GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment


References (23–175)

Supplementary Materials
www.sciencemag.org/cgi/content/full/science.1235488/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S22
Tables S1 to S27
Rietveld et al. Page 2

1Department of Applied Economics, Erasmus School of Economics, Erasmus University Rotterdam, 3000 DR Rotterdam, The Netherlands 2Department of Epidemiology, Erasmus Medical Center, Rotterdam 3000 CA, The Netherlands 3Queensland Institute of Medical Research, 300 Herston Road, Brisbane, Queensland 4006, Australia 4Institute for Behavioral Genetics, University of Colorado Boulder, Boulder, CO 80309–0447, USA 5University of Queensland Diamantina Institute, The University of Queensland, Princess Alexandra Hospital, Brisbane, Queensland 4102, Australia 6Estonian Genome Center, University of Tartu, Tartu 51010, Estonia 7School of Psychology, University of Queensland, Brisbane, Queensland 4072, Australia 8Department of Genetics, University Medical Center Groningen, University of Groningen, 9713 GZ Groningen, The Netherlands 9Queensland Brain Institute, The University of Queensland, Brisbane, Queensland 4072, Australia 10Department of Biological Psychology, VU University Amsterdam, 1081 BT Amsterdam, The Netherlands 11Department of Psychiatry, Washington University School of Medicine, St. Louis, MO 63110, USA 12Institute of Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Germany 13Department of Epidemiology, University Medical Center Groningen, University of Groningen, 9700 RB Groningen, The Netherlands 14Genetic Epidemiology Unit, Department of Epidemiology, Erasmus Medical Center, Rotterdam 3000 CA, the Netherlands 15Heart and Vascular and Lerner Research Institutes, Cleveland Clinic, Cleveland, OH 44195, USA 16Institute for Community Medicine, University Medicine Greifswald, Greifswald 17489, Germany 17Samuel Lunenfeld Research Institute, Mount Sinai Hospital, University of Toronto, Toronto, Ontario M5G 1X5, Canada 18Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI 48109–2029, USA 19Division of Biostatistics, University of Minnesota, Minneapolis, MN 55455, USA 20Rush University Medical Center, Rush Alzheimer’s Disease Center, Chicago, IL 60612, USA 21Centre for Cognitive Aging and Cognitive Epidemiology, The University of Edinburgh, Edinburgh EH8 9JZ, Scotland, UK 22Department of Functional Genomics, VU University Amsterdam and VU Medical Center, 1081 HV Amsterdam, The Netherlands 23Machine Learning Group, Intelligent Systems, Institute for Computing and Information Sciences, Faculty of Science, Radboud University Nijmegen, 6500 GL Nijmegen, The Netherlands 24Institute for Molecular Medicine Finland, University of Helsinki, Helsinki 00014, Finland 25Public Health Genomics Unit, Department of Chronic Disease Prevention, The National Institute for Health and Welfare, Helsinki 00014, Finland 26California Pacific Medical Center Research Institute, San Francisco, CA 94107–1728, USA 27Department of Genes and Environment, Division of Epidemiology, Norwegian Institute of Public Health, Nydalen, N-0403 Oslo, Norway 28Molecular Epidemiology, Department of Medical Sciences, Uppsala University, 751 85 Uppsala, Sweden 29Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, 751 23 Uppsala, Sweden 30Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, 171 77 Stockholm, Sweden 31Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK
Else Kroener-Fresenius-Centre for Nutritional Medicine, Technische Universität München, 81675 Munich, Germany 33Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Germany 34Department of Neurology, Erasmus Medical Center, Rotterdam 3000 CA, The Netherlands 35Department of Obstetrics and Gynecology, Institute of Public Health, Sahlgrenska Academy, Sahlgrenska University Hospital, Gothenburg, 413 45, Sweden 36Centre for Population Health Sciences, The University of Edinburgh, Edinburgh EH8 9AG, UK 37Institute of Health Sciences, University of Oulu, Oulu 90014, Finland 38Biocenter Oulu, University of Oulu, Oulu 90014, Finland 39Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK 40Department of Medicine, University of Split, 21000 Split, Croatia 41Department of Medical Genetics, University of Lausanne, 1005 Lausanne, Switzerland 42Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland 43Institute of Behavioral Sciences, University of Helsinki, Helsinki 00014, Finland 44Department of Epidemiology & Prevention, Division of Public Health Sciences, Wake Forest University Health Sciences, Winston-Salem, NC 27157–1063, USA 45Department of Biostatistical Sciences, Division of Public Health Sciences, Wake Forest University Health Sciences, Winston-Salem, NC 27157–1063, USA 46Division for Neurogeriatrics, Department of Neurology, Medical University of Graz, Graz 8036, Austria 47School of Social and Community Medicine, University of Bristol, Bristol BS8 2PR, UK 48Institute of Social and Preventive Medicine, Lausanne University Hospital, 1005 Lausanne, Switzerland 49National Institute on Aging, National Institutes of Health, Baltimore, MD 20892, USA 50Hunter Medical Research Institute and Faculty of Health, University of Newcastle, Newcastle, NSW 2308, Australia 51Division of General Neurology, Department of Neurology, General Hospital and Medical University of Graz, Graz 8036, Austria 52Department of Psychiatry, VU University Medical Center, 1081 HL Amsterdam, The Netherlands 53Department of Twin Research and Genetic Epidemiology, King’s College London, London SE1 7EH, UK 54Institute of Molecular Biology and Biochemistry, Medical University of Graz, Graz 8036, Austria 55Icelandic Heart Association, Kopavogur 201, Iceland 56Department of Medicine, University of Iceland, Reykjavik 101, Iceland 57College of Medicine, Florida State University, Tallahassee, FL 32306–4300, USA 58Institute of Epidemiology and Social Medicine, University of Muenster, 48129 Muenster, Germany 59Department of Psychology, University of Tartu, Tartu 50410, Estonia 60Geriatric Unit, Azienda Sanitaria Firenze, 50125 Florence, Italy 61Department of Internal Medicine, University Hospital, 1011 Lausanne, Switzerland 62Department of Economics, Harvard University, Cambridge, MA 02138, USA 63Department of Health Sciences, Community & Occupational Medicine, University Medical Center Groningen, 9700 AD Groningen, The Netherlands 64Department of Psychology, Union College, Schenectady, NY 12308, USA 65Istituto di Ricerca Genetica e Biomedica, CNR, Monserrato, 09042, Cagliari, Italy 66Dipartimento di Scienze Biomediche, Università di Sassari, 07100 SS, Italy 67Department of Health Sciences Research, Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN 55905, USA 68Program in Translational Neuropsychiatric Genomics, Department of Neurology, Brigham and Women’s Hospital, Boston, MA 02115, USA 69School of Public Policy, University College London, London WC1H 9QU, UK 70Centre for Economic Performance, London School of Economics, London WC2A 2AE, UK 71Department of Psychology, The University of Edinburgh, Edinburgh EH8 9JZ, Scotland, UK 72Department of Nutrition and Dietetics, Harokopio University of Athens, Athens 17671, Greece 73LifeLines Cohort Study, University of Groningen, University Medical Center Groningen, 9700 RB Groningen, The Netherlands 74Department of General Practice and Primary Health Care, University of Helsinki, Helsinki 00014, Finland 75Unit of General Practice, Helsinki University Central Hospital, Helsinki 00280, Finland 76Folkhälsan Research Center, Helsinki 00250, Finland 77Vaasa Central Hospital, Vaasa 65130, Finland 78MRC Centre for Causal Analyses in Translational Epidemiology, School of Social and Community Medicine, University of Bristol, Bristol BS8 2PR, UK 79Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI 48106, USA 80Division of Biology and Biomedical Sciences, Washington University, St. Louis, MO 63110–
1093, USA 81Faculty of Behavioral and Social Sciences, University of Groningen, 9747 AD Groningen, The Netherlands 82Department of Psychology, University of Minnesota, Minneapolis, MN 55455–0344, USA 83Institute of Health Economics and Health Care Management, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Germany 84Hannover Unified Biobank, Hannover Medical School, 30625 Hannover, Germany 85Department of Epidemiology and Biostatistics, MRC-HPA Centre for Environment and Health, Imperial College London, London W2 1PG, UK 86Unit of Primary Care, Oulu University Hospital, Oulu 90020, Finland 87Department of Children and Young People and Families, National Institute for Health and Welfare, Oulu 90101, Finland 88Department of Clinical Physiology, Tampere University Hospital and University of Tampere School of Medicine, Tampere 33520, Finland 89Department of Public Health, University of Helsinki, 00014 Helsinki, Finland 90Department of Mental Health and Substance Abuse Services, National Institute for Health and Welfare, 00300 Helsinki, Finland 91Ontario Institute for Cancer Research, Toronto, Ontario M5G 0A3, Canada 92Department of Clinical Chemistry, Finnlab Laboratories, Tampere University Hospital, Tampere 33520, Finland 93Human Genetics, Genome Institute of Singapore, Singapore 138672, Singapore 94Institute of Human Genetics, Helmholtz Centre Munich, German Research Center for Environmental Health, 85764 Neuherberg, Germany 95Department of Health, Functional Capacity and Welfare, National Institute for Health and Welfare, Helsinki 00271, Finland 96Western Australia Sleep Disorders Research Institute, Sir Charles Gairdner Hospital, Perth, Western Australia 6009, Australia 97Department of Medicine, University of Toronto, Toronto, Ontario M5S 1A8, Canada 98Women’s College Research Institute, University of Toronto, Toronto, Ontario M5G 1N8, Canada 99Department of Medical Genetics, University of Helsinki, 00014 Helsinki, Finland 100Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku 20520, Finland 101Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku 20520, Finland 102Department of Internal Medicine, Erasmus Medical Center, Rotterdam 3000 CA, The Netherlands 103Department of Economics, University of Minnesota, Minneapolis, MN 55455–0462, USA 104Chronic Disease Epidemiology Unit, Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki 00271, Finland 105School of Oral and Dental Sciences, University of Bristol, Bristol BS1 2LY, UK 106Alzheimer Scotland Dementia Research Centre, The University of Edinburgh, Edinburgh EH8 9JZ, Scotland, UK 107Department of Computer Science, University of California, Los Angeles, CA 90095, USA 108Department of Economics, Oulu Business School, University of Oulu, Oulu 90014, Finland 109Interfaculty Institute for Genetics and Functional Genomics, Department of Functional Genomics, University Medicine Greifswald, Greifswald 17487, Germany 110Department of Child and Adolescent Psychiatry, Erasmus Medical Center, 3000 CB Rotterdam, The Netherlands 111Division of Welfare and Health Promotion, National Institute for Health and Welfare, Helsinki 00271, Finland 112Department of Medicine, Turku University Hospital, Turku 20520, Finland 113Institute of Medical Informatics, Biometry and Epidemiology, Chair of Epidemiology, Ludwig-Maximilians-Universität, 81377 Munich, Germany 114Klinikum Grosshadern, 81377 Munich, Germany 115Institute of Epidemiology I, Helmholtz Zentrum München, German Research Centre for Environmental Health, 85764 Neuherberg, Germany 116Department of Sociology, New York University, New York, NY 10012, USA 117Department of Economics, Harvard University, Cambridge, MA 02138, USA 118Petrie-Flom Center for Health Law Policy, Biotechnology, & Bioethics, Harvard Law School, Cambridge, MA 02138, USA 119Nelson A. Rockefeller Institute of Government, State University of New York, Albany, NY 12203–1003, USA 120Department of Clinical Genetics, VU University Medical Center, 1081 BT Amsterdam, The Netherlands 121Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam 3000 DR, The Netherlands 122Department of Economics, Stockholm School of Economics, Stockholm 113 83, Sweden 123Panteia, Zoetermeer 2701 AA, Netherlands 124GSCM-Montpellier Business School, Montpellier 34185, France 125Centre for Medical Systems Biology, Leiden University Medical Center, 2300 RC Leiden, The Netherlands 126Department of Science. Author manuscript; available in PMC 2013 August 23.
Abstract

A genome-wide association study of educational attainment was conducted in a discovery sample of 101,069 individuals and a replication sample of 25,490. Three independent SNPs are genome-wide significant (rs9320913, rs11584700, rs4851266), and all three replicate. Estimated effects sizes are small ($R^2 \approx 0.02\%$), approximately 1 month of schooling per allele. A linear polygenic score from all measured SNPs accounts for $\approx 2\%$ of the variance in both educational attainment and cognitive function. Genes in the region of the loci have previously been associated with health, cognitive, and central nervous system phenotypes, and bioinformatics analyses suggest the involvement of the anterior caudate nucleus. These findings provide promising candidate SNPs for follow-up work, and our effect size estimates can anchor power analyses in social-science genetics.

Twin and family studies suggest that a broad range of psychological traits (1), economic preferences (2–4), and social and economic outcomes (5) are moderately heritable. Discovery of genetic variants associated with such traits leads to insights regarding the biological pathways underlying human behavior. If the predictive power of a set of genetic variants considered jointly is sufficiently large, then a “risk score” that aggregates their effects could be useful to control for genetic factors that are otherwise unobserved, or to identify populations with certain genetic propensities, for example in the context of medical intervention (6).

To date, however, few if any robust associations between specific genetic variants and social-scientific outcomes have been identified likely because existing work [for review see (7)] has relied on samples that are too small [for discussion, see (4, 6, 8, 9)]. In this paper, we apply to a complex behavioral trait—educational attainment—an approach to gene discovery that has been successfully applied to medical and physical phenotypes (10), namely meta-analyzing data from multiple samples. The phenotype of educational attainment is available in many samples with genotyped subjects (5). Educational attainment is influenced by many known environmental factors, including public policies. Educational attainment is strongly associated with social outcomes, and there is a well-documented health-education gradient (5, 11). Estimates suggest that around 40% of the variance in educational attainment is explained by genetic factors (5). Furthermore, educational attainment is moderately correlated with other heritable characteristics (1), including cognitive function (12) and personality traits related to persistence and self-discipline (13).

To create a harmonized measure of educational attainment, we coded study-specific measures using the International Standard Classification of Education (ISCED 1997) scale (14). We analyzed a quantitative variable defined as an individual’s years of schooling ($EduYears$) and a binary variable for college completion ($College$). $College$ may be more comparable across countries, whereas $EduYears$ contains more information about individual differences within countries.

A genome-wide association study (GWAS) meta-analysis was performed across 42 cohorts in the discovery phase. The overall discovery sample comprises 101,069 individuals for $EduYears$ and 95,427 for $College$. Analyses were performed at the cohort level according to a pre-specified analysis plan, which restricted the sample to Caucasians (to help reduce...
stratification concerns). Educational attainment was measured at an age at which subjects were very likely to have completed their education [over 95% of the sample was at least 30; (5)]. On average, subjects have 13.3 years of schooling, and 23.1% have a college degree. To enable pooling of GWAS results, all studies conducted analyses with data imputed to the HapMap 2 CEU (r22.b36) reference set. To guard against population stratification, the first four principal components of the genotypic data were included as controls in all the cohort-level analyses. All study-specific GWAS results were quality controlled, cross-checked, and meta-analyzed using single genomic control and a sample-size weighting scheme at three independent analysis centers.

At the cohort level, there is little evidence of general inflation of p-values. As in previous GWA studies of complex traits (15), the Q-Q plot of the meta-analysis exhibits strong inflation. This inflation is not driven by specific cohorts and is expected for a highly polygenic phenotype even in the absence of population stratification (16).

From the discovery phase, we identified one genome-wide significant locus (rs9320913, p = 4.2 × 10^{-9}) and three suggestive loci (defined as p < 10^{-6}) for EduYears. For College, we identified two genome-wide significant loci (rs11584700, p = 2.1 × 10^{-9}, and rs4851266, p = 2.2 × 10^{-9}) and an additional four suggestive loci (Table 1). We conducted replication analyses in 12 additional, independent cohorts that became available after the completion of the discovery meta-analysis, using the same pre-specified analysis plan. For both EduYears and College, the replication sample comprises 25,490 individuals.

For each of the ten loci that reached at least suggestive significance, we brought forward for replication the SNP with the lowest p-value. The three genome-wide significant SNPs replicate at the Bonferroni-adjusted 5% level, with point estimates of the same sign and similar magnitude (Fig. 1 and Table 1). The seven loci that did not reach genome-wide significance did not replicate (the effect went in the anticipated direction in 5 out of 7 cases). The meta-analytic findings are not driven by extreme results in a small number of cohorts (see $p_{het}$ in Table 1), by cohorts from a specific geographic region (figs. S7 to S15), or by a single sex (figs. S3 to S6). Given the high correlation between EduYears and College (5), it is unsurprising that the set of SNPs with low p-values exhibit considerable overlap in the two analyses (tables S8 and S9).

The observed effect sizes of the three replicated individual SNPs are small [see (5) for discussion]. For EduYears, the strongest effect identified (rs9320913) explains 0.022% of phenotypic variance in the replication sample. This $R^2$ corresponds to a difference of ~1 months of schooling per allele. For college completion, the SNP with the strongest estimated effect (rs11584700) has an odds ratio of 0.912 in the replication sample, equivalent to a 1.8 percentage-point difference per allele in the frequency of completing college.

We subsequently conducted a “combined stage” meta-analysis, including both the discovery and replication samples. This analysis revealed additional genome-wide significant SNPs: four for EduYears and three for College. Three of these newly genome-wide significant SNPs (rs1487441, rs11584700, rs4851264) are in linkage disequilibrium with the replicated SNPs. The remaining four are located in different loci and warrant replication attempts in future research: rs7309, a 3’UTR variant in TANK; rs11687170, close to GBX2; rs1056667, a 3’UTR variant in BTN1A1; and rs13401104 in ASB18.

Using the results of the combined meta-analyses of discovery and replication cohorts, we conducted a series of complementary and exploratory supplemental analyses to aid in interpreting and contextualizing the results: gene-based association tests; eQTL analyses of brain and blood tissue data; pathway analysis; functional annotation searches; enrichment analysis for cell-type-specific overlap with H3K4me3 chromatin marks; and predictions of
likely gene function using gene-expression data. Table S20 summarizes promising candidate loci identified through follow-up analyses (5). Two regions in particular showed convergent evidence from functional annotation, blood cis-eQTL analyses, and gene-based tests: chromosome 1q32 (including LRRN2, MDM4, and PIK3C2B) and chromosome 6 near the Major Histocompatibility Complex (MHC). We also find evidence that in anterior caudate cells, there is enrichment of H3K4me3 chromatin marks (believed to be more common in active regulatory regions) in the genomic regions implicated by our analyses (fig. S20).

Many of the implicated genes have previously been associated with health, central nervous system, or cognitive-process phenotypes in either human-GWAS or model-animal studies (table S22). Gene co-expression analysis revealed that several implicated genes (including BSN, GBX2, LRRN2, and PIK3C2B) are likely involved in pathways related to cognitive processes (such as learning and long-term memory) and neuronal development or function (table S21).

Although the effects of individual SNPs on educational attainment are small, many of their potential uses in social science depend on their combined explanatory power. To evaluate the combined explanatory power, we constructed a linear polygenic score (5) for each of our two education measures using the meta-analysis results (combining discovery and replication), excluding one cohort. We tested these scores for association with educational attainment in the excluded cohort. We constructed the scores using SNPs whose nominal p-values fall below a certain threshold, ranging from $5 \times 10^{-8}$ (only the genome-wide significant SNPs were included) to 1 (all SNPs were included).

We replicated this procedure with two of the largest cohorts in the study, both of which are family-based samples (QIMR and STR). The results suggest that educational attainment is a highly polygenic trait (Fig. 2 and table S23): the amount of variance accounted for increases as the p-value threshold becomes less conservative (i.e., includes more SNPs). The linear polygenic score from all measured SNPs accounts for \( \approx 2\% \) \((p = 1.0 \times 10^{-29})\) of the variance in EduYears in the STR sample and \( \approx 3\% \) \((p = 7.1 \times 10^{-24})\) in the QIMR sample.

To explore one of the many potential mediating endophenotypes, we examined how much the same polygenic scores (constructed to explain EduYears or College) could explain individual differences in cognitive function. While it would have been preferable to explore a richer set of mediators, this variable was available in STR, a dataset where we had access to the individual-level genotypic data. Cognitive function had been measured in a subset of males using the Swedish Enlistment Battery (used for conscription) (5, 17). The estimated \( R^2 \approx 2.5\% \) \((p < 1.0 \times 10^{-8})\) for cognitive function is actually slightly larger than the fraction of variance in educational attainment captured by the score in the STR sample. One possible interpretation is that some of the SNPs used to construct the score matter for education through their stronger, more direct effects on cognitive function (5). A mediation analysis (table S24) provides tentative evidence consistent with this interpretation.

The polygenic score remains associated with educational attainment and cognitive function in within-family analyses (table S25). Thus, these results appear robust to possible population stratification.

If the size of the training sample used to estimate the linear polygenic score increased, the explanatory power of the score in the prediction sample would be larger because the coefficients used for constructing the score would be estimated with less error. In (5), we report projections of this increase. We also assess, at various levels of explanatory power, the benefits from using the score as a control variable in a randomized educational intervention (5). An asymptotic upper bound for the explanatory power of a linear polygenic score is the additive genetic variance across individuals captured by current SNP
microarrays. Using combined data from STR and QIMR, we estimate that this upper bound is 22.4% (S.E. = 4.2%) in these samples (5) (table S12).

Placed in the context of the GWAS literature (10), our largest estimated SNP effect size of 0.02% is over an order of magnitude smaller than those observed for height and BMI: 0.4% (15) and 0.3% (18) respectively. While our linear polygenic score for education achieves an $R^2$ of 2% estimated from a sample of 120,000, a score for height reached 10% estimated from a sample of 180,000 (15), and a score for BMI using only the top 32 SNPs reached 1.4% (18). Taken together, our findings suggest that the genetic architecture of complex behavioral traits is far more diffuse than that of complex physical traits.

Existing claims of “candidate gene” associations with complex social-science traits have reported widely varying effect sizes—many with $R^2$ values more than one hundred times larger than those we find (4, 6). For complex social-science phenotypes that are likely to have a genetic architecture similar to educational attainment, our estimate of 0.02% can serve as a benchmark for conducting power analyses and evaluating the plausibility of existing findings in the literature.

The few GWAS studies conducted to date in social-science genetics have not found genome-wide significant SNPs that replicate consistently (19, 20). One commonly proposed solution is to gather better measures of the phenotypes in more environmentally homogenous samples. Our findings demonstrate the feasibility of a complementary approach: identify a phenotype that, although more distal from genetic influences, is available in a much larger sample [see (5) for a simple theoretical framework and power analysis]. The genetic variants uncovered by this “proxy-phenotype” methodology can then serve as a set of empirically-based candidate genes in follow-up work, such as tests for associations with well-measured endophenotypes (e.g., personality, cognitive function), research on gene-environment interactions, or explorations of biological pathways.

In social-science genetics, researchers must be especially vigilant to avoid misinterpretations. One of the many concerns is that a genetic association will be mischaracterized as “the gene for X,” encouraging misperceptions that genetically influenced phenotypes are immune to environmental intervention [for rebuttals, see (21, 22)] and misperceptions that individual SNPs have large effects (which our evidence contradicts). If properly interpreted, identifying SNPs and constructing polygenic scores are steps toward usefully incorporating genetic data into social-science research.

**Supplementary Material**

Refer to Web version on PubMed Central for supplementary material.

**Acknowledgments**

This research was carried out under the auspices of the Social Science Genetic Association Consortium (SSGAC), a cooperative enterprise among medical researchers and social scientists that coordinates genetic association studies for social science variables. Data for our analyses come from many studies and organizations, some of which are subject to an MTA (5). Results from the meta-analysis are available at the website of the consortium, www.ssgac.org. The formation of the SSGAC was made possible by an EAGER grant from the NSF and a supplemental grant from the NIH/OBSSR (SES-1064089). This research was also funded in part by the Söderbergh Foundation (E9/11), the NIA/NIH through grants P01-AG005842, P01-AG005842-2082, P30-AG012810, and T32-AG000186-23 and the Intramural Research Program of the NIA/NIH. For a full list of acknowledgments, see (5).
References and Notes


5. Please see the supplementary materials on Science Online.


30. SCAN, SNP and CNV Annotation Database. 2012. www.scandb.org/


72. Barrett JC, et al. UK IBD Genetics Consortium; Wellcome Trust Case Control Consortium 2, Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region. Nat Genet. 2009; 41:1330.10.1038/ng.483 [PubMed: 19915572]


77. Anderson CA, et al. Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. Nat Genet. 2011; 43:246.10.1038/ng.764 [PubMed: 21297633]


86. Hayes BJ, Visscher PM, Goddard ME. Increased accuracy of artificial selection by using the realized relationship matrix. Genet Res. 2009; 91:47.10.1017/S0016672308009981


89. Fryer RG. Financial incentives and student achievement: Evidence from randomized trials. Q J Econ. 2011; 126:1755.10.1093/qje/qgr045


Fig. 1.
Regional association plots of replicated loci associated with educational attainment [(A): rs9320913, (B): rs11584700, (C): rs4851266]. The plots are centered on the SNPs with the lowest p-values in the discovery stage (purple diamond). The $R^2$ values are from the CEU HapMap 2 samples. The CEU HapMap 2 recombination rates are indicated with a blue line on the right-hand y-axis. The figures were created with LocusZoom (http://csg.sph.umich.edu/locuszoom/).
Fig. 2.
Solid lines show results from regressions of *EduYears* on linear polygenic scores in a set of unrelated individuals from the QIMR ($N = 3526$) and STR ($N = 6770$) cohorts. Dashed lines show results from regressions of *Cognitive function* on linear polygenic scores in a sample from STR ($N = 1419$). The scores are constructed from the meta-analysis for either *EduYears* or *College*, excluding the QIMR and STR cohorts.
Table 1

The results of the GWAS meta-analysis for the independent signals reaching \( p < 10^{-6} \) in the discovery stage.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr</th>
<th>Position (bp)</th>
<th>Nearest gene</th>
<th>Effective allele</th>
<th>Frequency</th>
<th>Beta/OR (EduYears)</th>
<th>P-value (EduYears)</th>
<th>I²</th>
<th>P_het</th>
<th>Beta/OR (College)</th>
<th>P-value (College)</th>
<th>Beta/OR (College) – sex-specific</th>
<th>P-value (College) – sex-specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs9320913</td>
<td>6</td>
<td>9801454</td>
<td>LOC100129158</td>
<td>A</td>
<td>0.483</td>
<td>0.106</td>
<td>4.19×10⁻⁹</td>
<td>0.101</td>
<td>3.50×10⁻¹⁰</td>
<td>0.095</td>
<td>1.87×10⁻⁸</td>
<td>0.100</td>
<td>1.43×10⁻⁶</td>
</tr>
<tr>
<td>rs3783006</td>
<td>13</td>
<td>9759310</td>
<td>STK24</td>
<td>C</td>
<td>0.454</td>
<td>0.096</td>
<td>2.29×10⁻⁷</td>
<td>0.098</td>
<td>8.45×10⁻⁸</td>
<td>0.088</td>
<td>1.44×10⁻⁷</td>
<td>0.108</td>
<td>3.35×10⁻⁷</td>
</tr>
<tr>
<td>rs8049439</td>
<td>16</td>
<td>28745016</td>
<td>ATXN2L</td>
<td>T</td>
<td>0.581</td>
<td>0.090</td>
<td>7.12×10⁻⁷</td>
<td>0.065</td>
<td>1.55×10⁻⁷</td>
<td>0.086</td>
<td>1.43×10⁻⁶</td>
<td>0.078</td>
<td>1.90×10⁻⁷</td>
</tr>
<tr>
<td>rs1318378</td>
<td>5</td>
<td>10558587</td>
<td>SLCO6A1</td>
<td>A</td>
<td>0.878</td>
<td>-0.136</td>
<td>7.40×10⁻⁷</td>
<td>0.091</td>
<td>1.37×10⁻⁴</td>
<td>0.097</td>
<td>1.43×10⁻⁷</td>
<td>0.080</td>
<td>5.92×10⁻⁷</td>
</tr>
<tr>
<td>rs11984700</td>
<td>1</td>
<td>202843606</td>
<td>LRRN2</td>
<td>A</td>
<td>0.780</td>
<td>0.521</td>
<td>2.07×10⁻⁶</td>
<td>0.912</td>
<td>8.24×10⁻¹⁵</td>
<td>0.934</td>
<td>6.11×10⁻⁸</td>
<td>0.911</td>
<td>2.12×10⁻⁹</td>
</tr>
<tr>
<td>rs4852166</td>
<td>2</td>
<td>101814911</td>
<td>LOC150577</td>
<td>T</td>
<td>0.396</td>
<td>1.050</td>
<td>2.20×10⁻⁹</td>
<td>1.049</td>
<td>5.33×10⁻¹¹</td>
<td>1.054</td>
<td>1.55×10⁻⁸</td>
<td>1.052</td>
<td>6.74×10⁻⁸</td>
</tr>
<tr>
<td>rs2054125</td>
<td>4</td>
<td>19093906</td>
<td>PLCL1</td>
<td>T</td>
<td>0.064</td>
<td>1.468</td>
<td>5.55×10⁻⁸</td>
<td>1.098</td>
<td>2.12×10⁻⁵</td>
<td>1.264</td>
<td>1.74×10⁻²</td>
<td>1.503</td>
<td>1.95×10⁻⁷</td>
</tr>
<tr>
<td>rs3227</td>
<td>6</td>
<td>3370273</td>
<td>ITPR3</td>
<td>C</td>
<td>0.498</td>
<td>1.043</td>
<td>6.02×10⁻⁸</td>
<td>1.010</td>
<td>3.24×10⁻⁴</td>
<td>1.046</td>
<td>9.44×10⁻⁵</td>
<td>1.029</td>
<td>1.37×10⁻⁵</td>
</tr>
<tr>
<td>rs4073894</td>
<td>7</td>
<td>104254200</td>
<td>LHFPL3</td>
<td>A</td>
<td>0.207</td>
<td>1.076</td>
<td>4.41×10⁻⁷</td>
<td>0.076</td>
<td>5.55×10⁻⁶</td>
<td>1.050</td>
<td>2.18×10⁻²</td>
<td>1.073</td>
<td>1.74×10⁻⁵</td>
</tr>
<tr>
<td>rs1264026</td>
<td>4</td>
<td>176833266</td>
<td>GPM6A</td>
<td>A</td>
<td>0.890</td>
<td>1.041</td>
<td>4.94×10⁻⁷</td>
<td>1.000</td>
<td>7.48×10⁻⁶</td>
<td>1.038</td>
<td>1.59×10⁻⁵</td>
<td>1.031</td>
<td>7.61×10⁻⁴</td>
</tr>
</tbody>
</table>

The rows in bold are the independent signals reaching \( p < 5 \times 10^{-8} \) in the discovery stage. "Frequency" refers to allele-frequency in the combined-stage meta-analysis. "Beta/OR" refers to the effect size in the EduYears analysis and to the Odds Ratio in the College analysis. All \( p \)-values are from the sample-size-weighted meta-analysis (fixed effects). The \( p \)-value in the replication stage meta-analysis was calculated from a one-sided test. \( I² \) represents the % heterogeneity of effect size between the discovery stage studies. \( P_het \) is the heterogeneity \( p \)-value.