure 7. Application of fgwas to *EduYears*. See Supplementary Information section 4.2 for further details

a, The results of single-annotation models. “Enrichment” refers to the factor by which the prior odds of association at an LD-defined region must be multiplied if the region bears the given annotation; this factor is estimated using an empirical Bayes method applied to all SNPs in the GWAS meta-analysis regardless of statistical significance. Annotations were derived from ENCODE and a number of other data sources. Plotted are the base-2 logarithms of the enrichments and their 95% confidence intervals. Multiple instances of the same annotation correspond to independent replicates of the same experiment. **b**, The results of combining multiple annotations and applying model selection and cross-validation. Although the maximum-likelihood estimates are plotted, model selection was performed with penalized likelihood. **c**, Reweighting of GWAS loci. Each point represents an LD-

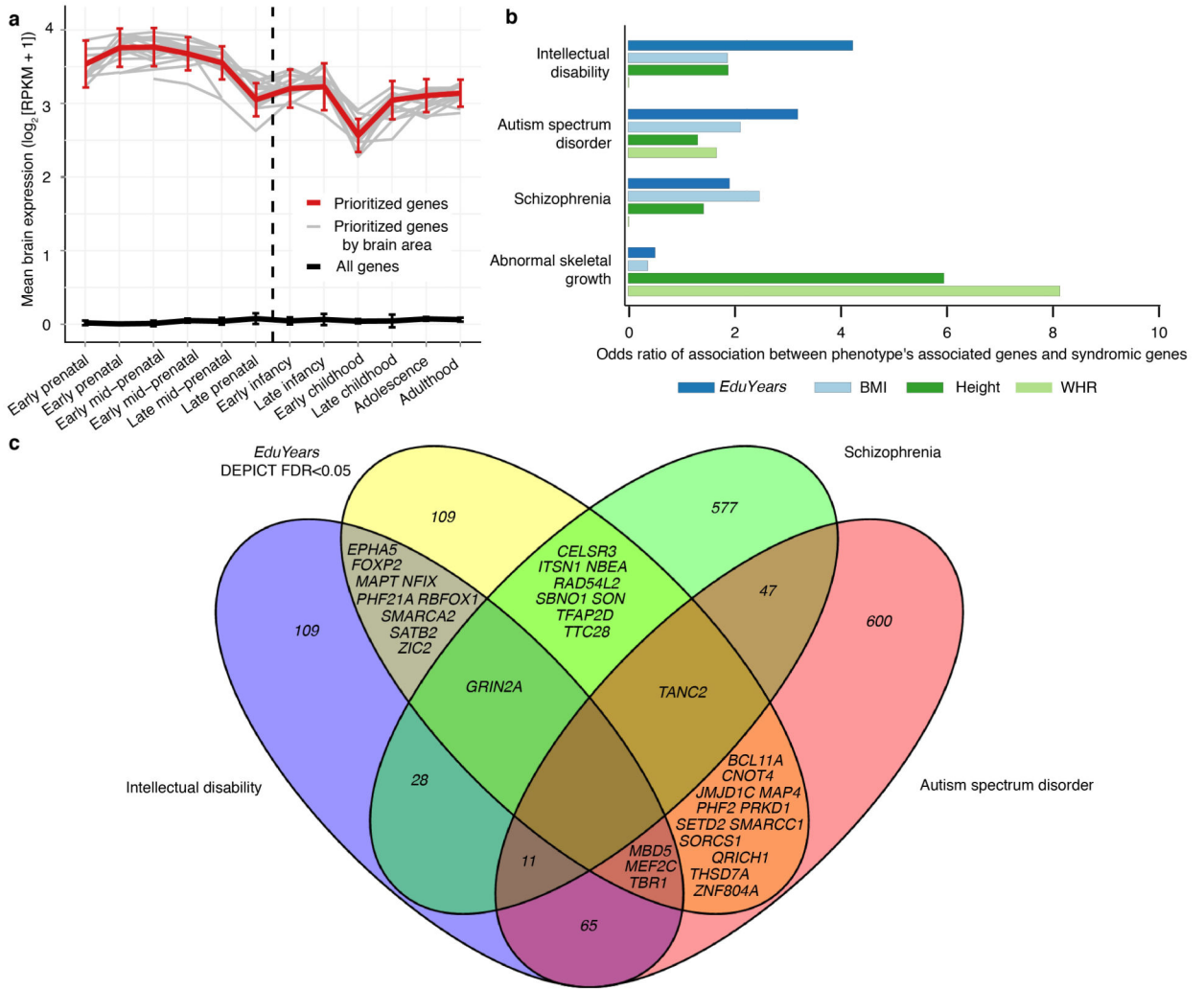
defined region of the genome, and shown are the regional posterior probabilities of association (PPAs). The *x*-axis give the PPA calculated from the GWAS summary statistics alone, whereas the *y*-axis gives the PPA upon reweighting on the basis of the annotations in b. The orange points represent genomic regions where the PPA is equivalent to the standard GWAS significance threshold only upon reweighting.



Extended Data Figure 8. Tissue-level biological annotation

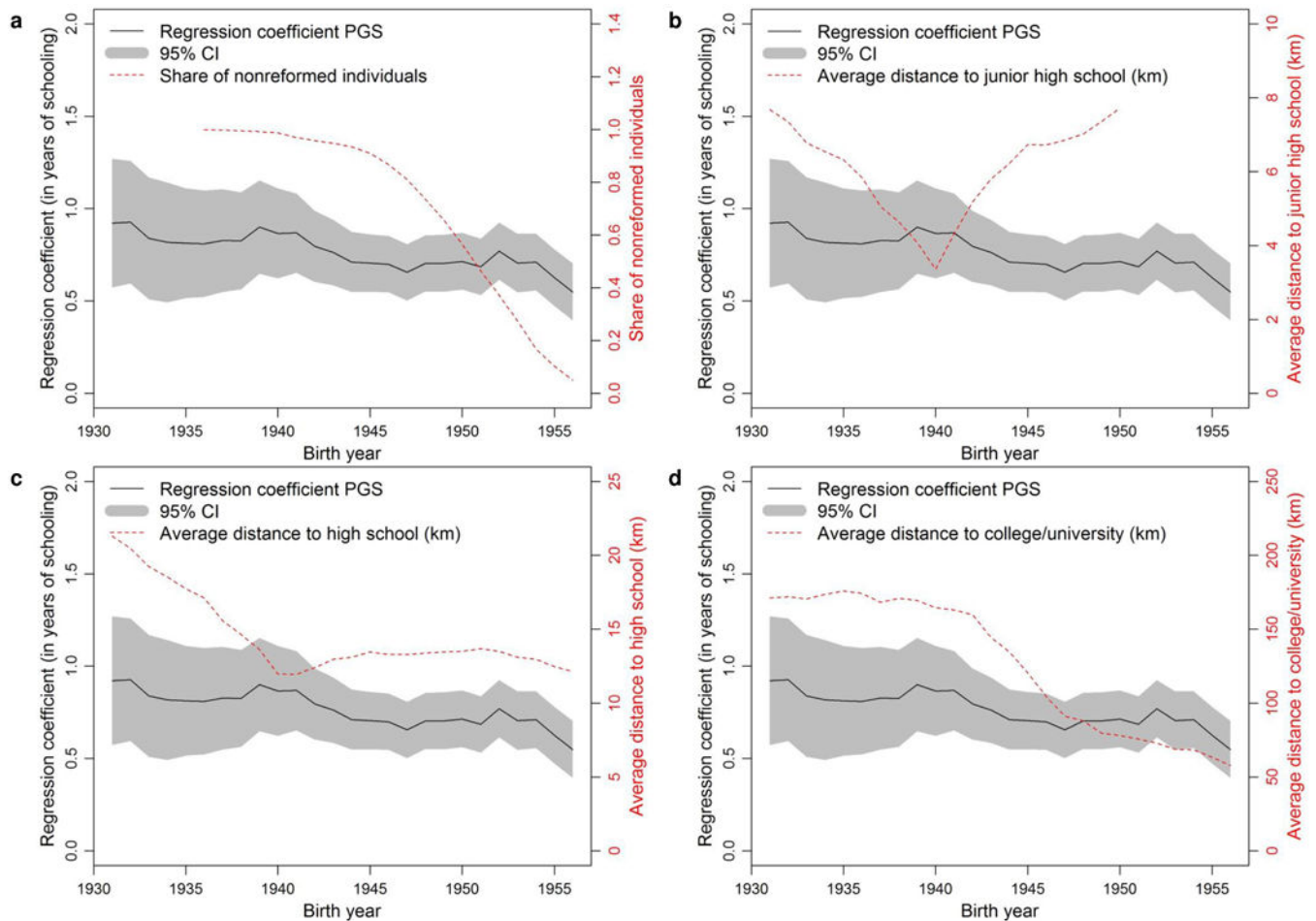
a, The enrichment factor for a given tissue type is the ratio of variance explained by SNPs in that group to the overall fraction of SNPs in that group. To benchmark the estimates for *Edu Years*, we compare the enrichment factors to those obtained when we use the largest GWAS conducted to date on body mass index, height, and waist-to-hip ratio adjusted for BMI. The estimates were produced with the LDSC python software, using the LD Scores and functional annotations introduced in Finucane et al. (2015) and the HapMap3 SNPs with MAF > 0.05. Each of the 10 enrichment calculations for a particular cell type is performed independently, while each controlling for the 52 functional annotation categories in the full baseline model. The error bars show the 95% confidence intervals. **b**, We took measurements of gene expression by the Genotype-Tissue Expression (GTEx) Consortium and determined whether the genes overlapping *Edu Years*-associated loci are significantly overexpressed (relative to genes in random sets of loci matched by gene density) in each of

37 tissue types. These types are grouped in the panel by organ. The colored bars corresponding to tissues where there is significant overexpression. The y-axis is the significance on a $-\log_{10}$ scale.



Extended Data Figure 9. Gene-level biological annotation

a, The DEPICT-prioritized genes for *EduYears* measured in the BrainSpan Developmental Transcriptome data (red curve) are more strongly expressed in the brain prenatally rather than postnatally. The DEPICT-prioritized genes exhibit similar gene-expression levels across different brain regions (gray lines). Analyses were based on \log_2 -transformed RNA-Seq data. Error bars represent 95% confidence intervals. **b**, For each phenotype and disorder, we calculated the overlap between the phenotype's DEPICT-prioritized genes and genes believed to harbor *de novo* mutations causing the disorder. The bars correspond to odds ratios. *EduYears*, years of education; BMI, body mass index; WHR, waist-to-hip ratio adjusted for BMI. **c**, DEPICT-prioritized genes in *EduYears*-associated loci exhibit substantial overlap with genes previously reported to harbor sites where mutations increase risk of intellectual disability and autism spectrum disorder (Supplementary Table 4.6.1).



Extended Figure 10. The predictive power of a polygenic score (PGS) varies in Sweden by birth cohort

Five-year rolling regressions of years of education on the PGS (left axis in all four panels), share of individuals not affected by the comprehensive school reform (**a**, right axis), and average distance to nearest junior high school (**b**, right axis), nearest high school (**c**, right axis) and nearest college/university (**d**, right axis). The shaded area displays the 95% confidence intervals for the PGS effect.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Aysu Okbay^{1,2,3,*}, Jonathan P. Beauchamp^{4,*}, Mark A. Fontana^{5,*}, James J. Lee^{6,*}, Tune H. Pers^{7,8,9,10,*}, Cornelius A. Rietveld^{1,2,3,*}, Patrick Turley^{4,*}, Guo-Bo Chen¹¹, Valur Emilsson^{12,13}, S. Fleur W. Meddens^{14,3,15}, Sven Oskarsson¹⁶, Joseph K. Pickrell¹⁷, Kevin Thom¹⁸, Pascal Timshel^{19,8}, Ronald de Vlaming^{1,2,3}, Abdel Abdellaoui²⁰, Tarunveer S. Ahluwalia^{21,9,22}, Jonas Bacelis²³, Clemens Baumbach^{24,25}, Gyda Bjornsdottir⁹⁵, Johannes H. Brandsma²⁶, Maria Pina

Concas²⁷, Jaime Derringer²⁸, Nicholas A. Furlotte²⁹, Tessel E. Galesloot³⁰, Giorgia Giroto³¹, Richa Gupta³², Leanne M. Hall^{33,34}, Sarah E. Harris^{35,36}, Edith Hofer^{37,38}, Momoko Horikoshi^{39,40}, Jennifer E. Huffman⁴¹, Kadri Kaasik⁴², Ioanna P. Kalafati⁴³, Robert Karlsson⁴⁴, Augustine Kong⁹⁵, Jari Lahti^{42,45}, Sven J. van der Lee², Christiaan de Leeuw^{14,46}, Penelope A. Lind⁴⁷, Karl-Oskar Lindgren¹⁶, Tian Liu⁴⁸, Massimo Mangino^{49,50}, Jonathan Marten⁴¹, Evelin Mihailov¹¹⁴, Michael B. Miller⁶, Peter J. van der Most⁵¹, Christopher Oldmeadow^{52,53}, Antony Payton^{54,55}, Natalia Pervjakova^{56,114}, Wouter J. Peyrot⁵⁷, Yong Qian⁵⁸, Olli Raitakari⁵⁹, Rico Rueedi^{60,61}, Erika Salvi⁶², Borge Schmidt⁶³, Katharina E. Schraut⁶⁴, Jianxin Shi⁶⁵, Albert V. Smith^{66,67}, Raymond A. Poot²⁶, Beate Pourcain St^{68,69}, Alexander Teumer⁷⁰, Gudmar Thorleifsson⁹⁵, Niek Verweij⁷¹, Dragana Vuckovic³¹, Juergen Wellmann⁷², Harm-Jan Westra^{73,74,8}, Jingyun Yang^{75,76}, Wei Zhao⁷⁷, Zhihong Zhu¹¹, Behrooz Z. Alizadeh^{51,78}, Najaf Amin², Andrew Bakshi¹¹, Sebastian E. Baumeister^{70,79}, Ginevra Biino⁸⁰, Klaus Bønnelykke²¹, Patricia A. Boyle^{75,81}, Harry Campbell⁶⁴, Francesco P. Cappuccio⁸², Gail Davies^{35,83}, Jan-Emmanuel De Neve⁸⁴, Panos Deloukas^{85,86}, Ilja Demuth^{87,88}, Jun Ding⁵⁸, Peter Eibich^{89,90}, Lewin Eisele⁶³, Niina Eklund⁵⁶, David M. Evans^{68,184}, Jessica D. Faul⁹¹, Mary F. Feitosa⁹², Andreas J. Forstner^{93,94}, Ilaria Gandin³¹, Bjarni Gunnarsson⁹⁵, Bjarni V. Halldórsson^{95,96}, Tamara B. Harris⁹⁷, Andrew C. Heath⁹⁸, Lynne J. Hocking⁹⁹, Elizabeth G. Holliday^{52,53}, Georg Homuth¹⁰⁰, Michael A. Horan¹⁰¹, Jouke-Jan Hottenga²⁰, Philip L. de Jager^{102,103,8}, Peter K. Joshi⁶⁴, Astanand Jugessur¹⁰⁴, Marika A. Kaakinen¹⁰⁵, Mika Kähönen^{106,107}, Stavroula Kanoni⁸⁵, Liisa Keltigangas-Järvinen⁴², Lambertus A.L.M. Kiemenev³⁰, Ivana Kolcic¹⁰⁸, Seppo Koskinen⁵⁶, Aldi T. Kraja⁹², Martin Kroh⁸⁹, Zoltan Kutalik^{109,60,61}, Antti Latvala³², Lenore J. Launer¹¹⁰, Maël P. Lebreton^{15,111}, Douglas F. Levinson¹¹², Paul Lichtenstein⁴⁴, Peter Lichtner¹¹⁸, David C.M. Liewald^{35,83}, LifeLines Cohort Study¹¹³, Anu Loukola³², Pamela A. Madden⁹⁸, Reedik Mägi¹¹⁴, Tomi Mäki-Opas⁵⁶, Riccardo E. Marioni^{35,115,11}, Pedro Marques-Vidal¹¹⁶, Gerardus A. Meddens¹¹⁷, George McMahon⁶⁸, Christa Meisinger²⁵, Thomas Meitinger¹¹⁸, Yusplitri Milaneschi⁵⁷, Lili Milani¹¹⁴, Grant W. Montgomery¹¹⁹, Ronny Myhre¹⁰⁴, Christopher P. Nelson^{33,34}, Dale R. Nyholt^{120,119}, William E.R. Ollier⁵⁴, Aarno Palotie^{121,8,122,123,124,125}, Lavinia Paternoster⁶⁸, Nancy L. Pedersen⁴⁴, Katja E. Petrovic³⁷, David J. Porteous³⁶, Katri Räikkönen^{42,45}, Susan M. Ring⁶⁸, Antonietta Robino¹²⁶, Olga Rostapshova^{4,127}, Igor Rudan⁶⁴, Aldo Rustichini¹²⁸, Veikko Salomaa⁵⁶, Alan R. Sanders^{129,130}, Antti-Pekka Sarin^{124,131}, Helena Schmidt^{132,37}, Rodney J. Scott^{133,53}, Blair H. Smith¹³⁴, Jennifer A. Smith⁷⁷, Jan A. Staessen^{135,136}, Elisabeth Steinhagen-Thiessen⁸⁷, Konstantin Strauch^{137,138}, Antonio Terracciano¹³⁹, Martin D. Tobin¹⁴⁰, Sheila Ulivi¹²⁶, Simona Vaccargiu²⁷, Lydia Quaye⁴⁹, Frank J.A. van Rooij^{2,141}, Cristina Venturini^{49,50}, Anna A.E. Vinkhuyzen¹¹, Uwe Völker¹⁰⁰, Henry Völzke⁷⁰, Judith M. Vonk⁵¹, Diego Vozzi¹²⁶, Johannes Waage^{21,22}, Erin B. Ware^{77,142}, Gonneke Willemsen²⁰, John R. Attia^{52,53}, David A. Bennett^{75,76}, Klaus Berger⁷¹, Lars Bertram^{143,144}, Hans Bisgaard²¹, Dorret I. Boomsma²⁰, Ingrid B. Borecki⁹², Ute Bultmann¹⁴⁵, Christopher F. Chabris¹⁴⁶, Francesco Cucca¹⁴⁷, Daniele Cusi^{62,148}, Ian J. Deary^{35,83}, George V. Dedoussis⁴³, Cornelia M. van Duijn², Johan G. Eriksson^{149,45}, Barbara Franke¹⁵⁰,

Lude Franke¹⁵⁵, Paolo Gasparini^{31,126,151}, Pablo V. Gejman^{129,130}, Christian Gieger²⁴, Hans-Jürgen Grabe^{152,153}, Jacob Gratten¹¹, Patrick J.F. Groenen¹⁵⁴, Vilmundur Gudnason^{12,67}, Pim van der Harst^{71,155,156}, Caroline Hayward^{41,157}, David A. Hinds²⁹, Wolfgang Hoffmann⁷⁰, Elina Hyppönen^{158,159,160}, William G. Iacono⁶, Bo Jacobsson^{23,104}, Marjo-Riitta Järvelin^{161,162,163,164}, Karl-Heinz Jöckel⁶³, Jaakko Kaprio^{32,124,56}, Sharon L.R. Kardia⁷⁷, Terho Lehtimäki^{165,166}, Steven F. Lehrer^{167,168}, Patrik K.E. Magnusson⁴⁴, Nicholas G. Martin¹⁶⁹, Matt McGue⁶, Andres Metspalu^{114,170}, Neil Pendleton^{171,172}, Brenda W.J.H. Penninx⁵⁷, Markus Perola^{56,114}, Nicola Pirastu³¹, Mario Pirastu²⁷, Ozren Polasek^{173,64}, Danielle Posthuma^{14,174}, Christine Power¹⁶⁰, Michael A. Province⁹², Nilesh J. Samani^{33,34}, David Schlessinger⁵⁸, Reinhold Schmidt³⁷, Thorkild I.A. Sørensen^{175,9,68}, Tim D. Spector⁴⁹, Kari Stefansson^{95,67}, Unnur Thorsteinsdottir^{95,67}, A. Roy Thurik^{1,176,177,3}, Nicholas J. Timpson⁶⁸, Henning Tiemeier^{2,178,179}, Joyce Y. Tung²⁹, André G. Uitterlinden^{180,22}, Veronique Vitart⁴¹, Peter Vollenweider¹¹⁶, David R. Weir⁹¹, James F. Wilson^{64,41}, Alan F. Wright⁴¹, Dalton C. Conley^{181,182}, Robert F. Krueger⁶, George Davey Smith⁶⁸, Albert Hofman², David I. Laibson⁴, Sarah E. Medland⁴⁷, Michelle N. Meyer¹⁸³, Jian Yang^{11,184}, Magnus Johannesson¹⁸⁵, Peter M. Visscher^{11,184,#}, Tõnu Esko^{114,7,186,8,#}, Philipp D. Koellinger^{14,15,3,#}, David Cesarini^{18,187,#}, and Daniel J. Benjamin^{188,189,#}

Affiliations

¹Department of Applied Economics, Erasmus School of Economics, Erasmus University Rotterdam, 3062 PA, Rotterdam, The Netherlands ²Department of Epidemiology, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands ³Erasmus University Rotterdam Institute for Behavior and Biology, Rotterdam 3062 PA, The Netherlands ⁴Department of Economics, Harvard University, Cambridge, MA 02138, USA ⁵Department of Economics, University of Michigan, Ann Arbor, MI 48109-1220, USA ⁶Department of Psychology, University of Minnesota Twin Cities, Minneapolis, MN 55455, USA ⁷Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, MA 2116, USA ⁸Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA ⁹The Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, University of Copenhagen, Faculty of Health and Medical Sciences, Copenhagen, 2100, Denmark ¹⁰Statens Serum Institut, Department of Epidemiology Research, Copenhagen, DK 2300, Denmark ¹¹Queensland Brain Institute, The University of Queensland, Brisbane, QLD 4072, Australia ¹²Icelandic Heart Association, Kopavogur, 201, Iceland ¹³Faculty of Pharmaceutical Sciences, University of Iceland, 107 Reykjavik, Iceland ¹⁴Department of Complex Trait Genetics, VU University, Center for Neurogenomics and Cognitive Research, Amsterdam, 1081 HV, The Netherlands ¹⁵Amsterdam Business School, University of Amsterdam, Amsterdam, 1018 TV, The Netherlands ¹⁶Department of Government, Uppsala University, Uppsala, 751 20, Sweden ¹⁷New York Genome Center, New York, NY 10013, USA ¹⁸Department of Economics, New York University, New York, NY 10012, USA ¹⁹Center for Biological Sequence

Analysis, Department of Systems Biology, Technical University of Denmark Lyngby, 2800, Denmark ²⁰Department of Biological Psychology, VU University Amsterdam, Amsterdam, 1081 BT, The Netherlands ²¹COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, 2820, Denmark ²²Steno Diabetes Center, Gentofte, 2820, Denmark ²³Department of Obstetrics and Gynecology, Institute of Clinical Sciences, Sahlgrenska Academy, Gothenburg, SE 416 85, Sweden ²⁴Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, 85764, Germany ²⁵Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, 85764, Germany ²⁶Department of Cell Biology, Erasmus Medical Center Rotterdam, 3015 CN, The Netherlands ²⁷Istituto di Ricerca Genetica e Biomedica U.O.S. di Sassari, National Research Council of Italy, Sassari, 07100, Italy ²⁸Psychology, University of Illinois, IL 61820, Champaign, USA ²⁹23andMe, Inc., Mountain View, CA 94041, USA ³⁰Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, 6500 HB, The Netherlands ³¹Department of Medical, Surgical and Health Sciences, University of Trieste, Trieste, 34100, Italy ³²Department of Public Health, University of Helsinki, Helsinki, FI-00014, Finland ³³Department of Cardiovascular Sciences, University of Leicester, Leicester, LE3 9QP, UK ³⁴NIHR Leicester Cardiovascular Biomedical Research Unit, Glenfield Hospital, Leicester, LE3 9QP, UK ³⁵Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, EH8 9JZ, UK ³⁶Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK ³⁷Department of Neurology, General Hospital and Medical University Graz, Graz, 8036, Austria ³⁸Institute for Medical Informatics, Statistics and Documentation, General Hospital and Medical University Graz, Graz, 8036, Austria ³⁹Oxford Centre for Diabetes, Endocrinology & Metabolism, University of Oxford, Oxford, OX3 7LE, UK ⁴⁰Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK ⁴¹MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK ⁴²Institute of Behavioural Sciences, University of Helsinki, Helsinki, FI-00014, Finland ⁴³Nutrition and Dietetics, Health Science and Education, Harokopio University, Athens, 17671, Greece ⁴⁴Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, 171 77, Sweden ⁴⁵Folkhälsan Research Centre, Helsingfors, FI-00014, Finland ⁴⁶Institute for Computing and Information Sciences, Radboud University Nijmegen, Nijmegen, 6525 EC, The Netherlands ⁴⁷Quantitative Genetics, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4029, Australia ⁴⁸Lifespan Psychology, Max Planck Institute for Human Development, Berlin, 14195, Germany ⁴⁹Department of Twin Research and Genetic Epidemiology, King's College London, London, SE1 7EH, UK ⁵⁰NIHR Biomedical Research Centre, Guy's and St. Thomas' Foundation Trust, London, SE1 7EH, UK ⁵¹Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, 9700 RB, The Netherlands ⁵²Public Health Stream, Hunter Medical Research Institute,

New Lambton, NSW 2305, Australia ⁵³Faculty of Health and Medicine, University of Newcastle, Newcastle, NSW 2300, Australia ⁵⁴Centre for Integrated Genomic Medical Research, Institute of Population Health, The University of Manchester, Manchester, M13 9PT, UK ⁵⁵School of Psychological Sciences, The University of Manchester, Manchester, M13 9PL, UK ⁵⁶Department of Health, THL-National Institute for Health and Welfare, Helsinki, FI-00271, Finland ⁵⁷Psychiatry, VU University Medical Center & GGZ inGeest, Amsterdam, 1081 HL, The Netherlands ⁵⁸Laboratory of Genetics, National Institute on Aging, Baltimore, MD 21224, USA ⁵⁹Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, 20521, Finland ⁶⁰Department of Medical Genetics, University of Lausanne, Lausanne, 1005, Switzerland ⁶¹Swiss Institute of Bioinformatics, Lausanne, 1015, Switzerland ⁶²Department Of Health Sciences, University of Milan, Milano, 20142, Italy ⁶³Institute for Medical Informatics, Biometry and Epidemiology, University Hospital of Essen, Essen, 45147, Germany ⁶⁴Centre for Global Health Research, The Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, EH8 9AG, UK ⁶⁵Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD 20892-9780, USA ⁶⁶Icelandic Heart Association, Kopavogur, 201, Iceland ⁶⁷Faculty of Medicine, University of Iceland, Reykjavik, 101, Iceland ⁶⁸MRC Integrative Epidemiology Unit, University of Bristol, Bristol, BS8 2BN, UK ⁶⁹School of Oral and Dental Sciences, University of Bristol, Bristol, BS1 2LY, UK ⁷⁰Institute for Community Medicine, University Medicine Greifswald, Greifswald, 17475, Germany ⁷¹Department of Cardiology, University Medical Center Groningen, University of Groningen, Groningen, 9700 RB, The Netherlands ⁷²Institute of Epidemiology and Social Medicine, University of Muenster, Muenster, 48149, Germany ⁷³Divisions of Genetics and Rheumatology, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, MA 02115, Boston, USA ⁷⁴Partners Center for Personalized Genetic Medicine, Boston, MA 02115, USA ⁷⁵Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL 60612, USA ⁷⁶Department of Neurological Sciences, Rush University Medical Center, Chicago, IL 60612, USA ⁷⁷Department of Epidemiology, University of Michigan, Ann Arbor, MI 48109, USA ⁷⁸Department of Gastroenterology and Hepatology, University of Groningen, University Medical Center Groningen, Groningen, 9713 GZ, The Netherlands ⁷⁹Institute of Epidemiology and Preventive Medicine, University of Regensburg, Regensburg, D-93053, Germany ⁸⁰Institute of Molecular Genetics, National Research Council of Italy, Pavia, 27100, Italy ⁸¹Department of Behavioral Sciences, Rush University Medical Center, Chicago, IL 60612, USA ⁸²Warwick Medical School, University of Warwick, Coventry, CV4 7AL, UK ⁸³Department of Psychology, University of Edinburgh, Edinburgh, EH8 9JZ, UK ⁸⁴Saïd Business School, University of Oxford, Oxford, OX1 1HP, UK ⁸⁵William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, EC1M 6BQ, UK ⁸⁶Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah, 21589, Saudi Arabia ⁸⁷The Berlin Aging Study II;

Research Group on Geriatrics, Charité – Universitätsmedizin Berlin, Germany, Berlin, 13347, Germany ⁸⁸Institute of Medical and Human Genetics, Charité-Universitätsmedizin, Berlin, Berlin, 13353, Germany ⁸⁹German Socio- Economic Panel Study, DIW Berlin, Berlin, 10117, Germany ⁹⁰Health Economics Research Centre, Nuffield Department of Population Health, University of Oxford, Oxford, OX3 7LF, UK ⁹¹Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI 48109, USA ⁹²Department of Genetics, Division of Statistical Genomics, Washington University School of Medicine, St. Louis, MO 63018, USA ⁹³Institute of Human Genetics, University of Bonn, Bonn, 53127, Germany ⁹⁴Department of Genomics, Life and Brain Center, University of Bonn, Bonn, 53127, Germany ⁹⁵deCODE Genetics/Amgen Inc., Reykjavik, IS-101, Iceland ⁹⁶Institute of Biomedical and Neural Engineering, School of Science and Engineering, Reykjavik University, Reykjavik 101, Iceland ⁹⁷Laboratory of Epidemiology, Demography, National Institute on Aging, National Institutes of Health, Bethesda, MD 20892-9205, United States ⁹⁸Department of Psychiatry, Washington University School of Medicine, St. Louis, MO 63110, USA ⁹⁹Division of Applied Health Sciences, University of Aberdeen, Aberdeen, AB25 2ZD, UK ¹⁰⁰Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, 17475, Germany ¹⁰¹Manchester Medical School, The University of Manchester, Manchester, 9PT, UK ¹⁰²Program in Translational NeuroPsychiatric Genomics, Departments of Neurology & Psychiatry, Brigham and Women's Hospital, Boston, MA 02115, USA ¹⁰³Harvard Medical School, Boston, MA 02115, USA ¹⁰⁴Department of Genes and Environment, Norwegian Institute of Public Health, Oslo, N-0403, Norway ¹⁰⁵Department of Genomics of Common Disease, Imperial College London, London, W12 0NN, UK ¹⁰⁶Department of Clinical Physiology, Tampere University Hospital, Tampere, 33521, Finland ¹⁰⁷Department of Clinical Physiology, University of Tampere, School of Medicine, Tampere, 33014, Finland ¹⁰⁸Public Health, Medical School, University of Split, 21000 Split, Croatia ¹⁰⁹Institute of Social and Preventive Medicine, Lausanne University Hospital (CHUV), Lausanne, 1010, Switzerland ¹¹⁰Neuroepidemiology Section, National Institute on Aging, National Institutes of Health, Bethesda, MD 20892-9205, USA ¹¹¹Amsterdam Brain and Cognition Center, University of Amsterdam, 1018 XA, Amsterdam, The Netherlands ¹¹²Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA 94305-5797, USA ¹¹³LifeLines Cohort Study, University of Groningen, University Medical Center Groningen, Groningen, 9713 BZ, The Netherlands ¹¹⁴Estonian Genome Center, University of Tartu, Tartu, 51010, Estonia ¹¹⁵Medical Genetics Section, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK ¹¹⁶Department of Internal Medicine, Internal Medicine, Lausanne University Hospital (CHUV), Lausanne, 1011, Switzerland ¹¹⁷Tema BV, 2131 HE Hoofddorp, The Netherlands ¹¹⁸Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, 85764, Germany ¹¹⁹Molecular Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4029, Australia ¹²⁰Institute of Health and

Biomedical Innovation, Queensland Institute of Technology, Brisbane, QLD 4059, Australia ¹²¹Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA 02114, USA ¹²²The Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA ¹²³Psychiatric & Neurodevelopmental Genetics Unit, Department of Psychiatry, Massachusetts General Hospital, Boston, MA 02114, USA ¹²⁴Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, 00014, Finland ¹²⁵Department of Neurology, Massachusetts General Hospital, Boston, MA 02114, USA ¹²⁶Medical Genetics, Institute for Maternal and Child Health IRCCS “Burlo Garofolo”, Trieste, 34100, Italy ¹²⁷Social Impact, Arlington, VA 22201, USA ¹²⁸Department of Economics, University of Minnesota Twin Cities, Minneapolis, MN 55455, USA ¹²⁹Department of Psychiatry and Behavioral Sciences, NorthShore University HealthSystem, Evanston, IL 60201-3137, USA ¹³⁰Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, IL 60637, USA ¹³¹Public Health Genomics Unit, National Institute for Health and Welfare, Helsinki 00300, Finland ¹³²Research Unit for Genetic Epidemiology, Institute of Molecular Biology and Biochemistry, Center of Molecular Medicine, General Hospital and Medical University, Graz, Graz, 8010, Austria ¹³³Information Based Medicine Stream, Hunter Medical Research Institute, New Lambton, NSW 2305, Australia ¹³⁴Medical Research Institute, University of Dundee, Dundee, DD1 9SY, UK ¹³⁵Research Unit Hypertension and Cardiovascular Epidemiology, Department of Cardiovascular Science, University of Leuven, Leuven, 3000, Belgium ¹³⁶R&D VitaK Group, Maastricht University, Maastricht, 6229 EV, The Netherlands ¹³⁷Institute of Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, 85764, Germany ¹³⁸Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig Maximilians-Universität, Munich, 81377, Germany ¹³⁹Department of Geriatrics, Florida State University College of Medicine, Tallahassee, FL 32306, USA ¹⁴⁰Department of Health Sciences and Genetics, University of Leicester, Leicester, LE1 7RH, UK ¹⁴¹Department of Internal Medicine, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands ¹⁴²Research Center for Group Dynamics, Institute for Social Research, University of Michigan, Ann Arbor, MI 48104, USA ¹⁴³Platform for Genome Analytics, Institutes of Neurogenetics & Integrative and Experimental Genomics, University of Lübeck, Lübeck, 23562, Germany ¹⁴⁴Neuroepidemiology and Ageing Research Unit, School of Public Health, Faculty of Medicine, The Imperial College of Science, Technology and Medicine, London SW7 2AZ, UK ¹⁴⁵Department of Health Sciences, Community & Occupational Medicine, University of Groningen, University Medical Center Groningen, Groningen, 9713 AV, The Netherlands ¹⁴⁶Department of Psychology, Union College, Schenectady, NY 12308, USA ¹⁴⁷Istituto di Ricerca Genetica e Biomedica (IRGB), Consiglio Nazionale delle Ricerche, c/o Cittadella Universitaria di Monserrato, Monserrato, Cagliari, 9042, Italy ¹⁴⁸Institute of Biomedical Technologies, Italian National Research Council, Segrate (Milano), 20090, Italy ¹⁴⁹Department of General Practice and Primary Health Care, University

of Helsinki, Helsinki, 00014, Finland ¹⁵⁰Departments of Human Genetics and Psychiatry, Donders Centre for Neuroscience, Nijmegen, 6500 HB, The Netherlands ¹⁵¹Sidra, Experimental Genetics Division, Sidra, Doha 26999, Qatar ¹⁵²Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, 17475, Germany ¹⁵³Department of Psychiatry and Psychotherapy, HELIOS-Hospital Stralsund, Stralsund, 18437, Germany ¹⁵⁴Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, 3062 PA, The Netherlands ¹⁵⁵Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen. 9700 RB, The Netherlands ¹⁵⁶Durrer Center for Cardiogenetic Research, ICIN-Netherlands Heart Institute, Utrecht, 1105 AZ, The Netherlands ¹⁵⁷Generation Scotland, Centre for Genomics and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK ¹⁵⁸Centre for Population Health Research, School of Health Sciences and Sansom Institute, University of South Australia, SA5000, Adelaide, Australia ¹⁵⁹South Australian Health and Medical Research Institute, Adelaide, SA5000, Australia ¹⁶⁰Population, Policy and Practice, UCL Institute of Child Health, London, WC1N 1EH, UK ¹⁶¹Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment & Health, School of Public Health, Imperial College London, London, W2 1PG, UK ¹⁶²Center for Life Course Epidemiology, Faculty of Medicine, University of Oulu, Oulu, FI-90014, Finland ¹⁶³Unit of Primary Care, Oulu University Hospital, Oulu, 90029 OYS, Finland ¹⁶⁴Biocenter Oulu, University of Oulu, FI-90014 Oulu, Finland ¹⁶⁵Fimlab Laboratories, Tampere, 33520, Finland ¹⁶⁶Department of Clinical Chemistry, University of Tampere, School of Medicine, Tampere, 33014, Finland ¹⁶⁷Economics, NYU Shanghai, 200122, Pudong, China ¹⁶⁸Policy Studies, Queen's University, Kingston, K7L3N6, Canada ¹⁶⁹Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4029, Australia ¹⁷⁰Institute of Molecular and Cell Biology, University of Tartu, Tartu, 51010, Estonia ¹⁷¹Centre for Clinical and Cognitive Neuroscience, Institute Brain Behaviour and Mental Health, Salford Royal Hospital, Manchester, M6 8HD, UK ¹⁷²Manchester Institute Collaborative Research in Ageing, University of Manchester, Manchester, M13 9PL, UK ¹⁷³Faculty of Medicine, University of Split, Croatia, Split 21000, Croatia ¹⁷⁴Department of Clinical Genetics, VU Medical Centre, Amsterdam, 1081 HV, The Netherlands ¹⁷⁵Institute of Preventive Medicine, Bispebjerg and Frederiksberg Hospitals, The Capital Region, Frederiksberg, 2000, Denmark ¹⁷⁶Montpellier Business School, Montpellier, 34080, France ¹⁷⁷Panteia, Zoetermeer, 2715 CA, The Netherlands ¹⁷⁸Department of Psychiatry, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands ¹⁷⁹Department of Child and Adolescent Psychiatry, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands ¹⁸⁰Department of Internal Medicine, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands ¹⁸¹Department of Sociology, New York University, New York, NY 10012, USA ¹⁸²School of Medicine, New York University, NY 10016, New York, USA ¹⁸³Bioethics Program, Union Graduate College - Icahn School of Medicine at Mount Sinai, Schenectady, NY 12308, USA ¹⁸⁴The University of Queensland Diamantina Institute, The

Translational Research Institute, Brisbane, QLD 4102, Australia ¹⁸⁵Department of Economics, Stockholm School of Economics, Stockholm, 113 83, Sweden
¹⁸⁶Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
¹⁸⁷Research Institute for Industrial Economics, Stockholm, 10215, Sweden
¹⁸⁸Department of Economics, Cornell University, Ithaca, NY 14853, USA ¹⁸⁹Center for Economic and Social Research, University of Southern California, Los Angeles, CA 90089-3332, USA

Acknowledgments

This research was carried out under the auspices of the Social Science Genetic Association Consortium (SSGAC). The SSGAC seeks to facilitate studies that investigate the influence of genes on human behavior, well-being, and social-scientific outcomes using large genome-wide association study meta-analyses. The SSGAC also provides opportunities for replication and promotes the collection of accurately measured, harmonized phenotypes across cohorts. The SSGAC operates as a working group within the CHARGE consortium. This research has also been conducted using the UK Biobank Resource. This study was supported by funding from the Ragnar Söderberg Foundation (E9/11), the Swedish Research Council (421-2013-1061), The Jan Wallander and Tom Hedelius Foundation, an ERC Consolidator Grant (647648 EdGe), the Pershing Square Fund of the Foundations of Human Behavior, and the NIA/NIH through grants P01-AG005842, P01-AG005842-20S2, P30-AG012810, and T32-AG000186-23 to NBER, and R01-AG042568 to USC. We thank Samantha Cunningham, Nishanth Galla and Justin Rashtian for research assistance. A full list of acknowledgments is provided in the supplementary materials.

References

1. Rietveld CA, et al. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science*. 2013; 340:1467–1471. [PubMed: 23722424]
2. Rietveld CA, et al. Replicability and robustness of GWAS for behavioral traits. *Psychol Sci*. 2014; 25:1975–1986. [PubMed: 25287667]
3. Yang J, et al. Genomic inflation factors under polygenic inheritance. *Eur J Hum Genet*. 2011; 19:807–812. [PubMed: 21407268]
4. Bulik-Sullivan BK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015; 47:291–295. [PubMed: 25642630]
5. Sudlow C, et al. UK Biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015; 12:e1001779. [PubMed: 25826379]
6. Fowler T, Zammit S, Owen MJ, Rasmussen F. A population-based study of shared genetic variation between premorbid IQ and psychosis among male twin pairs and sibling pairs from Sweden. *Arch Gen Psychiatry*. 2012; 69:460–466. [PubMed: 22566578]
7. Tambs K, Sundet JM, Magnus P, Berg K. Genetic and environmental contributions to the covariance between occupational status, educational attainment, and IQ: a study of twins. *Behav Genet*. 1989; 19:209–222. [PubMed: 2719624]
8. Thompson LA, Determan DK, Plomin R. Associations between cognitive abilities and scholastic achievement: Genetic overlap but environmental differences. *Psychol Sci*. 1991; 2:158–165.
9. Bulik-Sullivan B, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet*. 2015; 47:1236–1241. [PubMed: 26414676]
10. Ardlie KG, et al. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science*. 2015; 348:648–660. [PubMed: 25954001]
11. Pers TH, et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun*. 2015; 6:5890. [PubMed: 25597830]
12. Allen Institute for Brain Science. BrainSpan atlas of the developing human brain. 2015. <<http://www.brainspan.org/>>
13. Purcell SM, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009; 460:748–752. [PubMed: 19571811]

14. Krapohl E, et al. The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. *Proc Natl Acad Sci USA*. 2014; 111:15273–15278. [PubMed: 25288728]
15. Branigan AR, McCallum KJ, Freese J. Variation in the heritability of educational attainment: An international meta-analysis. *Northwest Univ Inst Policy Res Work Pap*. 2013; 92:109–140.
16. Heath AC, et al. Education policy and the heritability of educational attainment. *Nature*. 1985; 314:734–736. [PubMed: 4039415]

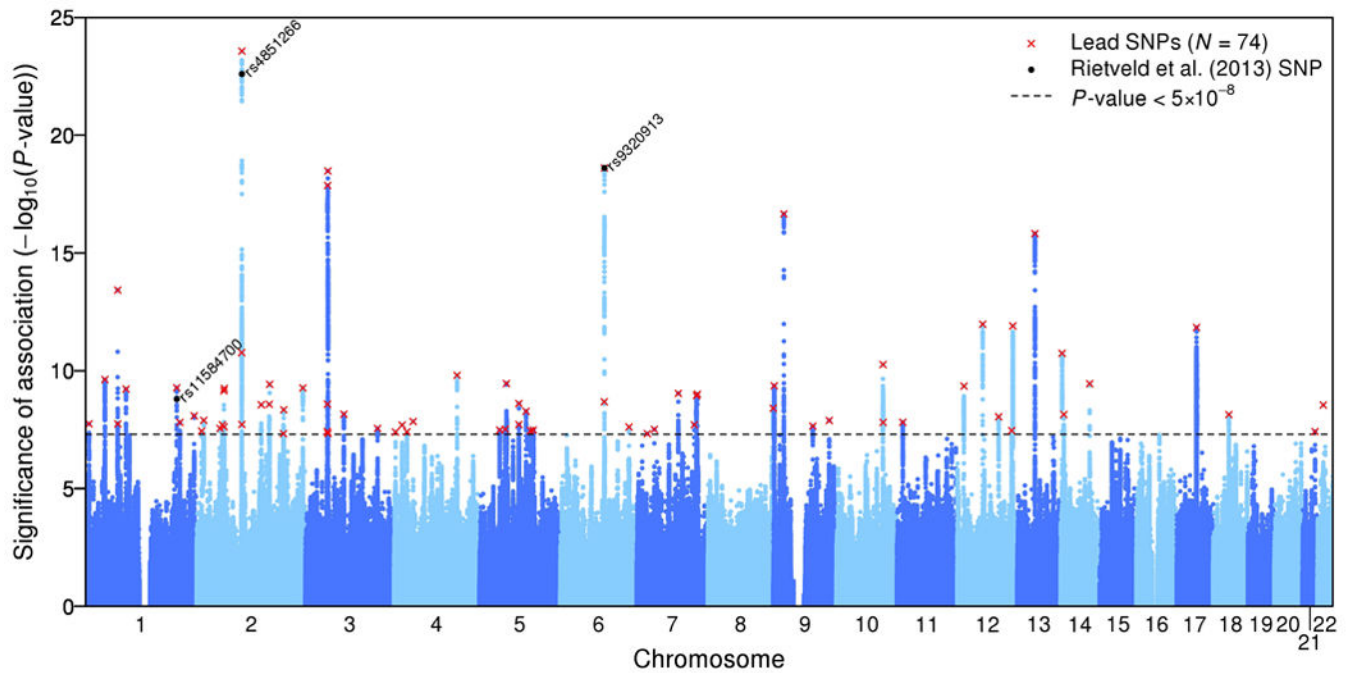


Figure 1. Manhattan plot for *EduYears* associations ($N = 293,723$)

The x -axis is chromosomal position, and the y -axis is the significance on a $-\log_{10}$ scale. The black line shows the genome-wide significance level (5×10^{-8}). The red x 's are the 74 approximately independent genome-wide significant associations (“lead SNPs”). The black dots labeled with rs numbers are the 3 Rietveld et al.¹ SNPs.

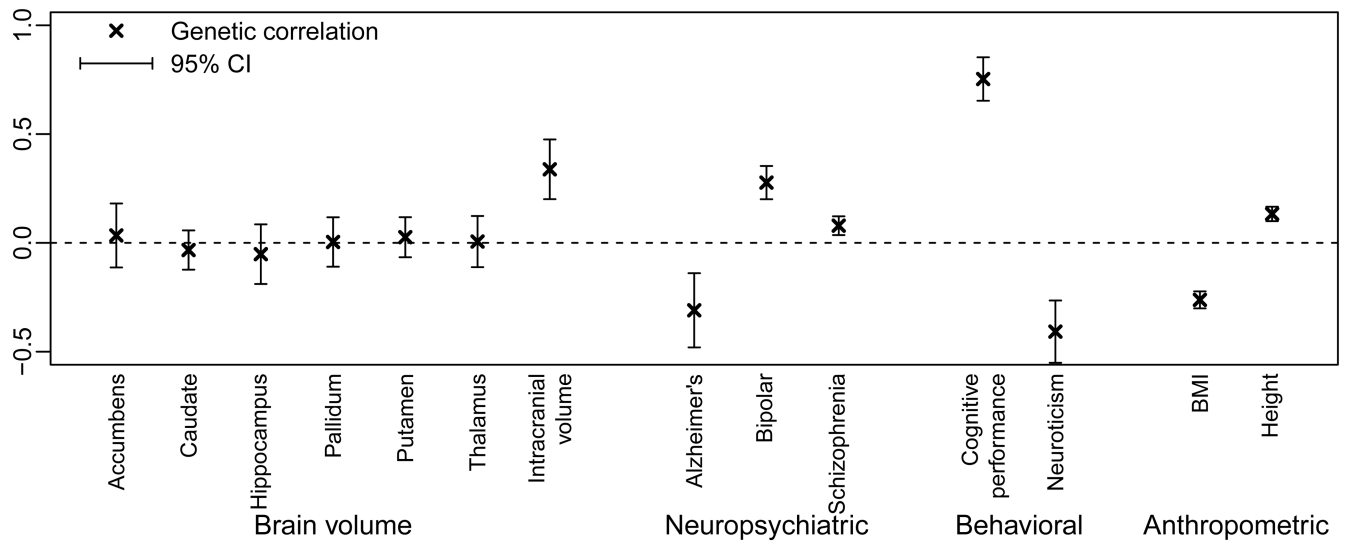


Figure 2. Genetic correlations between *EduYears* and other traits

Results from bivariate Linkage-Disequilibrium (LD) Score regressions⁹: estimates of genetic correlation with brain volume, neuropsychiatric, behavioral, and anthropometric phenotypes using published GWAS summary statistics. The error bars show the 95% confidence intervals.

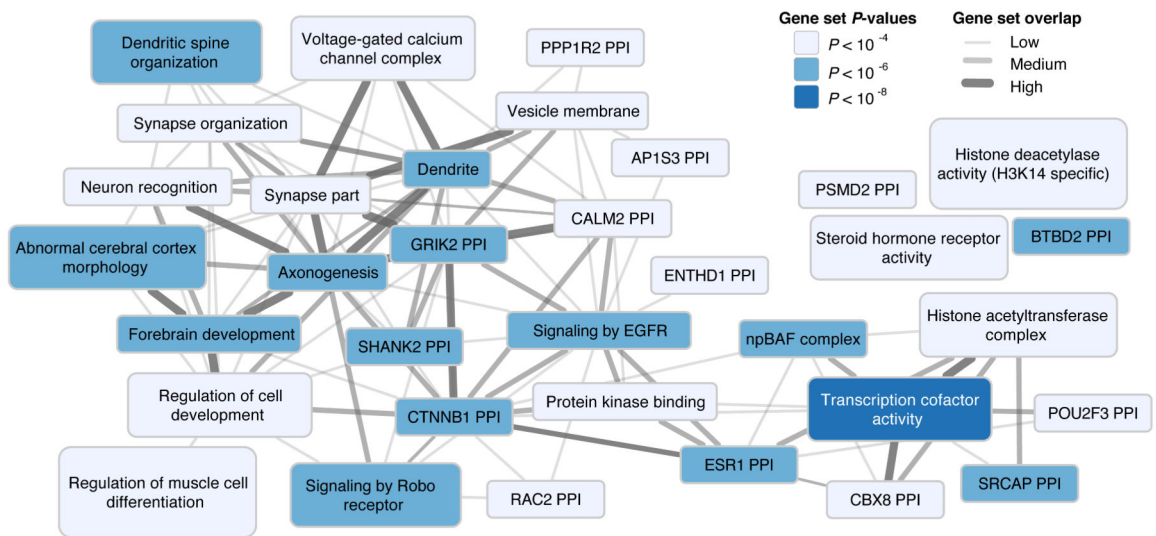


Figure 3. Overview of biological annotation

34 clusters of significantly enriched gene sets. Each cluster is named after one of its member gene sets. The color represents the P -value of the member set exhibiting the most statistically significant enrichment. Overlap between pairs of clusters is represented by an edge. Edge width represents the Pearson correlation ρ between the two vectors of gene membership scores ($\rho < 0.3$, no edge; $0.3 \leq \rho < 0.5$, thin edge; $0.5 \leq \rho < 0.7$, intermediate edge; $\rho \geq 0.7$, thick edge), where each cluster's vector is the vector for the gene set after which the cluster is named.

Table 1
Selected candidate genes implicated by bioinformatics analyses

Fifteen candidate genes implicated most consistently across various analyses. To assemble this list, each gene in a DEPICT-defined locus (Supplementary Information section 4.5) was assigned a score equal to the number of criteria it satisfies out of ten (see Supplementary Table 4.1 for details). The DEPICT prioritization *P*-value was used as the tiebreaker. “SNP”: the SNP in the gene’s locus with the lowest *P*-value in the *EduYears* meta-analysis. “Syndromic”: which, if any, of three neuropsychiatric disorders have been linked to *de novo* mutations in the gene (Supplementary Information section 4.6). “Top-ranking gene sets”: DEPICT reconstituted gene sets of which the gene is a top-20 member (Supplementary Table 4.5.1). The three most significant gene sets are shown if more than three are available. ID, intellectual disability; ASD, autism spectrum disorder; SCZ, schizophrenia.

Gene	SNP	Syndromic	Score	Top-ranking gene sets
<i>TBR1</i>	rs4500960	ID, ASD	6	Developmental biology, decreased brain size, abnormal cerebral cortex morphology
<i>MEF2C</i>	rs7277187	ID, ASD	5	ErbB signaling pathway, abnormal sternum ossification, regulation of muscle cell differentiation
<i>ZSWIM6</i>	rs61160187	–	5	Transcription factor binding, negative regulation of signal transduction, PI3K events in ErbB4 signaling
<i>BCL11A</i>	rs2457660	ASD	5	Dendritic spine organization, abnormal hippocampal mossy fiber morphology, SWI/SNF-type complex
<i>CELSR3</i>	rs11712056	SCZ	5	Dendrite morphogenesis, dendrite development, abnormal hippocampal mossy fiber morphology
<i>MAPT</i>	rs192818565	ID	5	Dendrite morphogenesis, abnormal hippocampal mossy fiber morphology, abnormal axon guidance
<i>SBNQ1</i>	rs7306755	SCZ	5	Protein serine/threonine phosphatase complex
<i>NBAS</i>	rs12987662	–	5	–
<i>NBEA</i>	rs9544418	SCZ	4	Developmental biology, signaling by Robo receptor, dendritic shaft
<i>SMARCA2</i>	rs1871109	ID	4	–
<i>MAP4</i>	rs11712056	ASD	4	Developmental biology, signaling by Robo receptor, SWI-SNF-type complex
<i>LINC00461</i>	rs10061788	–	4	Decreased brain size, abnormal cerebral cortex morphology, abnormal hippocampal mossy fiber morphology
<i>POU3F2</i>	rs9320913	–	4	Dendrite morphogenesis, developmental biology, decreased brain size
<i>RAD54L2</i>	rs11712056	SCZ	4	Decreased brain size, SWI/SNF-type complex, nBAF complex
<i>PLK2</i>	rs2964197	–	4	Negative regulation of signal transduction, PI3K events in ErbB4 signaling