

Please cite the Published Version

Okbay, Aysu, Beauchamp, Jonathan P, Fontana, Mark Alan, Lee, James J, Pers, Tune H, Rietveld, Cornelius A, Turley, Patrick, Chen, Guo-Bo, Emilsson, Valur, Meddens, S Fleur W, Oskarsson, Sven, Pickrell, Joseph K, Thom, Kevin, Timshel, Pascal, de Vlaming, Ronald, Abdellaoui, Abdel, Ahluwalia, Tarunveer S, Bacelis, Jonas, Baumbach, Clemens, Bjornsdottir, Gyda, Brandsma, Johannes H, Pina Concas, Maria, Derringer, Jaime, Furlotte, Nicholas A, Galesloot, Tessel E, Girotto, Giorgia, Gupta, Richa, Hall, Leanne M, Harris, Sarah E, Hofer, Edith, Horikoshi, Momoko, Huffman, Jennifer E, Kaasik, Kadri, Kalafati, Ioanna P, Karlsson, Robert, Kong, Augustine, Lahti, Jari, Lee, Sven J van der, deLeeuw, Christiaan, Lind, Penelope A, Lindgren, Karl-Oskar, Liu, Tian, Mangino, Massimo, Marten, Jonathan, Mihailov, Evelin, Miller, Michael B, van der Most, Peter J, Oldmeadow, Christopher, Payton, Antony, Pervjakova, Natalia, Peyrot, Wouter J, Qian, Yong, Raitakari, Olli, Rueedi, Rico, Salvi, Erika, Schmidt, Börge, Schraut, Katharina E, Shi, Jianxin, Smith, Albert V, Poot, Raymond A, St Pourcain, Beate, Teumer, Alexander, Thorleifsson, Gudmar, Verweij, Niek, Vuckovic, Dragana, Wellmann, Juergen, Westra, Harm-Jan, Yang, Jingyun, Zhao, Wei, Zhu, Zhihong, Alizadeh, Behrooz Z, Amin, Najaf, Bakshi, Andrew, Baumeister, Sebastian E, Biino, Ginevra, Bønnelykke, Klaus, Boyle, Patricia A, Campbell, Harry, Cappuccio, Francesco P, Davies, Gail, De Neve, Jan-Emmanuel, Deloukas, Panos, Demuth, Ilja, Ding, Jun, Eibich, Peter, Eisele, Lewin, Eklund, Niina, Evans, David M, Faul, Jessica D, Feitosa, Mary F, Forstner, Andreas J, Gandin, Ilaria, Gunnarsson, Bjarni, Halldórsson, Bjarni V, Harris, Tamara B, Heath, Andrew C, Hocking, Lynne J, Holliday, Elizabeth G, Homuth, Georg, Horan, Michael A, Hottenga, Jouke-Jan, de Jager, Philip L, Joshi, Peter K, Jugessur, Astanand, Kaakinen, Marika A, Kähönen, Mika, Kanoni, Stavroula, Keltigangas-Järvinen, Liisa, Kiemenev, Lambertus ALM, Kolcic, Ivana, Koskinen, Seppo, Kraja, Aldi T, Kroh, Martin, Kutalik, Zoltan, Latvala, Antti, Launer, Lenore J, Lebreton, Maël P, Levinson, Douglas F, Lichtenstein, Paul, Lichtner, Peter, Liewald, David CM, Cohort Study, LifeLines, Loukola, Anu, Madden, Pamela A, Mägi, Reedik, Mäki-Opas, Tomi, Marioni, Riccardo E, Marques-Vidal, Pedro, Meddens, Gerardus A, McMahon, George, Meisinger, Christa, Meitinger, Thomas, Milaneschi, Yusplriti, Milani, Lili, Montgomery, Grant W, Myhre, Ronny, Nelson, Christopher P, Nyholt, Dale R, Ollier, William ER, Palotie, Aarno, Paternoster, Lavinia, Pedersen, Nancy L, Petrovic, Katja E, Porteous, David J, Rääkkönen, Katri, Ring, Susan M, Robino, Antonietta, Rostapshova, Olga, Rudan, Igor, Rustichini, Aldo, Salomaa, Veikko, Sanders, Alan R, Sarin, Antti-Pekka, Schmidt, Helena, Scott, Rodney J, Smith, Blair H, Smith, Jennifer A, Staessen, Jan A, Steinhagen-Thiessen, Elisabeth, Strauch, Konstantin, Terracciano, Antonio, Tobin, Martin D, Ulivi, Sheila, Vaccargiu, Simona, Quaye, Lydia, van Rooij, Frank JA, Venturini, Cristina, Vinkhuyzen, Anna AE, Völker, Uwe, Völzke, Henry, Vonk, Judith M, Vozzi, Diego, Waage, Johannes, Ware, Erin B, Willemsen, Goncke, Attia, John R, Bennett, David A, Berger, Klaus, Bertram, Lars, Bisgaard, Hans, Boomsma, Dorret I, Borecki, Ingrid B, Bültmann, Ute, Chabris, Christopher F, Cucca, Francesco, Cusi, Daniele, Deary, Ian J, Dedoussis, George V, van Duijn, Cornelia M, Eriksson, Johan G, Franke, Barbara, Franke, Lude, Gasparini, Paolo, Gejman, Pablo V, Gieger, Christian, Grabe, Hans-Jörgen, Gratten, Jacob, Groenen, Patrick JF, Gudnason, Vilmundur, van der Harst, Pim, Hayward, Caroline, Hinds, David A, Hoffmann, Wolfgang, Hyppönen, Elina, Iacono, William G, Jacobsson, Bo, Järvelin, Marjo-Riitta, Jöckel, Karl-Heinz, Kaprio, Jaakko, Kardia, Sharon LR, Lehtimäki, Terho, Lehrer, Steven F, Magnusson, Patrik KE, Martin, Nicholas G, McGue, Matt, Metspalu, Andres, Pendleton, Neil, Penninx, Brenda WJH, Perola, Markus, Pirastu, Nicola, Pirastu, Mario, Polasek, Ozren, Posthuma, Danielle, Power, Christine, Province, Michael A, Samani, Nilesh J, Schlessinger, David, Schmidt, Reinhold, Sørensen, Thorkild IA, Spector, Tim D, Stefansson, Kari, Thorsteinsdottir, Unnur, Thurik, A Roy, Timpson, Nicholas J, Tiemeier, Henning, Tung, Joyce Y, Uitterlinden, André G, Vitart, Veronique, Vollenweider, Peter, Weir, David R, Wilson, James F, Wright, Alan F, Conley, Dalton C, Krueger, Robert F, Davey Smith, George, Hofman, Albert, Laibson, David I, Medland, Sarah E, Meyer, Michelle N, Yang, Jian, Johannesson, Magnus, Visscher, Peter M, Esko, Tõnu, Koellinger, Philipp D, Cesarini, David and Benjamin, Daniel J (2016) Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, 533 (7604). pp. 539-542. ISSN 0028-0836

DOI: <https://doi.org/10.1038/nature17671>

Publisher: Nature Research

Version: Accepted Version

Downloaded from: <https://e-space.mmu.ac.uk/625249/>

Usage rights: © In Copyright

Additional Information: This is an Accepted Manuscript of an article published in Nature.

Enquiries:

If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)

1 **Title: Genome-wide association study identifies 74 loci**
2 **associated with educational attainment**

3
4 **Authors:** All authors and their affiliations appear at the end of the paper

5
6 **Summary:** Educational attainment (EA) is strongly influenced by social and other
7 environmental factors, but genetic factors are also estimated to account for at least 20% of the
8 variation across individuals¹. We report the results of a genome-wide association study
9 (GWAS) for EA that extends our earlier discovery sample^{1,2} of 101,069 individuals to 293,723
10 individuals, and a replication in an independent sample of 111,349 individuals from the UK
11 Biobank. We now identify 74 genome-wide significant loci associated with number of years of
12 schooling completed. Single-nucleotide polymorphisms (SNPs) associated with educational
13 attainment are disproportionately found in genomic regions regulating gene expression in the
14 fetal brain. Candidate genes are preferentially expressed in neural tissue, especially during the
15 prenatal period, and enriched for biological pathways involved in neural development. Our
16 findings demonstrate that, even for a behavioral phenotype that is mostly environmentally
17 determined, a well-powered GWAS identifies replicable associated genetic variants that
18 suggest biologically relevant pathways. Because EA is measured in large numbers of
19 individuals, it will continue to be useful as a proxy phenotype in efforts to characterize the
20 genetic influences of related phenotypes, including cognition and neuropsychiatric disease.

21
22 **Main Text:**

23 We study educational attainment (EA), which is measured in all main analyses as the number
24 of years of schooling completed (*EduYears*, $N = 293,723$, mean = 14.33, SD = 3.61;
25 Supplementary Information sections 1.1-1.2). All genome-wide association studies (GWAS)
26 were performed at the cohort level in samples restricted to individuals of European descent
27 whose EA was assessed at or above age 30. A uniform set of quality-control (QC) procedures

1 was applied to the cohort-level summary statistics. In our GWAS meta-analysis of ~9.3M SNPs
2 from the 1000 Genomes Project, we used sample-size weighting and applied a single round of
3 genomic control at the cohort level.

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

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

14 To quantify the amount of population stratification in the GWAS estimates that remains even
15 after the stringent controls used by the cohorts (Supplementary Information section 1.4), we
16 used LD Score regression⁴. The regression results indicate that ~8% of the observed inflation
17 in the mean χ^2 is due to bias rather than polygenic signal (Extended Data Fig. 3a), suggesting
18 that stratification effects are small in magnitude. We also found evidence that the genetic
19 association signals taken as a whole replicate reliably in several within-family analyses
20 (Supplementary Information section 2 and Extended Data Fig. 3b).

21 To further test the robustness of our findings, we examined the within-sample and out-of-
22 sample replicability of SNPs reaching genome-wide significance (Supplementary
23 Information sections 1.7-1.8). We found that SNPs identified in the previous EA meta-analysis
24 replicated in the new cohorts included here, and conversely, that SNPs reaching genome-wide

1 significance in the new cohorts replicated in the old cohorts. For the out-of-sample replication
2 analyses of our 74 lead SNPs, we used the interim release of the U.K. Biobank⁵ (UKB) ($N =$
3 111,349). As shown in Extended Data Fig. 4, 72 out of the 74 lead SNPs have a consistent sign
4 ($P = 1.47 \times 10^{-19}$), 52 are significant at the 5% level ($P = 2.68 \times 10^{-50}$), and 7 reach genome-wide
5 significance in the U.K. Biobank dataset ($P = 1.41 \times 10^{-42}$). For comparison, the corresponding
6 expected numbers, assuming each SNP's true effect size is its estimated effect adjusted for the
7 winner's curse, are 71.4, 40.3, and 0.6. (Supplementary Information section 1.8.2). We also
8 find out-of-sample replicability of our overall GWAS results: the genetic correlation between
9 *EduYears* in our meta-analysis sample and in the UKB data is 0.95 (s.e. = 0.021; Supplementary
10 Table 1.14).

11 It is known that EA, cognitive performance, and many neuropsychiatric phenotypes are
12 phenotypically correlated, and several studies of twins find that the phenotypic correlations
13 partly reflect genetic overlap⁶⁻⁸ (Supplementary Information section 3.3.4). Here, we
14 investigate genetic correlation using our GWAS results for *EduYears* and published GWAS
15 results for 14 other phenotypes, using bivariate Linkage-Disequilibrium (LD) Score
16 regression⁹. First, we estimated genetic correlations with *EduYears*. As shown in Fig. 2, on
17 average, alleles associated with greater EA are also associated with increased cognitive
18 performance ($P = 9.9 \times 10^{-50}$) and intracranial volume ($P = 1.2 \times 10^{-6}$), increased risk of bipolar
19 disorder ($P = 7 \times 10^{-13}$), decreased risk of Alzheimer's ($P = 4 \times 10^{-4}$), and lower neuroticism (P
20 = 2.8×10^{-8}). We also found positive, statistically significant, but very small, genetic
21 correlations with height ($P = 5.2 \times 10^{-15}$) and risk of schizophrenia ($P = 3.2 \times 10^{-4}$).

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

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

7 To consider potential biological pathways, we first tested whether SNPs in particular regions
8 of the genome are implicated by our GWAS results. Unlike what has been found for other
9 phenotypes, SNPs in regions that are DNase I hypersensitive in the fetal brain are more likely
10 to be associated with *EduYears* by a factor of ~5 (95% confidence interval 2.89–7.07; Extended
11 Data Fig. 7). Moreover, the 15% of SNPs residing in regions associated with histones marked
12 in the central nervous system (CNS) explain 44% of the heritable variation (Extended Data Fig.
13 8a and Supplementary Table 4.4.2). This enrichment factor of ~3 for CNS ($P = 2.48 \times 10^{-16}$) is
14 greater than that of any of the other nine tissue categories in this analysis.

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

22 To investigate possible functions of the candidate genes from the GWAS associated loci, we
23 examined the extent of their overlap with groups of genes (“gene sets”) whose products are
24 known or predicted to participate in a common biological process¹¹. We found 283 gene sets
25 significantly enriched by the candidate genes identified in our GWAS (FDR < 0.05;

1 Supplementary Table 4.5.1). To facilitate interpretation, we used a standard procedure¹¹ to
2 group the 283 gene sets into “clusters” defined by degree of gene overlap. The resulting 34
3 clusters, shown in Fig. 3, paint a coherent picture, with many clusters corresponding to stages
4 of neural development: the proliferation of neural progenitor cells and their specialization (the
5 *cluster npBAF complex*), the migration of new neurons to the different layers of the cortex
6 (*forebrain development, abnormal cerebral cortex morphology*), the projection of axons from
7 neurons to their signaling targets (*axonogenesis, signaling by Robo receptor*), the sprouting of
8 dendrites and their spines (*dendrite, dendritic spine organization*), and neuronal signaling
9 and synaptic plasticity throughout the lifespan (*voltage-gated calcium channel complex,*
10 *synapse part, synapse organization*).

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

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

20 We constructed polygenic scores¹³ to assess the joint predictive power afforded by the GWAS
21 results (Supplementary Information section 5.2). Across our two holdout samples, the mean
22 predictive power of a polygenic score constructed from all measured SNPs is 3.2% ($P =$
23 1.18×10^{-39} ; Supplementary Table 5.2 and Supplementary Information section 5).

24 Studies of genetic analyses of behavioral phenotypes have been prone to misinterpretation,
25 such as characterizing identified associated variants as “genes for education.” Such

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

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

18

19 **Methods:** All methods are described in the Supplementary Information.

20

21

22 **References:**

- 23 1. Rietveld, C. A. *et al.* GWAS of 126,559 individuals identifies genetic variants
24 associated with educational attainment. *Science* **340**, 1467–1471 (2013).
- 25 2. Rietveld, C. A. *et al.* Replicability and robustness of GWAS for behavioral traits.
26 *Psychol. Sci.* **25**, 1975–1986 (2014).
- 27 3. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *Eur. J. Hum.*
28 *Genet.* **19**, 807–812 (2011).

- 1 4. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from
2 polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
- 3 5. Sudlow, C. *et al.* UK Biobank: An open access resource for identifying the causes of a
4 wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779
5 (2015).
- 6 6. Fowler, T., Zammit, S., Owen, M. J. & Rasmussen, F. A population-based study of
7 shared genetic variation between premorbid IQ and psychosis among male twin pairs
8 and sibling pairs from Sweden. *Arch. Gen. Psychiatry* **69**, 460–466 (2012).
- 9 7. Tambs, K., Sundet, J. M., Magnus, P. & Berg, K. Genetic and environmental
10 contributions to the covariance between occupational status, educational attainment,
11 and IQ: a study of twins. *Behav. Genet.* **19**, 209–222 (1989).
- 12 8. Thompson, L. A., Detterman, D. K. & Plomin, R. Associations between cognitive
13 abilities and scholastic achievement: Genetic overlap but environmental differences.
14 *Psychol. Sci.* **2**, 158–165 (1991).
- 15 9. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and
16 traits. *Nat. Genet.* **47**, 1236–1241 (2015).
- 17 10. Ardlie, K. G. *et al.* The Genotype-Tissue Expression (GTEx) pilot analysis:
18 Multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
- 19 11. Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using
20 predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
- 21 12. Allen Institute for Brain Science. BrainSpan atlas of the developing human brain.
22 (2015). at <<http://www.brainspan.org/>>
- 23 13. Purcell, S. M. *et al.* Common polygenic variation contributes to risk of schizophrenia
24 and bipolar disorder. *Nature* **460**, 748–752 (2009).
- 25 14. Krapohl, E. *et al.* The high heritability of educational achievement reflects many
26 genetically influenced traits, not just intelligence. *Proc. Natl. Acad. Sci. U. S. A.* **111**,
27 15273–15278 (2014).
- 28 15. Branigan, A. R., McCallum, K. J. & Freese, J. Variation in the heritability of
29 educational attainment: An international meta-analysis. *Northwest. Univ. Inst. Policy*
30 *Res. Work. Pap.* **92**, 109–140 (2013).
- 31 16. Heath, A. C. *et al.* Education policy and the heritability of educational attainment.
32 *Nature* **314**, 734–736 (1985).

33
34

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

37

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

15
16 **Author contributions** Study Design and Management: D.J.B., D.C., T.E., M.J., P.D.K. and
17 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.
18 Stratification: P.T., J.P.B., C.A.R. and J.Y. Genetic Overlap: J.P.B, M.A.F., P.T. Biological
19 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.,
20 G.A.M., T.H.P., J.K.P., P.Timshel, R.A.P., R.d.V. and H.J.W. Prediction and Mediation:
21 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:
22 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.,
23 M.N.M., G.D.S. and P.M.V. All authors contributed to and critically reviewed the manuscript.
24 Authors not listed above contributed to the recruitment, genotyping, or data processing for the

1 contributing components of the meta-analysis. For a full list of author contributions, see
2 Supplementary Information section 8.

3

4 **Author Information** Results can be downloaded from the SSGAC website
5 (<http://ssgac.org/Data.php>). Data for our analyses come from many studies and organizations,
6 some of which are subject to a MTA, and are listed in the Supplementary Information. Reprints
7 and permissions information is available at www.nature.com/reprints. The authors declare no
8 competing financial interests. Correspondence and requests for materials should be addressed
9 to D.J.B. (daniel.benjamin@gmail.com), D.C. (dac12@nyu.edu), P.D.K.
10 (p.d.koellinger@vu.nl), or P.M.V. (peter.visscher@uq.edu.au).

11

1 **Table 1 | Selected candidate genes implicated by bioinformatics analyses.** Fifteen
2 candidate genes implicated most consistently across various analyses. To assemble this list,
3 each gene in a DEPICT-defined locus (Supplementary Information section 4.5) was assigned
4 a score equal to the number of criteria it satisfies out of ten (see Supplementary Table 4.1 for
5 details). The DEPICT prioritization *P*-value was used as the tiebreaker. “SNP”: the SNP in
6 the gene’s locus with the lowest *P*-value in the *EduYears* meta-analysis. “Syndromic”: which,
7 if any, of three neuropsychiatric disorders have been linked to *de novo* mutations in the gene
8 (Supplementary Information section 4.6). “Top-ranking gene sets”: DEPICT reconstituted
9 gene sets of which the gene is a top-20 member (Supplementary Table 4.5.1). The three most
10 significant gene sets are shown if more than three are available. ID, intellectual disability;
11 ASD, autism spectrum disorder; SCZ, schizophrenia.

12

1

| 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>SBNO1</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 |

2

1 **Authors:**

2 Aysu Okbay^{1,2,3,*}, Jonathan P. Beauchamp^{4,*}, Mark A. Fontana^{5,*}, James J. Lee^{6,*}, Tune H.
3 Pers^{7,8,9,10,*}, Cornelius A. Rietveld^{1,2,3,*}, Patrick Turley^{4,*}, Guo-Bo Chen¹¹, Valur
4 Emilsson^{12,13}, S. Fleur W. Meddens^{14,3,15}, Sven Oskarsson¹⁶, Joseph K. Pickrell¹⁷, Kevin
5 Thom¹⁸, Pascal Timshel^{19,8}, Ronald de Vlaming^{1,2,3}, Abdel Abdellaoui²⁰, Tarunveer S.
6 Ahluwalia^{21,9,22}, Jonas Bacelis²³, Clemens Baumbach^{24,25}, Gyda Bjornsdottir⁹⁵, Johannes H.
7 Brandsma²⁶, Maria Pina Concas²⁷, Jaime Derringer²⁸, Nicholas A. Furlotte²⁹, Tessel E.
8 Galesloot³⁰, Giorgia Giotto³¹, Richa Gupta³², Leanne M. Hall^{33,34}, Sarah E. Harris^{35,36}, Edith
9 Hofer^{37,38}, Momoko Horikoshi^{39,40}, Jennifer E. Huffman⁴¹, Kadri Kaasik⁴², Ioanna P.
10 Kalafati⁴³, Robert Karlsson⁴⁴, Augustine Kong⁹⁵, Jari Lahti^{42,45}, Sven J. van der Lee²,
11 Christiaan de Leeuw^{14,46}, Penelope A. Lind⁴⁷, Karl-Oskar Lindgren¹⁶, Tian Liu⁴⁸, Massimo
12 Mangino^{49,50}, Jonathan Marten⁴¹, Evelin Mihailov¹¹⁴, Michael B. Miller⁶, Peter J. van der
13 Most⁵¹, Christopher Oldmeadow^{52,53}, Antony Payton^{54,55}, Natalia Pervjakova^{56,114}, Wouter J.
14 Peyrot⁵⁷, Yong Qian⁵⁸, Olli Raitakari⁵⁹, Rico Rueedi^{60,61}, Erika Salvi⁶², Børge Schmidt⁶³,
15 Katharina E. Schraut⁶⁴, Jianxin Shi⁶⁵, Albert V. Smith^{66,67}, Raymond A. Poot²⁶, Beate St
16 Pourcain^{68,69}, Alexander Teumer⁷⁰, Gudmar Thorleifsson⁹⁵, Niek Verweij⁷¹, Dragana
17 Vuckovic³¹, Juergen Wellmann⁷², Harm-Jan Westra^{73,74,8}, Jingyun Yang^{75,76}, Wei Zhao⁷⁷,
18 Zhihong Zhu¹¹, Behrooz Z. Alizadeh^{51,78}, Najaf Amin², Andrew Bakshi¹¹, Sebastian E.
19 Baumeister^{70,79}, Ginevra Biino⁸⁰, Klaus Bønnelykke²¹, Patricia A. Boyle^{75,81}, Harry
20 Campbell⁶⁴, Francesco P. Cappuccio⁸², Gail Davies^{35,83}, Jan-Emmanuel De Neve⁸⁴, Panos
21 Deloukas^{85,86}, Ilja Demuth^{87,88}, Jun Ding⁵⁸, Peter Eibich^{89,90}, Lewin Eisele⁶³, Niina Eklund⁵⁶,
22 David M. Evans^{68,184}, Jessica D. Faul⁹¹, Mary F. Feitosa⁹², Andreas J. Forstner^{93,94}, Ilaria
23 Gandin³¹, Bjarni Gunnarsson⁹⁵, Bjarni V. Halldórsson^{95,96}, Tamara B. Harris⁹⁷, Andrew C.
24 Heath⁹⁸, Lynne J. Hocking⁹⁹, Elizabeth G. Holliday^{52,53}, Georg Homuth¹⁰⁰, Michael A.
25 Horan¹⁰¹, Jouke-Jan Hottenga²⁰, Philip L. de Jager^{102,103,8}, Peter K. Joshi⁶⁴, Astanand

1 Jugessur¹⁰⁴, Marika A. Kaakinen¹⁰⁵, Mika Kähönen^{106,107}, Stavroula Kanoni⁸⁵, Liisa
2 Keltigangas-Järvinen⁴², Lambertus A.L.M. Kiemeny³⁰, Ivana Kolcic¹⁰⁸, Seppo Koskinen⁵⁶,
3 Aldi T. Kraja⁹², Martin Kroh⁸⁹, Zoltan Kutalik^{109,60,61}, Antti Latvala³², Lenore J. Launer¹¹⁰,
4 Maël P. Lebreton^{15,111}, Douglas F. Levinson¹¹², Paul Lichtenstein⁴⁴, Peter Lichtner¹¹⁸, David
5 C.M. Liewald^{35,83}, LifeLines Cohort Study¹¹³, Anu Loukola³², Pamela A. Madden⁹⁸, Reedik
6 Mägi¹¹⁴, Tomi Mäki-Opas⁵⁶, Riccardo E. Marioni^{35,115,11}, Pedro Marques-Vidal¹¹⁶, Gerardus
7 A. Meddens¹¹⁷, George McMahon⁶⁸, Christa Meisinger²⁵, Thomas Meitinger¹¹⁸, Yusplitri
8 Milaneschi⁵⁷, Lili Milani¹¹⁴, Grant W. Montgomery¹¹⁹, Ronny Myhre¹⁰⁴, Christopher P.
9 Nelson^{33,34}, Dale R. Nyholt^{120,119}, William E.R. Ollier⁵⁴, Aarno Palotie^{121,8,122,123,124,125},
10 Lavinia Paternoster⁶⁸, Nancy L. Pedersen⁴⁴, Katja E. Petrovic³⁷, David J. Porteous³⁶, Katri
11 Räikkönen^{42,45}, Susan M. Ring⁶⁸, Antonietta Robino¹²⁶, Olga Rostapshova^{4,127}, Igor Rudan⁶⁴,
12 Aldo Rustichini¹²⁸, Veikko Salomaa⁵⁶, Alan R. Sanders^{129,130}, Antti-Pekka Sarin^{124,131}, Helena
13 Schmidt^{132,37}, Rodney J. Scott^{133,53}, Blair H. Smith¹³⁴, Jennifer A. Smith⁷⁷, Jan A.
14 Staessen^{135,136}, Elisabeth Steinhausen-Thiessen⁸⁷, Konstantin Strauch^{137,138}, Antonio
15 Terracciano¹³⁹, Martin D. Tobin¹⁴⁰, Sheila Ulivi¹²⁶, Simona Vaccargiu²⁷, Lydia Quaye⁴⁹,
16 Frank J.A. van Rooij^{2,141}, Cristina Venturini^{49,50}, Anna A.E. Vinkhuyzen¹¹, Uwe Völker¹⁰⁰,
17 Henry Völzke⁷⁰, Judith M. Vonk⁵¹, Diego Vozzi¹²⁶, Johannes Waage^{21,22}, Erin B. Ware^{77,142},
18 Gonneke Willemsen²⁰, John R. Attia^{52,53}, David A. Bennett^{75,76}, Klaus Berger⁷¹, Lars
19 Bertram^{143,144}, Hans Bisgaard²¹, Dorret I. Boomsma²⁰, Ingrid B. Borecki⁹², Ute Bultmann¹⁴⁵,
20 Christopher F. Chabris¹⁴⁶, Francesco Cucca¹⁴⁷, Daniele Cusi^{62,148}, Ian J. Deary^{35,83}, George V.
21 Dedoussis⁴³, Cornelia M. van Duijn², Johan G. Eriksson^{149,45}, Barbara Franke¹⁵⁰, Lude
22 Franke¹⁵⁵, Paolo Gasparini^{31,126,151}, Pablo V. Gejman^{129,130}, Christian Gieger²⁴, Hans-Jürgen
23 Grabe^{152,153}, Jacob Gratten¹¹, Patrick J.F. Groenen¹⁵⁴, Vilmundur Gudnason^{12,67}, Pim van der
24 Harst^{71,155,156}, Caroline Hayward^{41,157}, David A. Hinds²⁹, Wolfgang Hoffmann⁷⁰, Elina
25 Hyppönen^{158,159,160}, William G. Iacono⁶, Bo Jacobsson^{23,104}, Marjo-Riitta Järvelin^{161,162,163,164},

1 Karl-Heinz Jöckel⁶³, Jaakko Kaprio^{32,124,56}, Sharon L.R. Kardia⁷⁷, Terho Lehtimäki^{165,166},
2 Steven F. Lehrer^{167,168}, Patrik K.E. Magnusson⁴⁴, Nicholas G. Martin¹⁶⁹, Matt McGue⁶,
3 Andres Metspalu^{114,170}, Neil Pendleton^{171,172}, Brenda W.J.H. Penninx⁵⁷, Markus Perola^{56,114},
4 Nicola Pirastu³¹, Mario Pirastu²⁷, Ozren Polasek^{173,64}, Danielle Posthuma^{14,174}, Christine
5 Power¹⁶⁰, Michael A. Province⁹², Nilesh J. Samani^{33,34}, David Schlessinger⁵⁸, Reinhold
6 Schmidt³⁷, Thorkild I.A. Sørensen^{175,9,68}, Tim D. Spector⁴⁹, Kari Stefansson^{95,67}, Unnur
7 Thorsteinsdottir^{95,67}, A. Roy Thurik^{1,176,177,3}, Nicholas J. Timpson⁶⁸, Henning Tiemeier^{2,178,179},
8 Joyce Y. Tung²⁹, André G. Uitterlinden^{180,2}, Veronique Vitart⁴¹, Peter Vollenweider¹¹⁶, David
9 R. Weir⁹¹, James F. Wilson^{64,41}, Alan F. Wright⁴¹, Dalton C. Conley^{181,182}, Robert F.
10 Krueger⁶, George Davey Smith⁶⁸, Albert Hofman², David I. Laibson⁴, Sarah E. Medland⁴⁷,
11 Michelle N. Meyer¹⁸³, Jian Yang^{11,184}, Magnus Johannesson¹⁸⁵, Peter M. Visscher^{11,184,#},
12 Tõnu Esko^{114,7,186,8,#}, Philipp D. Koellinger^{14,15,3,#}, David Cesarini^{18,187,#}, Daniel J.
13 Benjamin^{188,189,#}

14
15 * These authors contributed equally.

16 # Designed and oversaw the study.

¹ Department of Applied Economics, Erasmus School of Economics, Erasmus University
Rotterdam, 3062 PA, Rotterdam, The Netherlands

² Department of Epidemiology, Erasmus Medical Center, Rotterdam, 3015 GE, The
Netherlands

³ Erasmus University Rotterdam Institute for Behavior and Biology, Rotterdam 3062 PA, The
Netherlands

⁴ Department of Economics, Harvard University, Cambridge, MA 02138, USA

⁵ Department of Economics, University of Michigan, Ann Arbor, MI 48109-1220, USA

⁶ Department of Psychology, University of Minnesota Twin Cities, Minneapolis, MN 55455, USA

⁷ Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, MA 2116, USA

⁸ Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

⁹ The Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, University of Copenhagen, Faculty of Health and Medical Sciences, Copenhagen, 2100, Denmark

¹⁰ Statens Serum Institut, Department of Epidemiology Research, Copenhagen, DK 2300, Denmark

¹¹ Queensland Brain Institute, The University of Queensland, Brisbane, QLD 4072, Australia

¹² Icelandic Heart Association, Kopavogur, 201, Iceland

¹³ Faculty of Pharmaceutical Sciences, University of Iceland, 107 Reykjavík, Iceland

¹⁴ Department of Complex Trait Genetics, VU University, Center for Neurogenomics and Cognitive Research, Amsterdam, 1081 HV, The Netherlands

¹⁵ Amsterdam Business School, University of Amsterdam, Amsterdam, 1018 TV, The Netherlands

¹⁶ Department of Government, Uppsala University, Uppsala, 751 20, Sweden

¹⁷ New York Genome Center, New York, NY 10013, USA

¹⁸ Department of Economics, New York University, New York, NY 10012, USA

¹⁹ Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark Lyngby, 2800, Denmark

²⁰ Department of Biological Psychology, VU University Amsterdam, Amsterdam, 1081 BT, The Netherlands

²¹ COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, 2820, Denmark

²² Steno Diabetes Center, Gentofte, 2820, Denmark

²³ Department of Obstetrics and Gynecology, Institute of Clinical Sciences, Sahlgrenska Academy, Gothenburg, SE 416 85, Sweden

²⁴ Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, 85764, Germany

²⁵ Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, 85764, Germany

²⁶ Department of Cell Biology, Erasmus Medical Center Rotterdam, 3015 CN, The Netherlands

²⁷ Istituto di Ricerca Genetica e Biomedica U.O.S. di Sassari, National Research Council of Italy, Sassari, 07100, Italy

²⁸ Psychology, University of Illinois, IL 61820, Champaign, USA

²⁹ 23andMe, Inc., Mountain View, CA 94041, USA

³⁰ Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, 6500 HB, The Netherlands

³¹ Department of Medical, Surgical and Health Sciences, University of Trieste, Trieste, 34100, Italy

³² Department of Public Health, University of Helsinki, Helsinki, FI-00014, Finland

³³ Department of Cardiovascular Sciences, University of Leicester, Leicester, LE3 9QP, UK

³⁴ NIHR Leicester Cardiovascular Biomedical Research Unit, Glenfield Hospital, Leicester, LE3 9QP, UK

³⁵ Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, EH8 9JZ, UK

-
- ³⁶ Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK
- ³⁷ Department of Neurology, General Hospital and Medical University Graz, Graz, 8036, Austria
- ³⁸ Institute for Medical Informatics, Statistics and Documentation, General Hospital and Medical University Graz, Graz, 8036, Austria
- ³⁹ Oxford Centre for Diabetes, Endocrinology & Metabolism, University of Oxford, Oxford, OX3 7LE, UK
- ⁴⁰ Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK
- ⁴¹ MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK
- ⁴² Institute of Behavioural Sciences, University of Helsinki, Helsinki, FI-00014, Finland
- ⁴³ Nutrition and Dietetics, Health Science and Education, Harokopio University, Athens, 17671, Greece
- ⁴⁴ Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, 171 77, Sweden
- ⁴⁵ Folkhälsan Research Centre, Helsingfors, FI-00014, Finland
- ⁴⁶ Institute for Computing and Information Sciences, Radboud University Nijmegen, Nijmegen, 6525 EC, The Netherlands
- ⁴⁷ Quantitative Genetics, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4029, Australia
- ⁴⁸ Lifespan Psychology, Max Planck Institute for Human Development, Berlin, 14195, Germany
- ⁴⁹ Department of Twin Research and Genetic Epidemiology, King's College London, London, SE1 7EH, UK

⁵⁰ NIHR Biomedical Research Centre, Guy's and St. Thomas' Foundation Trust, London, SE1 7EH, UK

⁵¹ Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, 9700 RB, The Netherlands

⁵² Public Health Stream, Hunter Medical Research Institute, New Lambton, NSW 2305, Australia

⁵³ Faculty of Health and Medicine, University of Newcastle, Newcastle, NSW 2300, Australia

⁵⁴ Centre for Integrated Genomic Medical Research, Institute of Population Health, The University of Manchester, Manchester, M13 9PT, UK

⁵⁵ School of Psychological Sciences, The University of Manchester, Manchester, M13 9PL, UK

⁵⁶ Department of Health, THL-National Institute for Health and Welfare, Helsinki, FI-00271, Finland

⁵⁷ Psychiatry, VU University Medical Center & GGZ inGeest, Amsterdam, 1081 HL, The Netherlands

⁵⁸ Laboratory of Genetics, National Institute on Aging, Baltimore, MD 21224, USA

⁵⁹ Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, 20521, Finland

⁶⁰ Department of Medical Genetics, University of Lausanne, Lausanne, 1005, Switzerland

⁶¹ Swiss Institute of Bioinformatics, Lausanne, 1015, Switzerland

⁶² Department Of Health Sciences, University of Milan, Milano, 20142, Italy

⁶³ Institute for Medical Informatics, Biometry and Epidemiology, University Hospital of Essen, Essen, 45147, Germany

⁶⁴ Centre for Global Health Research, The Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, EH8 9AG, UK

-
- ⁶⁵ Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD 20892-9780, USA
- ⁶⁶ Icelandic Heart Association, Kopavogur, 201, Iceland
- ⁶⁷ Faculty of Medicine, University of Iceland, Reykjavik, 101, Iceland
- ⁶⁸ MRC Integrative Epidemiology Unit, University of Bristol, Bristol, BS8 2BN, UK
- ⁶⁹ School of Oral and Dental Sciences, University of Bristol, Bristol, BS1 2LY, UK
- ⁷⁰ Institute for Community Medicine, University Medicine Greifswald, Greifswald, 17475, Germany
- ⁷¹ Department of Cardiology, University Medical Center Groningen, University of Groningen, Groningen, 9700 RB, The Netherlands
- ⁷² Institute of Epidemiology and Social Medicine, University of Muenster, Muenster, 48149, Germany
- ⁷³ Divisions of Genetics and Rheumatology, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, MA 02115, Boston, USA
- ⁷⁴ Partners Center for Personalized Genetic Medicine, Boston, MA 02115, USA
- ⁷⁵ Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL 60612, USA
- ⁷⁶ Department of Neurological Sciences, Rush University Medical Center, Chicago, IL 60612, USA
- ⁷⁷ Department of Epidemiology, University of Michigan, Ann Arbor, MI 48109, USA
- ⁷⁸ Department of Gastroenterology and Hepatology, University of Groningen, University Medical Center Groningen, Groningen, 9713 GZ, The Netherlands
- ⁷⁹ Institute of Epidemiology and Preventive Medicine, University of Regensburg, Regensburg, D-93053, Germany
- ⁸⁰ Institute of Molecular Genetics, National Research Council of Italy, Pavia, 27100, Italy

-
- ⁸¹ Department of Behavioral Sciences, Rush University Medical Center, Chicago, IL 60612, USA
- ⁸² Warwick Medical School, University of Warwick, Coventry, CV4 7AL, UK
- ⁸³ Department of Psychology, University of Edinburgh, Edinburgh, EH8 9JZ, UK
- ⁸⁴ Saïd Business School, University of Oxford, Oxford, OX1 1HP, UK
- ⁸⁵ William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, EC1M 6BQ, UK
- ⁸⁶ Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah, 21589, Saudi Arabia
- ⁸⁷ The Berlin Aging Study II; Research Group on Geriatrics, Charité – Universitätsmedizin Berlin, Germany, Berlin, 13347, Germany
- ⁸⁸ Institute of Medical and Human Genetics, Charité-Universitätsmedizin, Berlin, Berlin, 13353, Germany
- ⁸⁹ German Socio- Economic Panel Study, DIW Berlin, Berlin, 10117, Germany
- ⁹⁰ Health Economics Research Centre, Nuffield Department of Population Health, University of Oxford, Oxford, OX3 7LF, UK
- ⁹¹ Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI 48109, USA
- ⁹² Department of Genetics, Division of Statistical Genomics, Washington University School of Medicine, St. Louis, MO 63018, USA
- ⁹³ Institute of Human Genetics, University of Bonn, Bonn, 53127, Germany
- ⁹⁴ Department of Genomics, Life and Brain Center, University of Bonn, Bonn, 53127, Germany
- ⁹⁵ deCODE Genetics/Amgen Inc., Reykjavik, IS-101, Iceland

⁹⁶ Institute of Biomedical and Neural Engineering, School of Science and Engineering,
Reykjavik University, Reykjavik 101, Iceland

⁹⁷ Laboratory of Epidemiology, Demography, National Institute on Aging, National Institutes
of Health, Bethesda, MD 20892-9205, United States

⁹⁸ Department of Psychiatry, Washington University School of Medicine, St. Louis, MO
63110, USA

⁹⁹ Division of Applied Health Sciences, University of Aberdeen, Aberdeen, AB25 2ZD, UK

¹⁰⁰ Interfaculty Institute for Genetics and Functional Genomics, University Medicine
Greifswald, Greifswald, 17475, Germany

¹⁰¹ Manchester Medical School, The University of Manchester, Manchester, 9PT, UK

¹⁰² Program in Translational NeuroPsychiatric Genomics, Departments of Neurology &
Psychiatry, Brigham and Women's Hospital, Boston, MA 02115, USA

¹⁰³ Harvard Medical School, Boston, MA 02115, USA

¹⁰⁴ Department of Genes and Environment, Norwegian Institute of Public Health, Oslo, N-
0403, Norway

¹⁰⁵ Department of Genomics of Common Disease, Imperial College London, London, W12
0NN, UK

¹⁰⁶ Department of Clinical Physiology, Tampere University Hospital, Tampere, 33521,
Finland

¹⁰⁷ Department of Clinical Physiology, University of Tampere, School of Medicine, Tampere,
33014, Finland

¹⁰⁸ Public Health, Medical School, University of Split, 21000 Split, Croatia

¹⁰⁹ Institute of Social and Preventive Medicine, Lausanne University Hospital (CHUV),
Lausanne, 1010, Switzerland

¹¹⁰ Neuroepidemiology Section, National Institute on Aging, National Institutes of Health, Bethesda, MD 20892-9205, USA

¹¹¹ Amsterdam Brain and Cognition Center, University of Amsterdam, 1018 XA, Amsterdam, The Netherlands

¹¹² Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA 94305-5797, USA

¹¹³ LifeLines Cohort Study, University of Groningen, University Medical Center Groningen, Groningen, 9713 BZ, The Netherlands

¹¹⁴ Estonian Genome Center, University of Tartu, Tartu, 51010, Estonia

¹¹⁵ Medical Genetics Section, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK

¹¹⁶ Department of Internal Medicine, Internal Medicine, Lausanne University Hospital (CHUV), Lausanne, 1011, Switzerland

¹¹⁷ Tema BV, 2131 HE Hoofddorp, The Netherlands

¹¹⁸ Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, 85764, Germany

¹¹⁹ Molecular Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4029, Australia

¹²⁰ Institute of Health and Biomedical Innovation, Queensland Institute of Technology, Brisbane, QLD 4059, Australia

¹²¹ Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA 02114, USA

¹²² The Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA.

-
- ¹²³ Psychiatric & Neurodevelopmental Genetics Unit, Department of Psychiatry, Massachusetts General Hospital, Boston, MA 02114, USA
- ¹²⁴ Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, 00014, Finland
- ¹²⁵ Department of Neurology, Massachusetts General Hospital, Boston, MA 02114, USA
- ¹²⁶ Medical Genetics, Institute for Maternal and Child Health IRCCS “Burlo Garofolo”, Trieste, 34100, Italy
- ¹²⁷ Social Impact, Arlington, VA 22201, USA
- ¹²⁸ Department of Economics, University of Minnesota Twin Cities, Minneapolis, MN 55455, USA
- ¹²⁹ Department of Psychiatry and Behavioral Sciences, NorthShore University HealthSystem, Evanston, IL 60201-3137, USA
- ¹³⁰ Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, IL 60637, USA
- ¹³¹ Public Health Genomics Unit, National Institute for Health and Welfare, Helsinki 00300, Finland
- ¹³² Research Unit for Genetic Epidemiology, Institute of Molecular Biology and Biochemistry, Center of Molecular Medicine, General Hospital and Medical University, Graz, Graz, 8010, Austria
- ¹³³ Information Based Medicine Stream, Hunter Medical Research Institute, New Lambton, NSW 2305, Australia
- ¹³⁴ Medical Research Institute, University of Dundee, Dundee, DD1 9SY, UK
- ¹³⁵ Research Unit Hypertension and Cardiovascular Epidemiology, Department of Cardiovascular Science, University of Leuven, Leuven, 3000, Belgium
- ¹³⁶ R&D VitaK Group, Maastricht University, Maastricht, 6229 EV, The Netherlands

-
- ¹³⁷ Institute of Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, 85764, Germany
- ¹³⁸ Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig Maximilians-Universität, Munich, 81377, Germany
- ¹³⁹ Department of Geriatrics, Florida State University College of Medicine, Tallahassee, FL 32306, USA
- ¹⁴⁰ Department of Health Sciences and Genetics, University of Leicester, Leicester, LE1 7RH, UK
- ¹⁴¹ Department of Internal Medicine, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands
- ¹⁴² Research Center for Group Dynamics, Institute for Social Research, University of Michigan, Ann Arbor, MI 48104, USA
- ¹⁴³ Platform for Genome Analytics, Institutes of Neurogenetics & Integrative and Experimental Genomics, University of Lübeck, Lübeck, 23562, Germany
- ¹⁴⁴ Neuroepidemiology and Ageing Research Unit, School of Public Health, Faculty of Medicine, The Imperial College of Science, Technology and Medicine, London SW7 2AZ, UK
- ¹⁴⁵ Department of Health Sciences, Community & Occupational Medicine, University of Groningen, University Medical Center Groningen, Groningen, 9713 AV, The Netherlands
- ¹⁴⁶ Department of Psychology, Union College, Schenectady, NY 12308, USA
- ¹⁴⁷ Istituto di Ricerca Genetica e Biomedica (IRGB), Consiglio Nazionale delle Ricerche, c/o Cittadella Universitaria di Monserrato, Monserrato, Cagliari, 9042, Italy
- ¹⁴⁸ Institute of Biomedical Technologies, Italian National Research Council, Segrate (Milano), 20090, Italy

-
- ¹⁴⁹ Department of General Practice and Primary Health Care, University of Helsinki, Helsinki, 00014, Finland
- ¹⁵⁰ Departments of Human Genetics and Psychiatry, Donders Centre for Neuroscience, Nijmegen, 6500 HB, The Netherlands
- ¹⁵¹ Sidra, Experimental Genetics Division, Sidra, Doha 26999, Qatar
- ¹⁵² Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, 17475, Germany
- ¹⁵³ Department of Psychiatry and Psychotherapy, HELIOS-Hospital Stralsund, Stralsund, 18437, Germany
- ¹⁵⁴ Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, 3062 PA, The Netherlands
- ¹⁵⁵ Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen. 9700 RB, The Netherlands
- ¹⁵⁶ Durrer Center for Cardiogenetic Research, ICIN-Netherlands Heart Institute, Utrecht, 1105 AZ, The Netherlands
- ¹⁵⁷ Generation Scotland, Centre for Genomics and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK
- ¹⁵⁸ Centre for Population Health Research, School of Health Sciences and Sansom Institute, University of South Australia, SA5000, Adelaide, Australia
- ¹⁵⁹ South Australian Health and Medical Research Institute, Adelaide, SA5000, Australia
- ¹⁶⁰ Population, Policy and Practice, UCL Institute of Child Health, London, WC1N 1EH, UK
- ¹⁶¹ Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment & Health, School of Public Health, Imperial College London, London, W2 1PG, UK
- ¹⁶² Center for Life Course Epidemiology, Faculty of Medicine, University of Oulu, Oulu, FI-90014, Finland

-
- ¹⁶³ Unit of Primary Care, Oulu University Hospital, Oulu, 90029 OYS, Finland
- ¹⁶⁴ Biocenter Oulu, University of Oulu, FI-90014 Oulu, Finland
- ¹⁶⁵ Fimlab Laboratories, Tampere, 33520, Finland
- ¹⁶⁶ Department of Clinical Chemistry, University of Tampere, School of Medicine, Tampere, 33014, Finland
- ¹⁶⁷ Economics, NYU Shanghai, 200122, Pudong, China
- ¹⁶⁸ Policy Studies, Queen's University, Kingston, K7L3N6, Canada
- ¹⁶⁹ Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4029, Australia
- ¹⁷⁰ Institute of Molecular and Cell Biology, University of Tartu, Tartu, 51010, Estonia
- ¹⁷¹ Centre for Clinical and Cognitive Neuroscience, Institute Brain Behaviour and Mental Health, Salford Royal Hospital, Manchester, M6 8HD, UK
- ¹⁷² Manchester Institute Collaborative Research in Ageing, University of Manchester, Manchester, M13 9PL, UK
- ¹⁷³ Faculty of Medicine, University of Split, Croatia, Split 21000, Croatia
- ¹⁷⁴ Department of Clinical Genetics, VU Medical Centre, Amsterdam, 1081 HV, The Netherlands
- ¹⁷⁵ Institute of Preventive Medicine, Bispebjerg and Frederiksberg Hospitals, The Capital Region, Frederiksberg, 2000, Denmark
- ¹⁷⁶ Montpellier Business School, Montpellier, 34080, France
- ¹⁷⁷ Panteia, Zoetermeer, 2715 CA, The Netherlands
- ¹⁷⁸ Department of Psychiatry, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands
- ¹⁷⁹ Department of Child and Adolescent Psychiatry, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands

¹⁸⁰ Department of Internal Medicine, Erasmus Medical Center, Rotterdam, 3015 GE, The Netherlands

¹⁸¹ Department of Sociology, New York University, New York, NY 10012, USA

¹⁸² School of Medicine, New York University, NY 10016, New York, USA

¹⁸³ Bioethics Program, Union Graduate College - Icahn School of Medicine at Mount Sinai, Schenectady, NY 12308, USA

¹⁸⁴ The University of Queensland Diamantina Institute, The Translational Research Institute, Brisbane, QLD 4102, Australia

¹⁸⁵ Department of Economics, Stockholm School of Economics, Stockholm, 113 83, Sweden

¹⁸⁶ Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

¹⁸⁷ Research Institute for Industrial Economics, Stockholm, 10215, Sweden

¹⁸⁸ Department of Economics, Cornell University, Ithaca, NY 14853, USA

¹⁸⁹ Center for Economic and Social Research, University of Southern California, Los Angeles, CA 90089-3332, USA

1 **Figure 1 | Manhattan plot for *EduYears* associations ($N = 293,723$).** The x -axis is
2 chromosomal position, and the y -axis is the significance on a $-\log_{10}$ scale. The black line
3 shows the genome-wide significance level (5×10^{-8}). The red x 's are the 74 approximately
4 independent genome-wide significant associations ("lead SNPs"). The black dots labeled
5 with rs numbers are the 3 Rietveld et al.¹ SNPs.

6

7

8 **Figure 2 | Genetic correlations between *EduYears* and other traits.** Results from bivariate
9 Linkage-Disequilibrium (LD) Score regressions⁹: estimates of genetic correlation with brain
10 volume, neuropsychiatric, behavioral, and anthropometric phenotypes using published GWAS
11 summary statistics. The error bars show the 95% confidence intervals.

12

13

14 **Figure 3 | Overview of biological annotation.** 34 clusters of significantly enriched gene sets.
15 Each cluster is named after one of its member gene sets. The color represents the P -value of
16 the member set exhibiting the most statistically significant enrichment. Overlap between pairs
17 of clusters is represented by an edge. Edge width represents the Pearson correlation ρ between
18 the two vectors of gene membership scores ($\rho < 0.3$, no edge; $0.3 \leq \rho < 0.5$, thin edge; $0.5 \leq \rho$
19 < 0.7 , intermediate edge; $\rho \geq 0.7$, thick edge), where each cluster's vector is the vector for the
20 gene set after which the cluster is named.

21

22

23

1 **Extended Data Figure 1 | Quantile-quantile plot of the genome-wide association meta-**
2 **analysis of 64 *EduYears* results files.** Observed and expected P -values are on a $-\log_{10}$ scale.
3 The grey region depicts the 95% confidence interval under the null hypothesis of a uniform P -
4 value distribution. The observed λ_{GC} is 1.28. (As reported in Supplementary Information
5 section 1.5.4, the unweighted mean λ_{GC} is 1.02, the unweighted median is 1.01, and the range
6 across cohorts is 0.95–1.15.)

7

8

9 **Extended Data Figure 2 | The distribution of effect sizes of the 74 lead SNPs. a,** SNPs
10 ordered by absolute value of the standardized effect of one more copy of the education-
11 increasing allele, with 95% confidence intervals. **b,** SNPs ordered by R^2 . Effects on *EduYears*
12 are benchmarked against the top 74 genome-wide significant hits identified in the largest
13 GWAS conducted to date of height and body mass index (BMI), and the 48 associations
14 reported for waist-to-hip ratio adjusted for BMI (WHR). These results are based on the
15 GIANT consortium's publicly available results for pooled analyses restricted to European-
16 ancestry individuals:

17 https://www.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium.

18

19

20 **Extended Data Figure 3 | Assessing the extent to which population stratification affects**
21 **the estimates from the GWAS. a,** LD Score regression plot with the summary statistics from
22 the GWAS. Each point represents an LD Score quantile for a chromosome (the x and y
23 coordinates of the point are the mean LD Score and the mean χ^2 statistic of variants in that
24 quantile). The facts that the intercept is close to one and that the χ^2 statistics increase linearly
25 with the LD Scores suggest that the bulk of the inflation in the χ^2 statistics is due to true

1 polygenic signal and not to population stratification. **b**, Estimates and 95% confidence intervals
2 from individual-level and WF regressions of *EduYears* on polygenic scores, for scores
3 constructed with sets of SNPs meeting different *P*-value thresholds. In addition to the analyses
4 shown here, we conduct a sign concordance test, and we decompose the variance of the
5 polygenic score. Overall, these analyses suggest that population stratification is unlikely to be
6 a major concern for our 74 lead SNPs. See Supplementary Information section 3 for additional
7 details.

8
9

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

17
18

19 **Extended Data Figure 5 | Q-Q plots for the 74 lead *EduYears* SNPs (or LD proxies) in**
20 **published GWAS of other phenotypes.** SNPs with concordant effects on both phenotypes are
21 pink, and SNPs with discordant effects are blue. SNPs outside the gray area pass Bonferroni-
22 corrected significance thresholds that correct for the total number of SNPs we tested ($P <$
23 $0.05/74 = 6.8 \times 10^{-4}$) and are labeled with their rs numbers. Observed and expected *P*-values are
24 on a $-\log_{10}$ scale. For the sign concordance test: * $P < 0.05$, ** $P < 0.01$, and *** $P < 0.001$.

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

8
9

10 **Extended Data Figure 7 | Application of fgwas to *EduYears*. See Supplementary**
11 **Information section 4.2 for further details. a,** The results of single-annotation models.
12 “Enrichment” refers to the factor by which the prior odds of association at an LD-defined
13 region must be multiplied if the region bears the given annotation; this factor is estimated using
14 an empirical Bayes method applied to all SNPs in the GWAS meta-analysis regardless of
15 statistical significance. Annotations were derived from ENCODE and a number of other data
16 sources. Plotted are the base-2 logarithms of the enrichments and their 95% confidence
17 intervals. Multiple instances of the same annotation correspond to independent replicates of the
18 same experiment. **b,** The results of combining multiple annotations and applying model
19 selection and cross-validation. Although the maximum-likelihood estimates are plotted, model
20 selection was performed with penalized likelihood. **c,** Reweighting of GWAS loci. Each point
21 represents an LD-defined region of the genome, and shown are the regional posterior
22 probabilities of association (PPAs). The x -axis give the PPA calculated from the GWAS
23 summary statistics alone, whereas the y -axis gives the PPA upon reweighting on the basis of
24 the annotations in b. The orange points represent genomic regions where the PPA is equivalent
25 to the standard GWAS significance threshold only upon reweighting.

1 **Extended Data Figure 8 | Tissue-level biological annotation. a,** The enrichment factor for a
2 given tissue type is the ratio of variance explained by SNPs in that group to the overall fraction
3 of SNPs in that group. To benchmark the estimates for *EduYears*, we compare the enrichment
4 factors to those obtained when we use the largest GWAS conducted to date on body mass
5 index, height, and waist-to-hip ratio adjusted for BMI. The estimates were produced with the
6 LDSC python software, using the LD Scores and functional annotations introduced in Finucane
7 et al. (2015) and the HapMap3 SNPs with MAF > 0.05. Each of the 10 enrichment calculations
8 for a particular cell type is performed independently, while each controlling for the 52
9 functional annotation categories in the full baseline model. The error bars show the 95%
10 confidence intervals. **b,** We took measurements of gene expression by the Genotype-Tissue
11 Expression (GTEx) Consortium and determined whether the genes overlapping *EduYears*-
12 associated loci are significantly overexpressed (relative to genes in random sets of loci matched
13 by gene density) in each of 37 tissue types. These types are grouped in the panel by organ. The
14 colored bars corresponding to tissues where there is significant overexpression. The y-axis is
15 the significance on a $-\log_{10}$ scale.

16

17

18 **Extended Data Figure 9 | Gene-level biological annotation. a,** The DEPICT-prioritized
19 genes for *EduYears* measured in the BrainSpan Developmental Transcriptome data (red curve)
20 are more strongly expressed in the brain prenatally rather than postnatally. The DEPICT-
21 prioritized genes exhibit similar gene-expression levels across different brain regions (gray
22 lines). Analyses were based on \log_2 -transformed RNA-Seq data. Error bars represent 95%
23 confidence intervals. **b,** For each phenotype and disorder, we calculated the overlap between
24 the phenotype's DEPICT-prioritized genes and genes believed to harbor *de novo* mutations
25 causing the disorder. The bars correspond to odds ratios. *EduYears*, years of education; BMI,

1 body mass index; WHR, waist-to-hip ratio adjusted for BMI. **c**, DEPICT-prioritized genes in
2 *EduYears*-associated loci exhibit substantial overlap with genes previously reported to harbor
3 sites where mutations increase risk of intellectual disability and autism spectrum disorder
4 (Supplementary Table 4.6.1).

5

6

7 **Extended Figure 10 | The predictive power of a polygenic score (PGS) varies in Sweden**
8 **by birth cohort.** Five-year rolling regressions of years of education on the PGS (left axis in all
9 four panels), share of individuals not affected by the comprehensive school reform (**a**, right
10 axis), and average distance to nearest junior high school (**b**, right axis), nearest high school (**c**,
11 right axis) and nearest college/university (**d**, right axis). The shaded area displays the 95%
12 confidence intervals for the PGS effect.

13

