Gene discovery and polygenic prediction from a 1.1-million-person GWAS of educational attainment

AUTHORS:

James J. Lee[1,†], Robbee Wedow[2,3,4,†], Aysu Okbay[5,6,*†], Edward Kong[7], Omeed Maghzian[7], Meghan Zacher[8], Tuan Anh Nguyen-Viet[9], Peter Bowers[7], Julia Sidorenko[10,11], Richard Karlsson Linnér[5,6], Mark Alan Fontana[12], Tushar Kundu[9], Chanwook Lee[7], Hui Li[7], Ruoxi Li[9], Rebecca Royer[9], Pascal N. Timshel[13,14], Raymond K. Walters[15,16], Emily A. Willoughby[1], Loïc Yengo[10], *23andMe Research Team*[17], *COGENT (Cognitive Genomics Consortium)*, *Social Science Genetic Association Consortium*, Maris Alver[11], Yanchun Bao[18], David W. Clark[19], Felix R. Day[20], Nicholas A. Furlotte[17], Peter K. Joshi[19], Kathryn E. Kemper[10], Aaron Kleinman[17], Claudia Langenberg[20], Reedik Mägi[11], Joey W. Trampush[21], Shefali Setia Verma[22], Yang Wu[10], Max Lam Zhan Yang[23], Jing Hua Zhao[20], Zhili Zheng[10,24], Jason D. Boardman[2,3,4], Harry Campbell[19], Jeremy Freese[25], Kathleen Mullan Harris[26,27], Caroline Hayward[28], Pamela Herd[29], Meena Kumari[18], Todd Lencz[30,31,32], Jian'an Luan[20], Anil K. Malhotra[30,31,32], Andres Metspalu[11], Lili Milani[11], Ken K. Ong[20], John R. B. Perry[20], David J. Porteous[33], Marylyn D. Ritchie[22], Melissa C. Smart[18], Blair H. Smith[34], Joyce Y. Tung[17], Nicholas J. Wareham[20], James F. Wilson[19,28], Jonathan P. Beauchamp[35], Dalton C. Conley[36], Tõnu Esko[11], Steven F. Lehrer[37,38,39], Patrik K. E. Magnusson[40], Sven Oskarsson[41], Tune H. Pers[13,14], Matthew R. Robinson[10,42], Kevin Thom[43], Chelsea Watson[9], Christopher F. Chabris[44], Michelle N. Meyer[45], David I. Laibson[7], Jian Yang[10,46], Magnus Johannesson[47], Philipp D. Koellinger[5,6], Patrick Turley[15,16,#], Peter M. Visscher[10,46,*#], Daniel J. Benjamin[9,39,48,*#], and David Cesarini[39,43,49,#]

[1] Department of Psychology, University of Minnesota Twin Cities, Minneapolis, Minnesota, USA
2 Department of Sociology, University of Colorado Boulder, Boulder, Colorado, USA
[3] Institute for Behavioral Genetics, University of Colorado Boulder, Boulder, Colorado, USA
[4] Institute of Behavioral Science, University of Colorado Boulder, Boulder, Colorado, USA
[5] Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands
[6] Institute for Behavior and Biology, Erasmus University Rotterdam, Rotterdam, the Netherlands
[7] Department of Economics, Harvard University, Cambridge, Massachusetts 02138, USA
[8] Department of Sociology, Harvard University, Cambridge, Massachusetts, USA
[9] Center for Economic and Social Research, University of Southern California, Los Angeles, California, USA
[10] Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia
[11] Estonian Genome Center, University of Tartu, Tartu, Estonia
[12] Hospital for Special Surgery, 535 E 70th Street, New York, New York 10021, USA
[13] 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
[14] Statens Serum Institut, Department of Epidemiology Research, Copenhagen 2300, Denmark

[15] Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA
[16] Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA
[17] 23andMe, Inc., Mountain View, California 94043, USA
[18] Institute for Social and Economic Research, University of Essex, Colchester, UK
[19] Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, Scotland
[20] MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge, Cambridge, UK
[21] BrainWorkup, LLC, Santa Monica, California, USA
[22] Biomedical and Translational Informatics, Geisinger Health System, Lewisburg, Pennsylvania, USA
[23] Institute of Mental Health, Singapore, Singapore
[24] The Eye Hospital, School of Ophthalmology & Optometry, Wenzhou Medical University, Wenzhou, Zhejiang, China
[25] Department of Sociology, Stanford University, Stanford, California, USA
[26] Department of Sociology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA
[27] Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA
[28] MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, Scotland
[29] La Follette School of Public Affairs, University of Wisconsin-Madison, Madison, Wisconsin, USA
[30] Departments of Psychiatry and Molecular Medicine, Hofstra Northwell School of Medicine, Hempstead, New York, USA
[31] Center for Psychiatric Neuroscience, Feinstein Institute for Medical Research, Manhasset, New York, USA
[32] Psychiatry Research, The Zucker Hillside Hospital, Glen Oaks, California, USA
[33] Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, Scotland
[34] Division of Population Health Sciences, Ninewells Hospital and Medical School, University of Dundee, Dundee, Scotland
[35] Department of Economics, University of Toronto, Toronto, Ontario, Canada
[36] Department of Sociology, Princeton University, Princeton, New Jersey, USA
[37] Department of Policy Studies, Queen's University, Kingston, Ontario, Canada
[38] Department of Economics, New York University Shanghai, Pudong, Shanghai, China
[39] National Bureau of Economic Research, Cambridge, MA, USA
[40] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
[41] Department of Government, Uppsala University, Uppsala, Sweden
[42] Department of Computational Biology, University of Lausanne, Lausanne, Switzerland
[43] Department of Economics, New York University, New York, New York, USA
[44] Autism and Developmental Medicine Institute, Geisinger Health System, Lewisburg, Pennsylvania, USA

[45] Center for Translational Bioethics and Health Care Policy, Geisinger Health System, Danville, Pennsylvania, USA
[46] Queensland Brain Institute, University of Queensland, Brisbane, Australia
[47] Department of Economics, Stockholm School of Economics, Stockholm, Sweden
[48] Department of Economics, University of Southern California, Los Angeles, California, USA
[49] Center for Experimental Social Science, New York University, New York, New York, USA

† These authors contributed equally.
# These authors jointly directed the work.

* Correspondence to Daniel Benjamin, daniel.benjamin@gmail.com, Aysu Okbay, a.okbay@vu.nl, and Peter Visscher, p.visscher@imb.uq.edu.au.

ABSTRACT

We conduct a large-scale genetic association analysis of educational attainment in a sample of ~1.1 million individuals and identify 1,271 independent genome-wide significant loci. For the loci taken together, we find evidence of heterogeneous effects across environments. The loci implicate genes involved in brain-development processes and neuron-to-neuron communication. In a separate analysis of the X chromosome, we identify 10 loci and estimate a SNP heritability of ~0.3% in both men and women, consistent with partial dosage compensation. A joint (multi-phenotype) analysis of educational attainment and three related cognitive phenotypes generates polygenic scores that explain 11-13% of the variance in educational attainment and 7-10% of the variance in cognitive performance. This prediction accuracy substantially increases the utility of polygenic scores as tools in research.

INTRODUCTION

Educational attainment (EA) is an important correlate of many social, economic, and health outcomes[1,2]. The largest GWAS of EA conducted to date identified 74 loci in a discovery sample of 293,723 individuals and reported that a 10-million-SNP polygenic score explained 3.2% of the variance in independent samples[3]. Here, we report results from (i) a meta-analysis of EA based on a much larger sample of 1,131,881 individuals and (ii) joint (multi-trait) analyses of EA and three genetically correlated phenotypes: cognitive (test) performance (N = 257,841), self-reported math ability (N = 564,698), and hardest math class completed (N = 430,445).

RESULTS

Main GWAS Results

In our primary GWAS, we study EA, which is measured as number of years of schooling completed (*EduYears*). All association analyses were performed at the cohort level in samples restricted to European-descent individuals. We applied a uniform set of quality-control procedures to all cohort-level results. Our final sample-size-weighted meta-analysis produced association statistics for ~10 million SNPs from phase 3 of the 1000 Genomes Project[4].

The quantile-quantile plot of the meta-analysis (**Supplementary Figure 1.1**) exhibits substantial inflation (λ_{GC} = 2.04). According to our LD Score regression[5] estimates, only a small share (~5%) of this inflation is attributable to bias (**Supplementary Figure 1.2**). We used the estimated LD Score intercept (1.11) to generate inflation-adjusted test statistics. **Fig. 1** shows the Manhattan plot of the resulting P values. Overall, our meta-analysis identified 1,271 approximately independent (pairwise $R^2 < 0.1$) SNPs at genome-wide significance ($P < 5\times10^{-8}$). Adjusted for winner's curse, the median effect size for these SNPs corresponds to 1.7 weeks of schooling per allele; at the 5th and 95th percentiles, 1.1 and 2.6 weeks, respectively.

We examined the replicability of 162 single-SNP associations ($P < 5\times10^{-8}$) from the combined discovery and replication sample (N = 405,073) of the largest previous study[3]. In the subsample of our data (N = 726,808) that did not contribute to the earlier study's analyses, the SNPs replicate at a rate that closely matches theoretical projections that account for sampling variation and winner's curse (**Supplementary Figure 1.4**).

To probe the robustness of our meta-analysis findings, we compared results from within-family association analyses conducted in four sibling cohorts to those from a meta-analysis that excluded the siblings. Our sample of 22,135 sibling pairs is too small to allow well-powered within-family association analyses of single SNPs but large enough for joint analyses of the lead SNPs. We find greater sign concordance than expected if GWAS results were driven primarily by stratification bias (**Supplementary Figure 2.1**; see **Supplementary Note** for a more detailed discussion).

Because educational institutions vary across places and time, the effects of specific SNPs may vary across environments. Consistent with such heterogeneity, we find that the inverse-variance-weighted mean genetic correlation of *EduYears* across pairs of cohorts in our sample is 0.72 (SE = 0.14), which is statistically distinguishable from one (*P* value = 0.03). Moreover, for the lead SNPs, we reject the joint null hypothesis of homogeneous cohort-level effects (*P* value = 9.7×10^{-12}; **Supplementary Figure 1.3**).

We supplemented our autosomal analyses with association analyses of SNPs on the X chromosome. We first conducted separate association analyses of males (*N* = 152,608) and females (*N* = 176,750) in the UK Biobank, finding a male-female genetic correlation close to unity. We also find nearly identical SNP heritability estimates for men and women, which is consistent with partial dosage compensation (i.e., on average the per-allele effect sizes are smaller in women) and implies that any contribution of common variants on the X chromosome to sex differences in the normal-range *variance* of cognitive phenotypes[6] is quantitatively negligible. Next, we conducted a large (*N* = 694,894) meta-analysis of summary statistics from mixed-sex analyses (**Supplementary Figure 4.1**). We identify 10 genome-wide significant loci and estimate SNP heritability due to the X chromosome of ~0.3%. This heritability is lower than that expected for an autosome of similar length (**Supplementary Figure 4.2**, **Supplementary Table 4.3**).

Biological Annotation

For biological annotation, we focus on the results from the autosomal meta-analysis of *EduYears*. Across an extensive set of analyses (see **Supplementary Figure 5.1** for a flowchart), all major conclusions from the largest previous GWAS of EA[3] continue to hold but are statistically stronger. For example, we applied the bioinformatics tool DEPICT[7] and found that, relative to other genes, genes near our lead SNPs are overwhelmingly enriched for expression in the central nervous system (**Fig. 2A**).

There are also many novel findings associated with the large number of genes newly implicated by our analyses: At the standard false discovery rate (FDR) threshold of 5%, the bioinformatics tool DEPICT[7] prioritizes 1,838 genes, a tenfold increase relative to the DEPICT results from an earlier GWAS of *EduYears*[3]. In what follows, we distinguish between the 1,703 "newly prioritized" genes and the 135 "previously prioritized" genes. The SOM contains an extensive analysis of many of the newly prioritized genes and their brain-related functions. Here we highlight two especially noteworthy regularities. First, whereas previously prioritized genes exhibited especially high expression in the brain prenatally, newly prioritized genes show elevated levels of expression both pre- and postnatally (**Fig. 2B**). Many of the newly prioritized genes encode proteins that carry out online brain functions such as neurotransmitter secretion, the activation of ion channels and metabotropic pathways, and synaptic plasticity. For a number of newly prioritized genes, **Fig. 3** illustrates the crucial roles in cellular neurophysiology[8] played by their protein products in the postnatal brain.

Second, even though glial cells are at least as numerous as neurons in the human brain[9], gene sets related to glial cells (astrocytes, myelination, and positive regulation of gliogenesis) are absent from those identified as positively enriched (**Supplementary Table 5.5**). Furthermore, using stratified LD Score regression[10], we estimated relatively weak enrichment of genes highly expressed in glial cells: 1.08-fold for astrocytes ($P = 0.07$) and 1.09-fold for oligodendrocytes ($P = 0.06$) versus 1.33-fold for neurons ($P = 2.89\times10^{-11}$). Because myelination increases the speed with which signals are transmitted along axons[11], the absence of enrichment of genes related to glial cells may weigh against the hypothesis that differences across people in cognition are driven by differences in transmission speed.

The results also raise a number of possible targets for functional studies. Among SNPs within 50 kb of lead SNPs, 127 of them are identified by the fine-mapping tool CAVIARBF[12] as likely causal SNPs (posterior probability > 0.9). Eight of these are non-synonymous, and one of these (rs61734410) is located in *CACNA1H*, which encodes the pore-forming subunit of a voltage-gated calcium channel that has been implicated in the trafficking of NMDA-type glutamate receptors[13].

Polygenic Prediction

Polygenic predictors derived from earlier GWAS have proven to be a valuable tool for researchers, especially in the social sciences[14,15]. We constructed polygenic scores for European-ancestry individuals in two prediction cohorts: the National Longitudinal Study of Adolescent to Adult Health (Add Health, $N = 4{,}775$), a representative sample of American adolescents; and the Health and Retirement Study (HRS, $N = 8{,}609$), a representative sample of Americans over age 50. We measure prediction accuracy by the "incremental R^2": the gain in coefficient of determination (R^2) when the score is added as a covariate to a regression of the phenotype on a set of baseline controls (sex, age, and 10 principal components of the genetic relatedness matrix).

All scores are based on results from a meta-analysis that excluded the prediction cohorts. Our first four scores were constructed from sets of LD-pruned SNPs associated with *EduYears* at various *P*-value thresholds: 5×10^{-8}, 5×10^{-5}, 5×10^{-3}, and 1 (i.e., all SNPs). In both cohorts, the predictive power is greater for scores constructed with less stringent thresholds (**Supplementary Figure 6.3**). The (sample-size weighted) mean incremental R^2 increases from 3.2% at $P < 5\times10^{-8}$ to 9.4% at $P \leq 1$. Our fifth score was generated from HapMap3 SNPs using the software LDpred[16]. Rather than dropping SNPs in LD with each other, LDpred weights each SNP by (an approximation to) its conditional effect, given other SNPs. This score had the greatest incremental R^2: 11.4%. Hereafter, we focus on the LDpred score.

To put the predictive power of this score in perspective, **Fig. 4A** shows the mean college completion rate by polygenic-score quintile. The difference between the bottom and top quintiles in Add Health and HRS is, respectively, 45 and 36 percentage points (see **Supplementary Figure 6.4** for analogous analyses of high school completion and grade retention). **Fig. 4B**

compares the incremental R^2 of the score to that of standard demographic variables. The score is a better predictor of *EduYears* than household income and a worse predictor than mother's or father's education. Controlling for all the demographic variables jointly, the score's incremental R^2 is 4.6% (**Supplementary Figure 6.7**). We also found that the score has substantial predictive power for a variety of other cognitive phenotypes measured in the prediction cohorts (**Supplementary Figure 6.1**). For example, it explains 9.2% of the variance in overall grade point average in Add Health.

Related Cognitive Phenotypes and MTAG

We also performed genome-wide association analyses of three complementary phenotypes: cognitive performance (*CP*, N = 257,841), self-reported math ability (*Math Ability*, N = 564,698), and highest math class taken (*Highest Math*, N = 430,445). For cognitive performance, we meta-analyzed published results from the COGENT Consortium[17] with results based on new analyses of the UKB. For the two math phenotypes, we studied new genome-wide analyses in samples of research participants from 23andMe. All analyses and quality-control procedures were harmonized to ensure comparability with the *EduYears* GWAS. We identified 225, 618, and 365 genome-wide significant loci with *CP*, *Math Ability*, and *Highest Math*, respectively (**Supplementary Figures 1.5-1.7**, **Supplementary Tables 1.8-1.10**).

We conducted a multi-trait analysis of *EduYears* and our supplementary phenotypes to further improve prediction accuracy. These phenotypes are well suited to joint analysis because their pairwise genetic correlations are high, in all cases exceeding 0.5 (**Supplementary Table 1.11**). We applied a recently developed method, Multi-Trait Analysis of GWAS, or MTAG[18], to summary statistics from the four phenotypes (again excluding the prediction cohorts). MTAG allows for sample overlap across the phenotypes and generates phenotype-specific association statistics. For all four phenotypes, MTAG increases the number of loci identified at genome-wide significance (**Supplementary Figures 1.8-1.12**, **Supplementary Table 1.14**). For example, MTAG identifies 661 loci associated with *CP*, and a host of follow-up analyses suggest that the false discovery rate is low.

Polygenic scores constructed from MTAG results are expected theoretically to outperform corresponding scores based on GWAS results under general conditions. **Fig. 4C** shows the incremental R^2 for the polygenic scores based on GWAS and MTAG association statistics (but otherwise constructed using identical methods) when the target phenotype is either *EduYears* (left panel) or *CP* (right panel). For *EduYears*, relative to the GWAS score, the MTAG score improves predictive power from 12.7% to 13.0% in Add Health and from 10.6% to 11.2% in the HRS. To measure prediction accuracy for cognitive performance, we used a third validation cohort, the Wisconsin Longitudinal Study (WLS), because it contains a measure with excellent retest reliability and psychometric properties similar to those used in our discovery GWAS of cognitive performance. In the WLS, the MTAG score predicts 9.7% of the variance in *CP*, a substantial improvement over the 7.0% predicted by the GWAS score—and approximately

double the prediction accuracy reported in three recent GWASs of cognitive performance[19–21]. In Add Health, where our measure of cognitive performance is the respondent's score on a test of verbal cognition, the incremental R^2s of the GWAS and MTAG scores are 5.1% and 6.9%, respectively.

DISCUSSION

For social science, the polygenic scores are the most important results of this paper. With their levels of predictive power—11-13% and 7-10% of the variance of EA and cognitive performance, respectively—they will be useful across at least three types of applications. First, it is now possible to conduct well-powered studies of the mechanisms by which genetic factors affect EA by examining associations between the scores and high-quality measures of endophenotypes in samples as small as those from laboratory experiments. Second, the polygenic scores can now generate non-trivial gains in statistical power when used as control variables in randomized-controlled trials of expensive interventions that aim to improve academic and cognitive outcomes [see the calculations in the SOM of Rietveld et al.[22]]. Third, genetic effects on educational attainment and cognitive performance have repeatedly been found to vary across environmental contexts[23,24], and the polygenic scores provide a new and powerful tool for researchers interested in exploring such gene-environment interactions.

CODE AVAILABILITY:

All software used to perform these analyses are available online.

URLs:

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

ACKNOWLEDGMENTS:

This research was carried out under the auspices of the Social Science Genetic Association Consortium (SSGAC). The research has also been conducted using the UK Biobank Resource under application numbers 11425 and 12512. This study was supported by funding from the Ragnar Söderberg Foundation (E9/11, E24/15), the Swedish Research Council (421-2013-1061), The Jan Wallander and Tom Hedelius Foundation, an ERC Consolidator Grant (647648 EdGe), the Pershing Square Fund for Research on~~f~~ 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. A full list of acknowledgments is provided in the Supplementary Note.

CONTRIBUTOR LIST FOR THE 23andMe RESEARCH TEAM: Michelle Agee, Babak Alipanahi, Adam Auton, Robert K. Bell, Katarzyna Bryc, Sarah L. Elson, Pierre Fontanillas, Nicholas A. Furlotte, David A. Hinds, Bethann S. Hromatka, Karen E. Huber, Aaron Kleinman, Nadia K. Litterman, Matthew H. McIntyre, Joanna L. Mountain, Carrie A.M. Northover, J. Fah Sathirapongsasuti, Olga V. Sazonova, Janie F. Shelton, Suyash Shringarpure, Chao Tian, Joyce Y. Tung, Vladimir Vacic, Catherine H. Wilson, and Steven J. Pitts.

AUTHOR CONTRIBUTIONS: D.J.B., D.C., P.T., and P.M.V. designed and oversaw the study. A.O. was the study's lead analyst, responsible for quality control and meta-analyses. Analysts who assisted A.O. in major ways include: E.K. (quality control), O.M. (COJO, MTAG, quality-control), T.A.N-V. (figure preparation), H.L. (quality control), C.L. (quality control), J.S. (UKB association analyses), and R.K.L. (UKB association analyses). P.B. and E.K. conducted the within-family association analyses. The cross-cohort heritability and genetic-correlation analyses were conducted by R.W. and M.Z. The analyses of the X chromosome in UK Biobank were conducted by J.S.; A.O. ran the meta-analysis. J.J.L. organized and oversaw the bioinformatics analyses, with assistance from T.E., E.K., K.T., T.H.P., and P.N.T. Polygenic-prediction analyses were designed and conducted by A.O., K.T., and R.W. Besides the contributions explicitly listed above, T.K., R.L., and R.R. conducted additional analyses for several subsections. C.W. helped with coordinating among the participating cohorts. J.P.B., D.C.C., T.E., M.J., J.J.L., P.D.K., D.I.L., S.F.L., S.O., M.R.R., K.T., and J.Y. provided helpful advice and feedback on various aspects of the study design. All authors contributed to and critically reviewed the manuscript. E.K., J.J.L., and R.W. made especially major contributions to the writing and editing.

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

REFERENCES

1. Conti, G., Heckman, J. & Urzua, S. The Education-Health Gradient. *Am. Econ. Rev.* **100,** 234–238 (2010).
2. Cutler, D. M. & Lleras-Muney, A. in *Making Americans Healthier: Social and Economic Policy as Health Policy* (eds. House, J., Schoeni, R., Kaplan, G. & Pollack, H.) (Russell Sage Foundation, 2008).
3. Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533,** 539–542 (2016).

4. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526,** 68–74 (2015).

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

6. Johnson, W., Carothers, A. & Deary, I. J. Sex Differences in Variability in General Intelligence: A New Look at the Old Question. *Perspect. Psychol. Sci.* **3,** 518–531 (2008).

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

8. Fain, G. L. *Molecular and Cellular Physiology of Neurons*. (Harvard University Press). doi:10.1017/CBO9781107415324.004

9. Azevedo, F. A. C. *et al.* Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *J. Comp. Neurol.* **513,** 532–541 (2009).

10. Finucane, H. K. *et al.* Partitioning heritability by functional category using GWAS summary statistics. *Nat. Genet.* **47,** 1228–1235 (2015).

11. Reed, T. E. & Jensen, A. R. Arm nerve conduction velocity (NCV), brain NCV, reaction time, and intelligence. *Intelligence* **15,** 33–47 (1991).

12. Chen, W., McDonnell, S. K., Thibodeau, S. N., Tillmans, L. S. & Schaid, D. J. Incorporating functional annotations for fine-mapping causal variants in a Bayesian framework using summary statistics. *Genetics* **204,** 933–958 (2016).

13. Wang, G. *et al.* CaV3.2 calcium channels control NMDA receptor-mediated transmission: a new mechanism for absence epilepsy. *Genes Dev.* **29,** 1535–51 (2015).

14. Domingue, B. W., Belsky, D. W., Conley, D., Harris, K. M. & Boardman, J. D. Polygenic Influence on Educational Attainment: Polygenic Influence on Educational Attainment: New evidence from The National Longitudinal Study of Adolescent to Adult Health. *AERA Open* **1,** 1–13 (2015).

15. Belsky, D. W. *et al.* The Genetics of Success. *Psychol. Sci.* **27,** 957–972 (2016).

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

17. Trampush, J. W. *et al.* GWAS meta-analysis reveals novel loci and genetic correlates for general cognitive function: a report from the COGENT consortium. *Mol. Psychiatry* **22,** 336–345 (2017).

18. Turley, P. *et al.* MTAG: Multi-Trait Analysis of GWAS. *Nat. Genet.* **in press,** (2017).

19. Sniekers, S. *et al.* Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nat Genet* **49,** 1107–1112 (2017).

20. Davies, G. *et al.* Ninety-nine independent genetic loci influencing general cognitive function include genes associated with brain health and structure (N = 280,360). *bioRxiv* (2017). at <http://www.biorxiv.org/content/early/2017/08/17/176511>

21. Savage, J. E. *et al.* GWAS meta-analysis (N=279,930) identifies new genes and functional

links to intelligence. *bioRxiv* (2017). at <https://www.biorxiv.org/content/early/2017/09/06/184853.1>

22. Rietveld, C. A. *et al.* GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science.* **340,** 1467–1471 (2013).

23. Branigan, A. R. *et al.* Variation in the Heritability of Educational Attainment: An International Meta-Analysis. *Soc. Forces* **92,** 109–140 (2013).

24. Heath, A. C. *et al.* Education policy and the heritability of educational attainment. *Nature* **314,** 734–736 (1985).

25. Kang, H. J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478,** 483–489 (2011).

Fig. 1. Manhattan Plot for GWAS of *EduYears* (*N* = 1,131,881). *P* values and the mean χ^2 shown in figure are based on inflation-adjusted test statistics. The *x*-axis is chromosomal position, and the *y*-axis is the significance on a $-\log_{10}$ scale. The dashed line marks the threshold for genome-wide significance ($P = 5\times10^{-8}$).

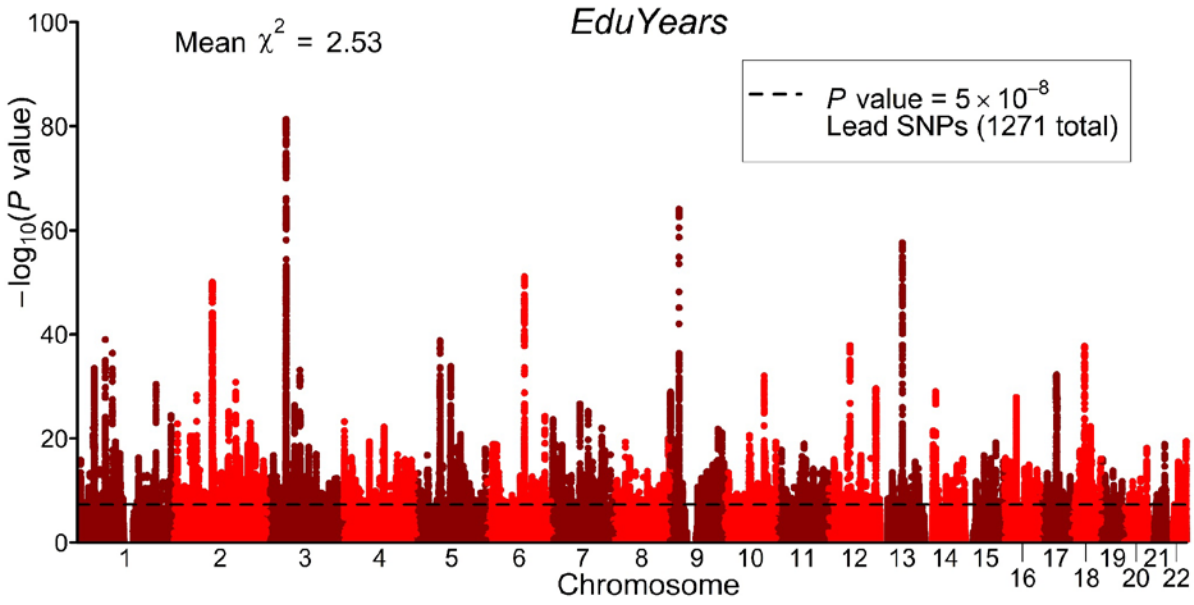


Fig. 2. Tissue-specific expression of genes in DEPICT-defined loci. (A) We took microarray measurements from the Gene Expression Omnibus[7] and determined whether the genes overlapping *EduYears*-associated loci are significantly overexpressed (relative to genes in random sets of loci) in each of 180 tissues/cell types. These types are grouped in the figure by Medical Subject Headings (MeSH) first-level term. The *y*-axis is the one-sided *P* value from DEPICT on a $-\log_{10}$ scale. The 28 dark bars correspond to tissues/cell types in which the genes are significantly overexpressed (FDR < 0.01), including all 22 classified as part of the central nervous system (see **Supplementary Table 5.1** for identifiers of all tissues/cell types). **(B)** Whereas genes prioritized by DEPICT in a previous analysis based on a smaller sample[3] tend to be more strongly expressed in the brain prenatally (red curve), the 1,703 newly prioritized genes show a flat trajectory of expression across development (blue curve). Both groups of DEPICT-prioritized genes show elevated levels of expression relative to protein-coding genes that are not prioritized (gray curve). Analyses were based on RNA-seq data from the BrainSpan Developmental Transcriptome[25]. Error bars represents 95% confidence intervals.

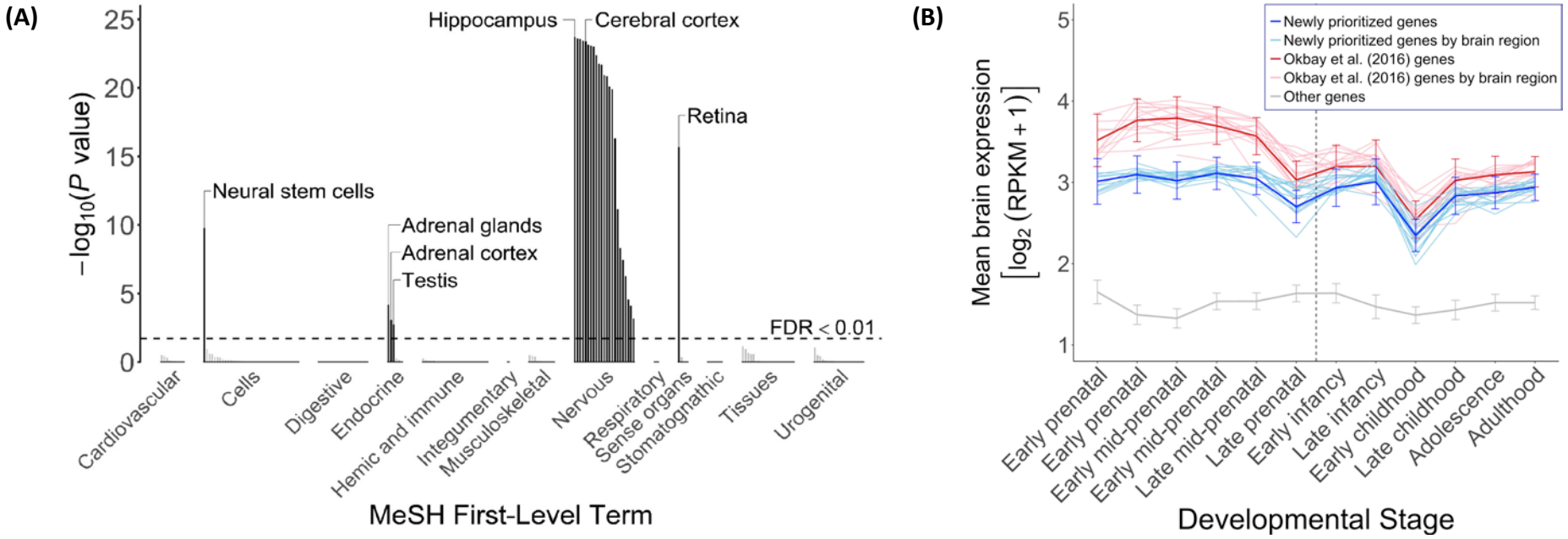


Fig. 3. Roles of selected newly prioritized genes involved in neuronal communication. The 59 genes listed in the figure were selected as follows. We began with the 30 gene-set clusters in **Supplementary Figure 5.2** and dropped those that include gene sets that were implicated in a previous study of *EduYears* [Supplementary Table 4.5.1 of Okbay et al.[3]]. Of the 8 clusters that remained, we retained the 4 related to neuronal communication ("DAG and IP_3 signaling," "associative learning," "post NMDA receptor activation events," "regulation of neurotransmitter levels"). We identified the 460 DEPICT-prioritized genes that belong to the exemplary gene sets that are members of these clusters (membership Z score > 2). Of these, the figure shows the 59 genes that appear in a figure or table of a well-known recent textbook[8]; these are genes whose functions are known and considered important for neuronal physiology.

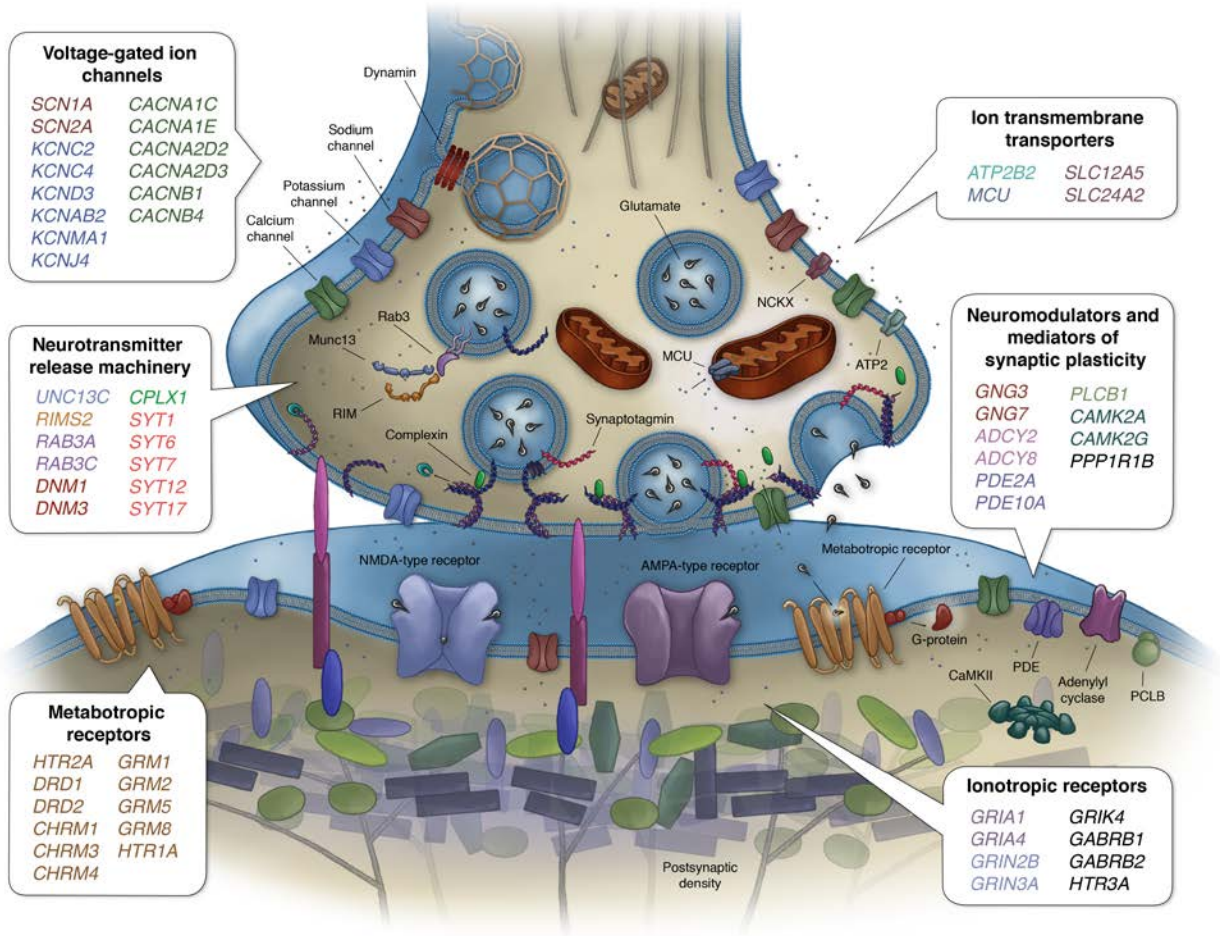
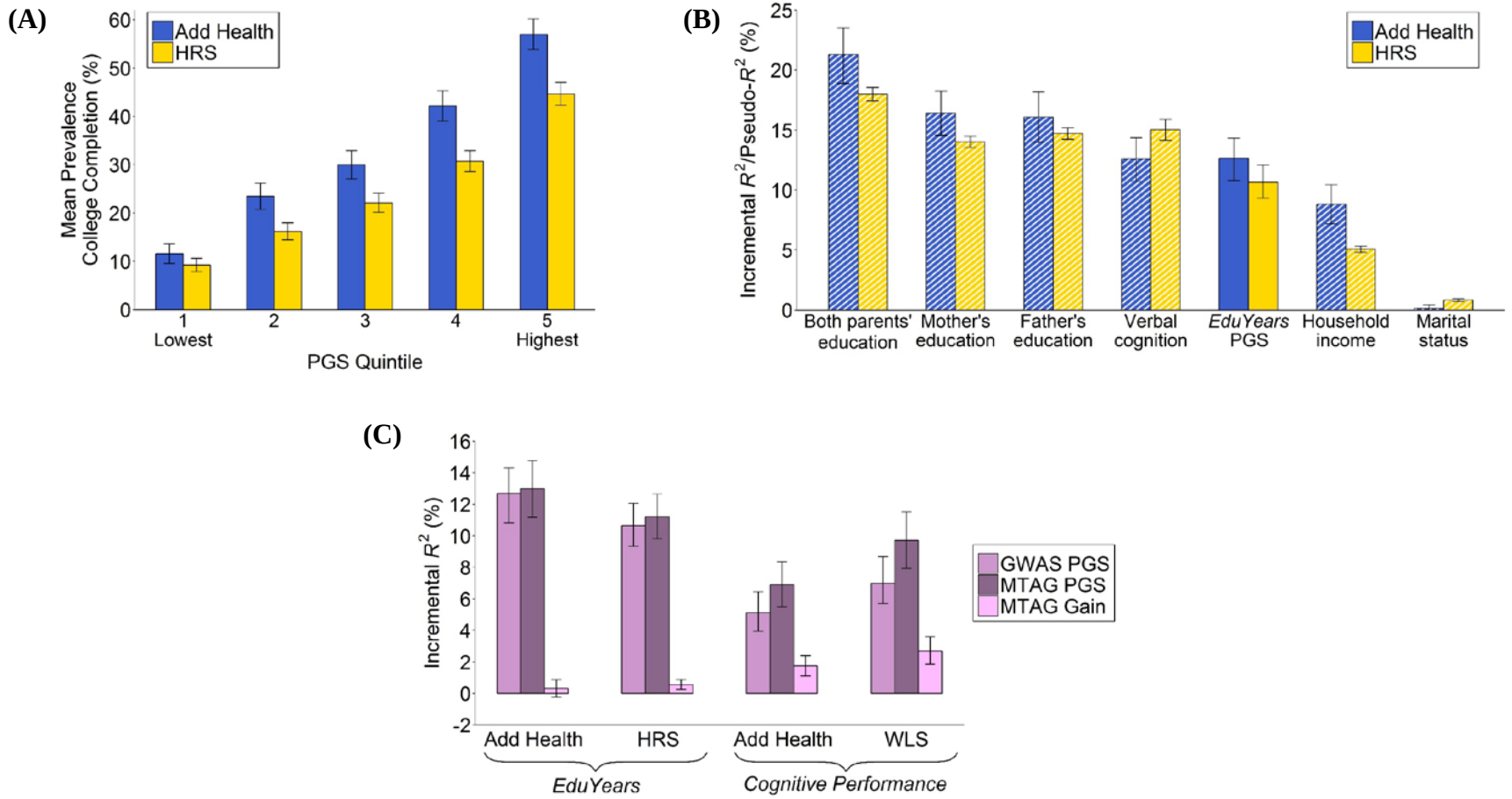


Fig. 4. Prediction Accuracy. (A) Mean prevalence of college completion by *EduYears* PGS quintile. **(B)** Incremental R^2 of the *EduYears* PGS on *EduYears* compared to that of other variables. (**C**) Incremental R^2 of the PGS for *EduYears* and *Cognitive Performance* constructed from the respective GWAS or MTAG summary statistics. Error bars show bootstrapped 95% confidence intervals with 1000 iterations each.



ONLINE METHODS

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

GWAS of educational attainment. The *EduYears* meta-analysis was performed by combining 59 cohort-level results files from Okbay et al.[3] with new results from 8 cohorts that did not contribute to Okbay et al.[3] and 4 cohorts that did contribute to it but genotyped additional samples after the publication of that article, enabling them to contribute analyses based on larger samples. In what follows, we focus on the 12 new results files.

Genome-wide association analyses were performed at the cohort level. Genotyping was performed using a range of common, commercially available genotyping arrays. Imputation was conducted using a reference panel from either the 1000 Genomes Project[26] or a larger panel subsequently released by Haplotype Reference Consortium[27]. The *EduYears* phenotype was constructed by mapping each major educational qualification that can be identified from the cohort's survey measure to an ISCED category, and imputing a years-of-education equivalent for each ISCED category.

Cohorts were asked to estimate, for each SNP, a linear regression of *EduYears* on the allele dose of the SNP, a vector of the first ten principal components of the variance-covariance matrix of the genotypic data, a vector of standardized controls, including a third-order polynomial in year of birth, an indicator for being female, and their interactions; and a vector of study-specific controls. All analyses were restricted to European-ancestry individuals that passed the cohort's quality control and whose *EduYears* was measured at an age of at least 30.

Prior to meta-analyses, the quality-control protocol and filters described in Okbay et al.[3] were applied to the results files. Detailed cohort descriptions for the 12 new results files, information about cohort-level phenotype measures, genotyping and imputation procedures, and association analyses are shown in **Supplementary Tables 1.1**-**1.4**. For analogous information on the 59 results files from Okbay et al., see Supplementary Tables 1.1, 1.3-1.5 of Okbay et al.[3].

We performed a sample-size-weighted meta-analysis of 71 cleaned cohort-level results files using the METAL software[28], applying a minimum sample-size filter of 500,000. Subsequently, we inflated the standard errors using the square root of the estimated intercept from an LD Score regression ($\sqrt{1.11}$) to adjust for non-independence.

To select independent genome-wide significant SNPs from our results, we used the standard clumping algorithm implemented in PLINK[29], clumping together all SNPs on a chromosome with a pairwise LD exceeding $R^2 = 0.1$. We estimate LD using the 1000 Genomes Project phase 3 genotyping data[26] that have been described in detail in a previous publication[30]. **Supplementary Table 1.5** shows the association results for the 1,271 approximately independent SNPs that reached genome-wide significance.

Sensitivity of results to alternative locus definitions. We conducted two follow-up analyses using a sample of approximately unrelated individuals (pairwise relatedness < 0.025) of European ancestry from *UKB* (N = 405,519). First, we re-ran our main clumping algorithm using

the *UKB* reference sample and identified 1,223 approximately independent SNPs at genome-wide significance. Second, we performed a conditional and joint multiple-SNP analysis (COJO)[31], aiming to explore secondary associations that may have been lost from clumping. Before analysis, we applied SNP filters recommended in the original COJO paper[31]. We performed COJO using the implementation found in the GCTA software (Version 1.90.0 beta). Model selection was performed using the stepwise selection process outlined in the original COJO paper[31] in which SNPs from across the genome are iteratively added to the model. We set the LD window to 100 Mb.

Our COJO analysis identified 765 variants at genome-wide significance. To help interpret this number, we also applied our clumping algorithm to the ~4.9M SNPs that passed COJO filters and found 1,053 lead SNPs when *UKB* is used as the reference sample. We classified each of the 765 COJO hits as either primary or secondary by applying our clumping algorithm to the list of COJO variants, again using *UKB* as our reference sample and an R^2 threshold of 0.1. We found that our clumping algorithm eliminated 60 SNPs from the original list of 765 COJO hits (pairwise $R^2 > 0.1$ with at least one COJO variant). We call these 60 variants secondary associations and the remaining 705 variants primary associations (**Supplementary Table 1.6**).

Estimating the Distribution of Effect Size. To calculate the 5th, 50th, and 95th percentile of the effect-size distribution of our lead SNPs, we first calculate the posterior distribution of each SNP's effect size using the winner's curse adjustment described in Section 1.8 in the Supplementary Methods of Okbay et al.[3] We then drew 10 simulated effect sizes from the posterior distribution for each lead SNP. To estimate the 5th percentile of the effect-size distribution, we used the 5th percentile of the simulated effect sizes. We calculated the 50th and 95th percentiles in the same way.

Replication of Okbay et al. Lead SNPs. We conducted a replication analysis of the 162 lead SNPs identified at genome-wide significance in Okbay et al.'s[3] pooled (discovery and replication) meta-analysis (N = 405,073). Of the 162 SNPs, 158 pass quality-control filters in our updated meta-analysis. To examine their out-of-sample replicability, we calculated approximate Z-statistics from the subsample of our data (N = 726,808) that was not included in Okbay et al. Let the Z-statistics of association from, respectively, Okbay et al., the new data, and our final EA3 meta-analysis, be denoted by Z_1, Z_2 and Z. Since our meta-analysis used sample-size weighting[28], Z_2 is implicitly defined by:

$$Z = \sqrt{\frac{N_1}{N}} Z_1 + \sqrt{\frac{N_2}{N}} Z_2,$$

where SNP subscripts have been dropped for notational convenience and N's are sample sizes.

Of the 158 SNPs, we find that 154 have matching signs in the new data (for the remaining four SNPs, the estimated effect is never statistically significant at $P < 0.10$). Of the 154 SNPs with matching signs, 143 are significant at $P < 0.01$, 119 are significant at $P < 10^{-5}$, and 97 are significant at $P < 5\times10^{-8}$. The replication results are shown graphically in **Supplementary Figure 1.4.**

To help interpret these results, we used the statistical framework from Section 1.8 in the Supplementary Methods of Okbay et al.[3] to calculate the expected replication record under the null that all 158 SNPs are true associations. The theoretical projections are based on shrinkage parameters estimated from Okbay et al. summary statistics (used to adjust the Okbay et al. effect sizes for winner's curse): $(\hat{\tau}^2, \hat{\pi}) = (5.02 \times 10^{-6}, 0.33)$.

Tests for Population Stratification. We used several methods to test for population stratification. First, we compared the estimated intercept from LD Score regression[5] to the average χ^2 test statistic among HapMap3 SNPs to produce an estimate of the share of inflation in the test statistics that is due to stratification (**Supplementary Figure 1.2** and **Supplementary Table 1.12**).

Second, we conducted within-family association analyses on a sample of 22,135 sibling pairs from *STR-Twingene, STR-SALTY, UKB,* and *WLS*. For each cohort, we standardized *EduYears* within each cohort and then residualized this variable using a vector of controls. We then regressed the sibling difference in the residuals on the sibling difference in genotype. We restricted analyses to SNPs with minor allele frequency above 5% in each of the sibling cohorts and meta-analyzed the cohort-level results using an inverse-variance-weighted meta-analysis. We used these within-family summary statistics alongside a discovery GWAS conducted in an independent sample to conduct two sets of tests: a sign test and a within-family regression test.

For the sign test, we followed Okbay et al.[30] to compare the signs of the within-family estimates to the signs of the discovery GWAS. We benchmarked our observed fraction of concordant signs against several null hypotheses: that the GWAS results are entirely driven by stratification (i.e., an expected sign concordance of 50%), that the GWAS results contain no stratification (but are corrected for winner's curse), and that the GWAS results are biased upward by assortative mating. We constructed these null distributions via simulation, and we conducted one-sided binomial tests where the alternative hypothesis is that the observed sign concordance falls short of each benchmark. We conducted this test for sets of approximately independent SNPs selected at the *P* value thresholds 5×10^{-8}, 5×10^{-5}, and 5×10^{-3} (**Supplementary Table 2.1** and **Supplementary Figure 2.1**).

We also performed a regression-based comparison of the within-family estimates and the GWAS estimates (**Supplementary Table 2.2** and **Supplementary Figure 2.2**). Further details on our within-family analyses, including a derivation for the assortative mating correction, can be found in the **Supplementary Note.**

Joint F-test of Heterogeneity. For the 1,271 lead SNPs, a test of homogenous effects across cohorts fails to reject the null at the Bonferroni-adjusted *P* value threshold of 0.05/1,271 for all SNPs barring one. We generated an omnibus test statistic for heterogeneity by summing the Cochran Q-statistics for heterogeneity across all 1,271 lead SNPs[32]. Because the software used for meta-analysis does not report Q-statistics, we inferred these values based on the reported heterogeneity *P* values. To do so, we treated each lead SNP as if it were available for each of the 71 cohorts in the meta-analysis, which implies that the Q-statistic for each lead SNP has a χ^2 distribution with 70 degrees of freedom. The sum of these Q-statistics is therefore

(approximately) χ^2-distributed with $70 \times 1{,}271 = 88{,}970$ degrees of freedom. This gave us an omnibus Q-statistic of 91,830, with corresponding *P* value equal to 9.68×10^{-12}.

Mean genetic correlation. To calculate the mean genetic correlation of *EduYears* across pairs of cohorts included in the meta-analysis, we first estimated the genetic correlation of *EduYears* across all unique pairs of cohorts with non-negative heritability estimates (**Supplementary Table 3.1**). To do so, we used bivariate LD Score regression[33] implemented by the LDSC software with a European reference population, filtered to HapMap3 SNPs. The estimated genetic correlation of *EduYears* between each of our 933 pairs of cohorts is shown in **Supplementary Table 3.2**.

We report the inverse-variance-weighted mean of the genetic-correlation estimates. The genetic correlation across pairs of cohorts will be correlated across all observations that share one of their cohorts in common. Therefore, to obtain correct standard errors, we used the node-jackknife variance estimator described by Cameron and Miller[34].

As detailed in **Supplementary Note**, we also estimated the amount of variation in SNP heritability of *EduYears* across cohorts, and we conducted analyses to assess the extent to which we can predict variation in SNP heritability and genetic correlation of *EduYears* based on several observable cohort characteristics (**Supplementary Tables 3.3** and **3.4**).

X chromosome. We performed association analyses of SNPs on the X chromosome in our two largest cohorts, *UKB* (*N* = 329,358) and *23andMe* (*N* = 365,536). In both cohorts, the association analyses were performed on a pooled male-female sample with male genotypes coded 0/2.

Except for this allele coding in males, all major aspects of the *23andMe* analysis were identical to those described for the autosomal analyses; see **Supplementary Tables 1.2-1.4** for details.

Imputed genotypes for the X chromosome were not included in the data officially released by *UKB*. We therefore imputed the data ourselves using the 1000 Genomes Project[26] as our reference panel. The *UKB* analyses were conducted in a sample of conventionally unrelated European-ancestry individuals, yielding a smaller sample size than the autosomal *UKB* analyses (**Supplementary Table 4.1**).

Both sets of association results underwent the same set of quality-control filters as the autosomal analyses prior to meta-analysis. Additionally, we dropped a small number of SNPs with male-female allele frequency differences above 0.005 in *UKB*. The meta-analysis was conducted in METAL[28], using sample-size weighting. Only SNPs that were present in both results files were used. To adjust the test statistics for bias, we inflated the standard errors by the LD Score regression intercept from our main autosomal analysis ($\sqrt{1.113}$).

Applying our clumping algorithm, we found 10 approximately independent SNPs at genome-wide significance (**Supplementary Table 4.2**). **Supplementary Figure 4.1** shows Manhattan and quantile-quantile plots from the meta-analysis.

Heritability of the X Chromosome and Dosage Compensation. SNP heritability for males and females from SNPs on the X chromosome was estimated solving for h_i^2 in

$$\mathrm{E}[\chi_i^2] = 1 + \frac{N_i h_i^2}{M_{\mathrm{eff}}},$$

where $i \in \{m, f\}$ indicates males or females, $\mathrm{E}[\chi_i^2]$ is the expected χ^2 statistic, h_i^2 is the SNP heritability for the X chromosome, N_i is the GWAS sample size, and M_{eff} is the effective number of loci (which is assumed to be the same in males and females).

We use $\gamma = h_m^2/h_f^2$ to denote the dosage compensation ratio. The ratio takes on a value between 0.5 (zero dosage compensation) and 2 (full dosage compensation). It is estimated as

$$\hat{\gamma} = \frac{(\hat{\chi}_m^2 - 1)N_f}{(\hat{\chi}_f^2 - 1)N_m},$$

where $\hat{\chi}_i^2$ is the mean χ^2 statistic.

Biological Annotation. We used DEPICT[7] (downloaded February 2016 from https://github.com/perslab/depict) to identify the tissues/cell types where the causal genes are strongly expressed, detect enrichment of gene sets, and prioritize likely causal genes. We ran DEPICT as described previously[3] with the following exceptions: we used 37,427 human Affymetrix HGU133a2.0 platform microarrays[7], discarded gene sets that were not well reconstituted[35], and relaxed the significance threshold for defining a matching SNP in the simulated null GWAS from 5×10^{-4} to 5×10^{-3}.

The list of genes we used is described in **Supplementary Table 5.3.** For more details on how the gene sets were constructed, see **Supplementary Note.** In addition to the results presented in the main text, we examined how enrichment of gene sets differs across phenotypes (**Supplementary Table 5.7**) and which functional systems are *least* implicated by DEPICT (**Supplementary Table 5.6**).

We tested the robustness of our DEPICT results using the bioinformatics tools MAGMA[36] and PANTHER[37,38]. For MAGMA, we used the "multi=snp-wise" option, mapping a SNP to a gene if it resides within the gene boundaries or 5kb of either endpoint. We estimated LD using a reference panel of Europeans in 1000 Genomes phase 3, and we defined a gene as significant if its joint *P* value falls below the threshold corresponding to FDR < 0.05 (**Supplementary Table 5.4**). For PANTHER, we used the binomial overrepresentation test with the DEPICT-prioritized genes as input (**Supplementary Table 5.8**).

We also used stratified LD Score regression to partition the heritability of the trait between SNP-level annotation categories. We devised three novel annotation types, described in the **Supplementary Note**. We first constructed a model using baseline annotations (**Supplementary Table 5.12**) and tested the heritability enrichment of various SNP-level annotations (**Supplementary Table 5.13** and **Supplementary Figure 5.4**), developmental stages (**Supplemental Table 5.11**), and cell types (**Supplementary Table 5.2**). We also applied LD Score regression to DEPICT-reconstituted gene sets (**Supplementary Table 5.9**) and binary gene sets (**Supplementary Table 5.10** and **Supplementary Figure 5.3**).

Finally, we used the tool CAVIARBF[12,39] to identify candidate causal SNPs. We used the 74 baseline annotations employed by stratified LD Score regression as well as 451 annotations from Pickrell's (2014) list[40]. We applied a MAF filter of 0.01 and a sample-size filter of 400,000, and we only considered SNPs within a 50-kb radius of a lead SNP. We computed exact Bayes factors by averaging over prior variances of 0.01, 0.1, and 0.5; we set the sample size to the mean sample size of our considered SNPs; and we added 0.2 to the main diagonal of the LD matrix because we used a reference panel for LD estimation. To incorporate annotations, we used the elastic net setting with parameters selected via 5-fold cross-validation. The resulting annotation effect sizes and list of candidate causal SNPs are given in **Supplementary Tables 5.14** and **5.15.** Regional association plots of four noteworthy candidates are shown in **Supplementary Figure 5.5.**

Defining Newly Prioritized Genes. We distinguished "newly prioritized genes" from those prioritized by DEPICT in Okbay et al.[3], as in **Figure 2B.** As in this previous work, we used expression data from the BrainSpan Developmental Transcriptome[25] and calculated the average expression in the brain of DEPICT-prioritized *EduYears* genes as a function of developmental stage.

Polygenic Prediction Methods. Prediction analyses were performed using the National Longitudinal Study of Adolescent to Adult Health (Add Health), the Health and Retirement Study (HRS), and the Wisconsin Longitudinal Study (WLS). Polygenic scores were constructed using HapMap3 SNPs that meet the following conditions: (i) the variant has a call rate greater than 98% in the prediction cohort; (ii) the variant has a minor allele frequency (MAF) greater than 1% in the prediction cohort; and (iii) the allele frequency discrepancy between the meta-analysis and the prediction cohort does not exceed 0.15. To calculate the SNP weights we use the software package LDpred[16], assuming a fraction of causal variants equal to 1, and then we construct the scores in PLINK.

All prediction exercises were performed with OLS regression of a phenotype on our score and a set of controls consisting of a full set of dummy variables for year of birth, an indicator variable for sex, a full set of interactions between sex and year of birth, and the first 10 principal components of the variance-covariance matrix of the genetic data.

Our measure of prediction accuracy is the incremental R^2. To calculate this value, we first regress a phenotype on our set of controls without the polygenic score. Next, we re-run the same regression but with the score included as a regressor. For quantitative phenotypes, our measure of predictive power is the change in R^2. For binary outcomes, we calculated the incremental pseudo-R^2 from a Probit regression. 95% confidence intervals around the incremental R^2's are bootstrapped with 1000 repetitions (**Supplementary Table 6.1** and **Supplementary Figures 6.1**, **6.2**, **6.5**, and **6.6)**.

Prediction of Other Phenotypes. In addition to *EduYears*, we also used our polygenic score to predict a number of other phenotypes. In the HRS and Add Health, we analyzed three binary variables related to educational attainment: (i) High School Completion, (ii) College Completion, and (iii) Grade Retention (i.e., retaking a grade).

In additional analyses in Add Health, we predicted an augmented version of the Peabody Picture Vocabulary test, measured when participants were 12–20 years old. In this test, an interviewer reads a word aloud, and a respondent selects the illustration that best fits the word's meaning. Eighty-seven items were included on this computer-adapted test, and Peabody scores were age-standardized. We also predicted a number of Grade Point Average variables from the third wave of Add Health, when transcripts were collected from respondents' high schools. From the transcripts, grade point averages (GPAs) are calculated using the common United States 0.0 to 4.0 range, both for Overall GPA and for subject-specific GPAs. We analyzed Overall GPA, Math GPA, Science GPA, and Verbal GPA, controlling for high school fixed effects.

In additional analyses in the HRS, we predicted several cognitive phenotypes. Total Cognition is the sum of four cognitive measures common across waves 3 through 10: an immediate word recall task, a delayed word recall task, a naming task, and a counting task, with a total score ranging from 0 to 35. Verbal Cognition measures the subject's ability to define five words. Each definition supplied is rated as incorrect (0), partially correct (1) or completely correct (2), resulting in a total score ranging from 0 to 10. To evaluate changes over time, we also studied wave-to-wave changes in Total Cognition and Verbal Cognition, $(x_t - x_{t-1})$. Our next cognitive outcome, Alzheimer's, is an indicator variable equal to one for subjects who report having been diagnosed with Alzheimer's disease, and 0 otherwise. Since the HRS data are longitudinal, the unit of analysis for our 4 cognitive outcomes is a person-year. For these analyses, because individual *i* took the cognitive tests at different ages, in our set of controls we replaced our person-specific age variable with age at assessment (which differs for individual *i* across the cognitive outcomes); we also clustered all standard errors at the person level.

In the WLS, we predicted cognition from a respondent's raw score on a Henmon-Nelson test of mental ability, a 30-minute multiple-choice test that consists of 90 individual verbal or quantitative items.

For all of these additional prediction exercises, results are shown in **Supplementary Table 6.1** and depicted in **Figure 4A** and **Supplementary Figures 6.1** and **6.4**.

Benchmarking the Predictive Power of the *Eduyears* Polygenic Score. To benchmark our score's predictive power, we compared its predictive power to the predictive power of other common variables: mother's education, father's education, both mother's and father's education, verbal cognition, household income, and a binary indicator for marital status. For each variable, we calculated the variable's incremental R^2 using the same procedures as those described above, with the same set of control variables. (For "mother's and father's education," we calculated the incremental R^2 from adding both variables as regressors.) The results of this analysis are shown in **Supplementary Table 6.2A** and depicted in **Figure 4B** and **Supplementary Figure 6.7**.

We also evaluated the attenuation in the incremental R^2 of the polygenic score in predicting *EduYears* using various sets of available demographic control variables one at a time: marital status, household income, mother's education, and father's education. We next controlled for both mother's and father's education, and finally we controlled for the full set of demographic controls. At each addition of a control or set of controls, we calculated the change in incremental R^2. The results of this analysis are shown in **Supplementary Table 6.2B** and **Supplementary Figure 6.7.**

GWAS of Cognitive Performance, Math Ability and Highest Math. The GWAS of *Math Ability* (N = 564,698) and *Highest Math* (N = 430,445) phenotypes were conducted exclusively among research participants of the personal genomics company *23andMe* who answered survey questions about their mathematical background. In our analyses of cognitive performance, we combined a published study of general cognitive ability (N = 35,298) conducted by the COGENT consortium[41] with new genome-wide association analyses of cognitive performance in the UK Biobank (N = 222,543). The phenotype measures are described in detail in **Supplementary Table 1.7**. Our new genome-wide analyses of *CP* in *UKB*, and *Math Ability* and *Highest Math* in *23andMe* were conducted using methods identical to those for *EduYears* in *UKB* and *23andMe* respectively (**Supplementary Table 1.4**).

For *Cognitive Performance (CP)*, we conducted a sample-size-weighted meta-analysis (N = 257,841), imposing a minimum-sample-size filter of 100,000. We similarly applied minimum-sample-size filters to the *Math Ability* (N > 500,000) and *Highest Math* (N > 350,000) results. We adjusted the test statistics using the estimated intercepts from LD Score regressions (1.073 for *Math Ability*, 1.105 for *Highest Math*, and 1.046 for *CP*). The summary statistics underwent quality control using the same procedures applied to *EduYears* results files.

The lists of approximately independent genome-wide significant SNPs were obtained by applying the same clumping algorithm used in the *EduYears* analyses (**Supplementary Tables 1.8-1.10**). Manhattan plots from the analyses are shown in **Supplementary Figures 1.5-1.7.**

MTAG of Cognitive Performance, Math Ability and Highest Math. We performed a joint analysis of our GWAS results on *EduYears*, *Cognitive Performance*, *Math Ability*, and *High Math* using MTAG[18]. **Supplementary Table 1.11** shows moderately high pairwise genetic correlations, ranging from 0.51 to 0.85, which motivate the multivariate analysis.

We applied the MTAG-recommended filters to the summary statistics, dropping (i) SNPs with minor allele frequency below 1% or (ii) SNPs with sample sizes below a cutoff (66.6% of the 90th percentile), leaving approximately 7.1 million SNPs found in all four results files. **Supplementary Table 1.13** provides the increases in effective sample size from using MTAG for each set of GWAS results.

Supplementary Table 1.14 lists all the SNPs that reach genome-wide significance in the MTAG analysis. **Supplementary Figures 1.8-1.11** show inverted Manhattan plots that compare the MTAG and GWAS results, restricted to the set of SNPs that pass MTAG filters.

Polygenic scores were constructed from MTAG results using the same procedures as for the GWAS results. **Supplementary Figure 6.8** and **Supplementary Tables 6.3** and **6.4** compare the predictive power of scores constructed from MTAG results in the Add Health and WLS cohorts (see **Supplementary Note** section 6.5 for details).

To examine the credibility of the MTAG-identified loci of our lowest-powered GWAS, *Cognitive Performance*, we conducted a replication analysis. We re-ran MTAG with GWAS results that exclude COGENT cohorts, and we used the COGENT meta-analysis as our replication sample. In addition to applying the MTAG filters above, we limited the analysis to SNPs for which the COGENT results file contains summary statistics based on analyses of at least 25,000 individuals. The MTAG-identified loci for *Cognitive Performance* from our

restricted sampled are reported in **Supplementary Table 1.15**. We calculated the expected replication record of the MTAG results, given sampling variation. The observed replication record in the COGENT meta-analysis is not much below the expected record.

ACCESSION CODES

Upon publication (or earlier, if permitted by the editor), summary statistics will be posted at www.thessgac.org/data. For analyses that include data from 23andMe, only up to 10,000 SNPs can be reported. Complete summary statistics for the *EduYears* GWAS omitting 23andMe are available. Only the lead SNPs are reported for the GWAS results for *Cognitive Performance*, *Math Ability*, and *Highest Math* and for the MTAG results of *EduYears*, *Cognitive Performance*, *Math Ability*, and *Highest Math*. For *EduYears* GWAS results that include 23andMe, clumped results for the 4,429 SNPs with the lowest *P* values are available. This number of SNPs was chosen such that the total number of SNP association results that include data from 23andMe is exactly 10,000.

METHODS-ONLY REFERENCES

26. The 1000 Genomes Project Consortium *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491,** 56–65 (2012).

27. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48,** (2016).

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

29. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4,** 1–16 (2015).

30. Okbay, A. *et al.* Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* **48,** 624–633 (2016).

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

32. Cochran, W. G. The Combination of Estimates from Different Experiments. *Biometrics* **10,** 101 (1954).

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

34. Cameron, A. C. & Miller, D. Robust inference with dyadic data. *mimeo* (2014). doi:10.1201/b10440

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

36. de Leeuw, C. A. *et al.* MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput. Biol.* **11,** e1004219 (2015).

37. Liu, J. Z. *et al.* A versatile gene-based test for genome-wide association studies. *Am. J. Hum. Genet.* **87,** 139–145 (2010).

38. Mi, H., Muruganujan, A., Casagrande, J. T. & Thomas, P. D. Large-scale gene function analysis with the PANTHER classification system. *Nat. Protoc.* **8,** 1551–1566 (2013).

39. Chen, W. *et al.* Fine mapping causal variants with an approximate Bayesian method using marginal test statistics. *Genetics* **200,** 719–736 (2015).

40. Pickrell, J. K. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am. J. Hum. Genet.* **94,** 559–573 (2014).

41. Trampush, J. *et al.* GWAS meta-analysis reveals novel loci and genetic correlates for general cognitive function: a report from the COGENT consortium. *Nat. Publ. Gr.* (2017). doi:10.1038/mp.2016.244