Common genetic variants associated with cognitive performance identified using the proxy-phenotype method

Cornelius A. Rietveld[a,b], Tõnu Esko[c,d,e,f], Gail Davies[g,h], Tune H. Pers[c,d], Patrick Turley[i], Beben Benyamin[j], Christopher F. Chabris[k], Valur Emilsson[l,m], Andrew D. Johnson[n], James J. Lee[o,p], Christiaan de Leeuw[q,r], Riccardo E. Marioni[g,j,s], Sarah E. Medland[t], Michael B. Miller[p], Olga Rostapshova[v], Sven J. van der Lee[w], Anna A. E. Vinkhuyzen[j], Najaf Amin[w], Dalton Conley[x], Jaime Derringer[y], Cornelia M. van Duijn[w,z], Rudolf Fehrmann[aa], Lude Franke[aa], Edward L. Glaeser[i], Narelle K. Hansell[bb], Caroline Hayward[s,cc], William G. Iacono[p], Carla Ibrahim-Verbaas[v,dd], Vincent Jaddoe[b,ee], Juha Karjalainen[aa], David Laibson[i], Paul Lichtenstein[i], David C. Liewald[g], Patrik K. E. Magnusson[ff], Nicholas G. Martin[u], Matt McGue[p], George McMahon[gg], Nancy L. Pedersen[ff], Steven Pinker[o], David J. Porteous[g,s], Danielle Posthuma[q,hh,ii], Fernando Rivadeneira[b,jj], Blair H. Smith[kk], John M. Starr[g,ll], Henning Tiemeier[b,hh], Nicholas J. Timpson[mm], Maciej Trzaskowski[nn], André G. Uitterlinden[b,jj], Frank C. Verhulst[hh], Mary E. Ward[gg], Margaret J. Wright[bb], George Davey Smith[mm], Ian J. Deary[g,h], Magnus Johannesson[oo], Robert Plomin[nn], Peter M. Visscher[j,pp], Daniel J. Benjamin[qq,1,2], David Cesarini[rr,ss,1], and Philipp D. Koellinger[a,b,tt,1,2]

[a]Department of Applied Economics, Erasmus School of Economics, Erasmus University, 3000 DR, Rotterdam, The Netherlands; Departments of [b]Epidemiology, [dd]Neurology, and [jj]Internal Medicine, [w]Genetic Epidemiology Unit, Department of Epidemiology and Biostatistics, and [ee]Generation R Study Group, Erasmus Medical Center, 3000 CA, Rotterdam, The Netherlands; [c]Division of Genetics and Endocrinology, Boston Children's Hospital, Boston, MA 02115; [d]Program in Medical and Population Genetics, Broad Institute of the Massachusetts Institute of Technology and Harvard, Cambridge, MA 02142; [e]Department of Genetics, Harvard Medical School, Boston, MA 02115; [f]Estonian Genome Center, University of Tartu, 51010 Tartu, Estonia; [g]Centre for Cognitive Ageing and Cognitive Epidemiology, [h]Department of Psychology, and [ll]Alzheimer Scotland Dementia Research Centre, University of Edinburgh, Edinburgh EH8 9JZ, United Kingdom; Departments of [i]Economics and [o]Psychology, Harvard University, Cambridge, MA 02138; [j]Queensland Brain Institute, University of Queensland, Brisbane, QLD 4072, Australia; [k]Department of Psychology, Union College, Schenectady, NY 12308; [l]Icelandic Heart Association, 201 Kopavogur, Iceland; [m]Faculty of Pharmaceutical Sciences, University of Iceland, 107 Reykjavík, Iceland; [n]Framingham Heart Study, National Heart, Lung, and Blood Institute, Framingham, MA 01702; [p]Department of Psychology, University of Minnesota, Minneapolis, MN 55455-0344; [q]Department of Complex Trait Genetics, VU University Amsterdam and VU Medical Center, 1081 HV, Amsterdam, The Netherlands; [r]Machine Learning Group, Intelligent Systems, Institute for Computing and Information Sciences, Faculty of Science, Radboud University, 6500 GL, Nijmegen, The Netherlands; [s]Centre for Genomic and Experimental Medicine and [cc]Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, United Kingdom; [t]Quantitative Genetics Laboratory, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4029, Australia; [u]Genetic Epidemiology Laboratory, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4029, Australia; [v]Harvard Kennedy School, Harvard University, Cambridge, MA 02139; [x]Department of Sociology and [rr]Center for Experimental Social Science, Department of Economics, New York University, New York, NY 10012; [y]Department of Psychology, University of Illinois, Urbana–Champaign, IL 61820; [z]Centre for Medical Systems Biology, Leiden University Medical Center, 2300 RC, Leiden, The Netherlands; [aa]Department of Genetics, University Medical Center Groningen, University of Groningen, 9713 GZ, Groningen, The Netherlands; [bb]Neuroimaging Genetics Group, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4029, Australia; [ff]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, 171 77 Stockholm, Sweden; [gg]School of Social and Community Medicine, University of Bristol, Bristol BS8 2PR, United Kingdom; [hh]Department of Child and Adolescent Psychiatry, Erasmus Medical Center, 3000 CB, Rotterdam, The Netherlands; [ii]Department of Clinical Genetics, VU University Medical Center, 1081 BT, Amsterdam, The Netherlands; [kk]Medical Research Institute, University of Dundee, Dundee DD2 4RB, United Kingdom; [mm]Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol BS8 2PR, United Kingdom; [nn]Medical Research Council Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, London SE5 8AF, United Kingdom; [oo]Department of Economics, Stockholm School of Economics, 113 83 Stockholm, Sweden; [pp]University of Queensland Diamantina Institute, University of Queensland, Princess Alexandra Hospital, Brisbane, QLD 4102, Australia; [qq]Department of Economics, Cornell University, Ithaca, NY 14853; [ss]Institute for the Interdisciplinary Study of Decision Making, New York University, New York, NY 10012; and [tt]Faculty of Economics and Business, University of Amsterdam, 1012 WX, Amsterdam, The Netherlands

Edited by Michael S. Gazzaniga, University of California, Santa Barbara, CA, and approved August 14, 2014 (received for review March 12, 2014)

We identify common genetic variants associated with cognitive performance using a two-stage approach, which we call the proxy-phenotype method. First, we conduct a genome-wide association study of educational attainment in a large sample (n = 106,736), which produces a set of 69 education-associated SNPs. Second, using independent samples (n = 24,189), we measure the association of these education-associated SNPs with cognitive performance. Three SNPs (rs1487441, rs7923609, and rs2721173) are significantly associated with cognitive performance after correction for multiple hypothesis testing. In an independent sample of older Americans (n = 8,652), we also show that a polygenic score derived from the education-associated SNPs is associated with memory and absence of dementia. Convergent evidence from a set of bioinformatics analyses implicates four specific genes (*KNCMA1*, *NRXN1*, *POU2F3*, and *SCRT*). All of these genes are associated with a particular neurotransmitter pathway involved in synaptic plasticity, the main cellular mechanism for learning and memory.

Twin and family studies have shown that at least a moderate share of variation in most facets of cognitive performance (i.e., performance by healthy individuals on cognitive tests) is associated with genetic factors (1, 2). However, despite considerable interest and effort, research to date has largely failed to identify common genetic variants associated with cognitive performance phenotypes (3–5) with the exception of *APOE*, which predicts cognitive decline in older individuals (6–8). Existing studies have

Author contributions: D.J.B., D. Cesarini, and P.D.K. designed research; C.A.R., T.E., G.D., T.H.P., P.T., B.B., V.E., A.D.J., J.J.L., C.d.L., R.E.M., S.E.M., M.B.M., O.R., S.J.v.d.L., A.A.E.V., N.A., D. Conley, J.D., R.F., L.F., C.H., C.I.-V., J.K., D.C.L., P.K.E.M., G.M., D.P., M.T., M.E.W., M.J., P.M.V., and D. Cesarini analyzed data; C.A.R., T.E., P.T., C.F.C., D.L., D.J.B., D. Cesarini, and P.D.K. wrote the paper; C.F.C., C.M.v.D., E.L.G., W.G.I., V.J., D.L., P.L., N.G.M., M.M., N.L.P., S.P., D.P., J.M.S., H.T., F.C.V., M.J.W., G.D.S., I.J.D., M.J., and R.P. performed data collection; J.J.L., M.B.M., C.M.v.D., N.K.H., P.K.E.M., D.J.P., B.H.S., J.M.S., H.T., N.J.T., M.J.W., I.J.D., and M.J. performed phenotyping; and G.D., M.B.M., C.M.v.D., C.H., V.J., D.C.L., P.K.E.M., N.G.M., D.J.P., F.R., N.J.T., and A.G.U. performed genotyping.

See *SI Appendix* for further details.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: Genetic summary data on which our work is based are posted on the website of our research consortium (www.ssgac.org).

[1]D.J.B., D. Cesarini, and P.D.K. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: daniel.benjamin@gmail.com or p.d.koellinger@uva.nl.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1404623111/-/DCSupplemental.

Significance

We identify several common genetic variants associated with cognitive performance using a two-stage approach: we conduct a genome-wide association study of educational attainment to generate a set of candidates, and then we estimate the association of these variants with cognitive performance. In older Americans, we find that these variants are jointly associated with cognitive health. Bioinformatics analyses implicate a set of genes that is associated with a particular neurotransmitter pathway involved in synaptic plasticity, the main cellular mechanism for learning and memory. In addition to the substantive contribution, this work also serves to show a proxy-phenotype approach to discovering common genetic variants that is likely to be useful for many phenotypes of interest to social scientists (such as personality traits).

relied on one of two research strategies. The first strategy is a candidate gene design, in which researchers test a small number of genetic variants for association with the phenotype of interest, typically based on hypotheses derived from the known biological functions of the candidate genes. The candidate gene associations that have been reported with cognitive performance (9), however, fail to replicate when larger samples are used (3). The second research strategy is a genome-wide association study (GWAS), in which researchers atheoretically test hundreds of thousands of SNPs for association with the phenotype and apply a threshold for genome-wide statistical significance—typically 5×10^{-8}—to account for multiple hypothesis testing. For physical and medical phenotypes, GWASs have identified many novel associations that replicate (10). GWASs on cognitive performance, however, have not yet identified any genome-wide significant associations (4, 5).

Here, we apply an alternative two-stage research strategy, which we call the proxy-phenotype method. In the first stage, we conduct a GWAS on a proxy phenotype to identify a relatively small set of SNPs that are associated with the proxy phenotype. In the second stage, these SNPs serve as candidates that are tested in independent samples for association with the phenotype of interest at a significance threshold corrected for the number of proxy-associated SNPs. In the study reported here, our phenotype of interest is cognitive performance, for which we use Spearman's measure of general cognitive ability (usually abbreviated as *g*; it is the general factor measured by a battery of diverse cognitive tests) (4). Our proxy phenotype is educational attainment measured by self-reported years of schooling.

Rietveld et al. (11) had suggested the strategy of using SNPs associated with educational attainment as "empirically-based candidate genes" for association with cognitive performance (11); here, we conduct that analysis and further develop the methodology for doing so. *SI Appendix* contains our formal framework, which builds on that in ref. 11, as well as power calculations under a range of assumptions. According to the framework, educational attainment is a good proxy phenotype for cognitive performance, because cognitive performance is strongly genetically influenced and causally affects educational attainment; also, much larger samples are available for GWASs on educational attainment. The high genetic correlation (estimated to be roughly 0.65 or higher) (12–14) between the two traits does not have straightforward implications for the statistical power to identify specific SNPs influencing cognitive performance. It does, however, imply that a polygenic score associated with educational attainment will be associated with cognitive performance; thus, it may be viewed as providing an additional suggestive justification for the approach to identifying specific SNPs.

Results

In our first stage, we conducted a GWAS of educational attainment in a pooled Education Sample of 106,736 individuals. We used the same data, analysis protocol, and quantitative years of schooling measure as in ref. 11, except that we omit cohorts with high-quality measures of cognitive performance; we, instead, include these cohorts in the subsequent Cognitive Performance Sample. We chose our inclusion threshold of $P < 10^{-5}$ for selecting candidate SNPs based on ex ante power calculations, with a goal to maximize the number of true positives among the candidates (*SI Appendix*). Pruning for linkage disequilibrium the 927 SNPs that reach this threshold resulted in 69 approximately independent SNPs (*SI Appendix*).

In our second stage, we tested these 69 education-associated SNPs for association with cognitive performance in the Cognitive Performance Sample, which comprises 24,189 genotyped subjects from 11 cohorts (*SI Appendix*). The specific cognitive tests differ across cohorts, but the cognitive performance measure in every cohort is calculated as Spearman's *g* (*SI Appendix*); previous research has found that *g* values from different test batteries are highly correlated, especially if the batteries have many tests or if the test is specifically constructed to measure *g* (15–17). We tested each SNP individually for association with cognitive performance using ordinary least squares, controlling for sex, age, and (depending on the cohort) at least four principal components of the genome-wide data (to reduce confounding from population stratification). At the cohort level, the analyses were conducted according to a prespecified plan that we preregistered on the Open Science Framework (https://osf.io/z7fe2/). The cohorts' results were then meta-analyzed using an inverse-variance weighting scheme. Two independent teams of analysts cross-checked and verified the results.

To confirm that the education-based first stage identifies reasonable candidate SNPs for cognitive performance, Fig. 1 plots the standardized regression coefficients from the regression of years of schooling on the education-associated SNPs in the Education Sample (with the reference allele chosen to ensure that the coefficient is positive) against the standardized coefficients from the second-stage regression of cognitive performance on the SNPs in the Cognitive Performance Sample. The direction of the effect coincides in 53 of 69 cases (two-sided binomial test, $P = 9.10 \times 10^{-6}$), indicating that this context is a good context for applying the proxy-phenotype method. We were surprised that the correlation between the effect size on educational attainment and the effect size on cognitive performance is negative ($\rho = -0.25$, $P = 0.03$), although not significantly after dropping a possible outlier (the bottommost point of the figure; $\rho = -0.14$, $P = 0.26$). If the population correlation is truly negative, within our theoretical framework, it suggests that SNPs that affect cognitive performance more strongly tend to affect other factors that matter for educational attainment (such as personality traits) less strongly and vice versa (*SI Appendix*).

To provide a benchmark for evaluating our list of education-associated candidate SNPs, we generated (through a prespecified algorithm) a list of theory-based candidate SNPs for cognitive performance drawn from published findings in the candidate gene literature (*SI Appendix*). (This list does not include the SNPs comprising the *APOE* haplotype, because these SNPs were not available in the cohort GWAS results.) After applying the same pruning procedure as for the education-associated SNPs, our list of theory-based SNPs contains 24 independent SNPs, of which only one is in a genomic region close to an education-associated SNP. Fig. 2 overlays Q–Q plots for the theory-based and education-associated candidates. The education-associated candidates taken altogether are more strongly associated with cognitive performance than would be expected by chance ($z = 5.98$, $P = 1.12 \times 10^{-9}$). Whereas a visual inspection of the plot

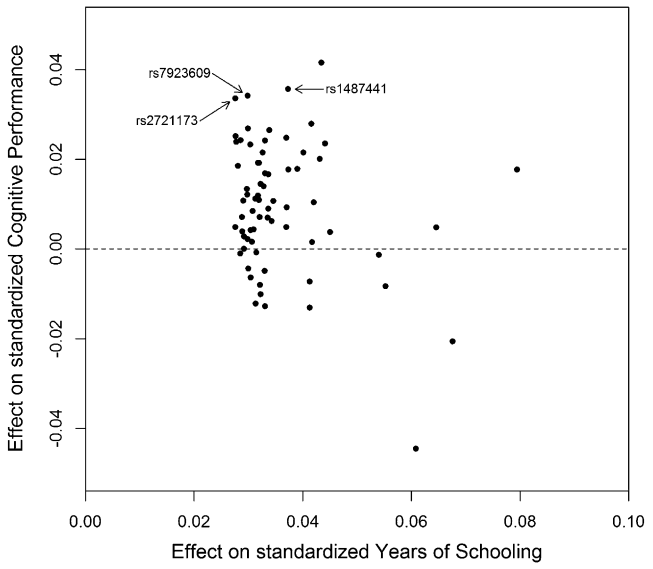

Fig. 1. The relationship between standardized coefficients from the first-stage regression of years of schooling on the education-associated SNPs in the Education Sample (*x* axis) and standardized coefficients from the second-stage regression of cognitive performance on these SNPs in the Cognitive Performance Sample (*y* axis). The reference allele is chosen such that the coefficient on years of schooling is positive. Each point represents 1 of the 69 education-associated SNPs. (The cloud of points is bounded away from zero effect on years of schooling, because the criterion for including an SNP was reaching $P < 10^{-5}$ in the GWAS on years of schooling in the Education Sample.) Because the SD of years of schooling is ∼3, a coefficient of 0.03—a typical size for a years of schooling standardized coefficient (before correcting for the winner's curse)—means that each reference allele is associated with an increase of 0.03 × 3 ∼ 0.09 y of educational attainment. In conventional IQ units that have an SD of 15, a standardized regression coefficient on cognitive performance of 0.03 corresponds to ∼0.45 IQ points.

suggests that the theory-based candidates exhibit some association with cognitive performance, we cannot reject the null hypothesis for any SNP individually or all of them taken together ($z = 1.19$, $P = 0.12$).

The top three education-associated SNPs—rs1487441, rs7923609, and rs2721173—show clear separation from the others in Fig. 2 and are significantly associated with cognitive performance after Bonferroni correction for multiple hypothesis testing (Table 1). Consistent with the negative correlation in Fig. 1, these SNPs are different from the three SNPs that reached genome-wide significance for association with educational attainment in the analyses in ref. 11. After adjusting the estimated effect sizes of the SNPs (each $R^2 \sim 0.0006$) for the winner's curse, we estimate each as $R^2 \sim 0.0002$ (*SI Appendix*), or in terms of coefficient magnitude, each additional reference allele for each SNP is associated with an ∼0.02 SD increase in cognitive performance [or 0.3 points on the typical intelligence quotient (IQ) scale]. $R^2 \sim 0.0002$ is about the same as the R^2 value for the known SNP associations with educational attainment (11) but far smaller than the largest effect sizes for complex physical traits, such as height ($R^2 \sim 0.004$) and body mass index ($R^2 \sim 0.003$) (18, 19).

Power calculations that we report in *SI Appendix* help shed light on why the proxy-phenotype method succeeded in identifying SNPs, whereas GWASs to date on cognitive performance have not. A GWAS in our Cognitive Performance Sample of $n = 24{,}189$—which is larger than the largest GWASs ($n = 17{,}989$ in ref. 4 and $n = 3{,}511$ in ref. 5)—would have had power of 0.06% to identify any given SNP with an association that has $R^2 = 0.0002$. In contrast, our proxy-phenotype approach had power of 12%. Given this power and the rather stringent significance threshold (0.05/69 ∼ 0.00072), Bayesian calculations using reasonable assumptions regarding priors suggest that the posterior probabilities that these three SNPs are associated with cognitive performance are high (*SI Appendix*).

Turning from specific SNPs to the set of all 69 education-associated SNPs, we assess the explanatory power of a linear polygenic score that aggregates their coefficients (*SI Appendix*). In pooled results from four family-based cohorts (4,463 individuals in total), we find that the score is significantly associated with cognitive performance ($P = 8.17 \times 10^{-4}$), with R^2 ranging approximately from 0.2% to 0.4% across samples. Using only within-family variation, the pooled coefficient has the same sign but is smaller and has a larger SE ($P = 0.36$). Thus, we cannot rule out that some of the score's explanatory power is because of population stratification, although even without stratification, the nonsignificance of the within-family coefficient is not surprising given the low power of this test (*SI Appendix*).

Next, we explore whether educational attainment might serve as a proxy phenotype for cognitive health phenotypes (as opposed to cognitive performance in the normal range). Our sample comprises 8,652 European descent individuals over the age of 50 y from the Health and Retirement Study (HRS) (*SI Appendix*). We confirm that, for 60 of 69 SNPs available in the HRS data, the direction of the effects on educational attainment generally coincides with the direction of the effects on the two cognitive health phenotypes that we study: total word recall, which is a test for memory problems (two-sided binomial test, $P = 0.0067$) and total mental status, which is a battery that screens for early signs of dementia ($P = 0.0775$). We obtain the weights for a polygenic score by conducting a de novo meta-GWAS analysis of educational attainment just as in the first stage described above, but this time, we exclude the HRS from the Education Sample.

Fig. 3 shows that the score is associated with both of the cognitive health phenotypes. The strength of the protective effect is approximately constant across age categories from age 50–80 y and becomes weaker for total word recall after age 80 y. These associations are essentially unaffected when we control for up to

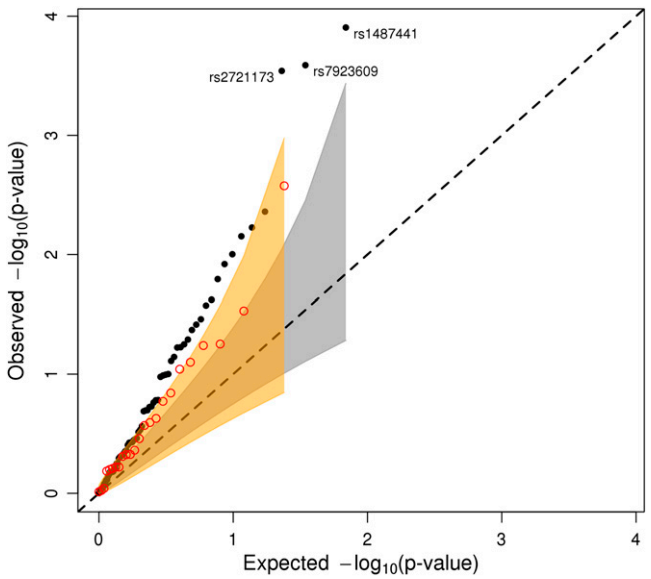

Fig. 2. A Q–Q plot for a regression of cognitive performance on the education-associated SNPs (black circles) with a 95% confidence interval around the null hypothesis (gray shaded region) and a Q–Q plot for a regression of cognitive performance on the theory-based SNPs (red circles) with a 95% confidence interval around the null hypothesis (orange shaded region).

Table 1. SNPs significantly associated with cognitive performance after Bonferroni correction (full results are in *SI Appendix*, Table S4)

						Years of schooling (Education Sample)		Cognitive performance (Cognitive Performance Sample)	
SNP	Chromosome	Base pair position	Nearest gene	Effective allele	Allele frequency	Standardized coefficient	*P* value	Standardized coefficient	*P* value
rs1487441	6	98660615	*LOC100129158*	A	0.473	0.026	1.78×10^{-9}	0.036	1.24×10^{-4}
rs7923609	10	64803828	*JMJD1C*	A	0.521	−0.021	1.06×10^{-6}	−0.034	2.58×10^{-4}
rs2721173	8	145715237	*LRRC14*	T	0.473	−0.020	8.61×10^{-6}	−0.034	2.88×10^{-4}

The chromosome and base pair position are from the National Center for Biotechnology Information genome annotation (build 36), and the nearest gene is from the SCAN database (www.scandb.org/newinterface/about.html). Allele frequency refers to the Cognitive Performance Sample.

20 principal components of the genome-wide data, suggesting that the associations are not driven by population stratification (20). The R^2 values of these associations range roughly from 0.2% to 0.4% (similar magnitudes as in the analysis of cognitive performance in the family-based cohorts). When we control for years of schooling, the estimated effect of the score falls roughly in one-half but remains statistically significant (*SI Appendix*). The score is not associated with cognitive decline (i.e., the change in a cognitive phenotype across longitudinal survey waves), except for total word recall after age 80 y.

Finally, we used 14 (of 69) education-associated SNPs that are nominally significantly associated with cognitive performance ($P < 0.05$) to explore possible biological pathways in a set of bioinformatic analyses (*SI Appendix*); 2 of 14 SNPs are in gene deserts, but the other 12 SNPs are in close vicinity to at least one gene predicted (based on its expression profile) to be involved in the nervous system (*SI Appendix*). Among the most promising genes across these loci are *KNCMA1*, *NRXN1*, *POU2F3*, and *SCRT*, all of which are predicted to be involved in a glutamate neurotransmission pathway [labeled in REACTOME as "unblocking of NMDA receptor, glutamate binding, and activation" (21)] that is involved in synaptic plasticity, a cellular mechanism for learning and memory. Using different methods (but some overlapping data), this same pathway has previously been implicated in human cognitive performance (21).

Discussion

This paper makes two contributions. The first contribution is that we show that the proxy-phenotype method generates positive findings in a domain in which neither candidate gene nor GWAS approaches have so far made substantial progress. Similar approaches have sometimes been used in prior work (e.g., to find rare structural variants associated with cognition) (22), and there is existing work focused on the related idea of increasing statistical power in GWAS by analyzing correlated phenotypes jointly (23, 24).

We propose that the proxy-phenotype method, if systematically applied in social science genetics, could be a useful complement to traditional gene discovery methods (such as GWAS) in cases where it affords greater statistical power. In this case, it does so, because (*i*) much larger genotyped samples are available for educational attainment than cognitive performance and (*ii*) some genetic variants are likely to be associated with educational attainment because of their more direct, stronger relationships with cognitive performance. For the same reasons, educational attainment might similarly serve as a proxy phenotype for personality traits, such as persistence and self-control. In other contexts, the proxy-phenotype method may be better powered for different reasons. For example, for behavioral phenotypes with substantial measurement error—such as smoking, drinking, exercise, or eating habits—the proxy phenotype could be a medical outcome associated with the behavior (e.g., pulmonary disease for smoking or cirrhosis for alcohol consumption). We also note that, although our analysis plan specified that cohorts look up a relatively small set of education-associated SNPs in their existing GWAS results on cognitive performance, researchers with access to full GWAS results on the phenotype of interest could implement a more powerful version of the proxy-phenotype method. For example, first-stage results on the proxy phenotype could inform priors that are updated using GWAS results on the phenotype of interest.

We caution that the proxy-phenotype method (like theory-based candidate SNP approaches) could generate an unacceptably high rate of false positives if it were applied when underpowered and if results were reported selectively. To minimize this danger, we propose a set of best practices that proxy-phenotype studies should follow: researchers should (*i*) conduct power calculations ex ante to justify the use of the method for a particular phenotype of interest and report these calculations in the supplemental information, (*ii*) circulate an analysis plan to all cohorts before conducting any analysis and register the plan in a public repository, (*iii*) commit to publishing all findings from the study, including null results, and (*iv*) conduct Bayesian calculations of the credibility of any findings. We followed these procedures in this paper. Although replication of findings in an independent cohort would be ideal, we anticipate that it will often be infeasible given the unavailability of genotyped samples that may motivate the proxy-phenotype approach in the first place.

The second contribution of this paper is to identify common genetic variants associated with cognitive phenotypes. Knowing

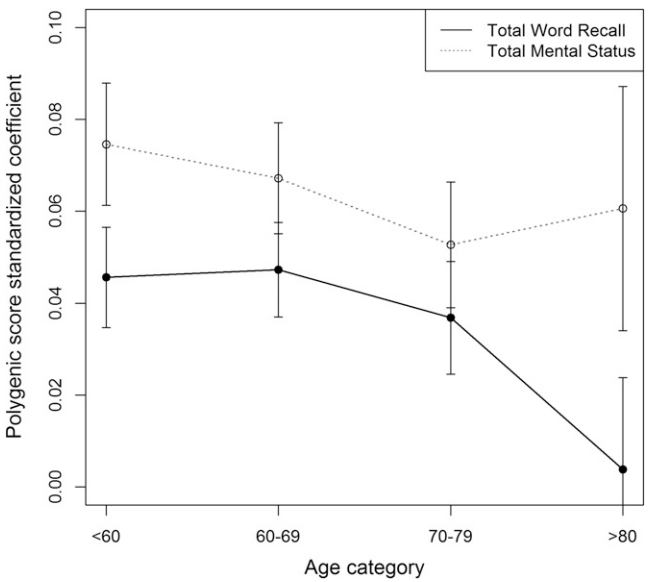

Fig. 3. Coefficients from regression of standardized cognitive phenotype (total word recall or total mental status) on standardized polygenic score within age category controlling for sex and clustering SEs by individual (details in *SI Appendix*, section 14). Error bars show ±1 SE.

the three significant SNPs is not useful for predicting any particular individual's cognitive performance because the effect sizes are far too small, but it does enable follow-up research (e.g., pinpointing the causal variants and then conducting KO experiments in animals) that may ultimately shed light on biological pathways underlying cognitive variation. The polygenic scores constructed from our results may prove useful for studying gene–environment interactions. In future work, the magnitude of explained variance will increase as researchers gain access to datasets with even larger first-stage samples. Our results suggest that such scores hold promise for eventually identifying individuals whose cognitive health at older ages is at greatest risk, which could allow for appropriate preparation and (if possible) preventative intervention.

Materials and Methods

The first stage of our two-stage procedure consisted of conducting a GWAS meta-analysis on years of schooling in a pooled Education Sample (n = 106,736) using the same analysis plan as in the work by Rietveld et al. (11) and the same cohorts, except for omitting the individuals who we include in the second stage. To obtain our set of education-associated SNPs, we selected all SNPs with P value $< 10^{-5}$ from the first-stage meta-analysis results and then pruned for linkage disequilibrium. The second stage of our two-stage procedure consisted of conducting a meta-analysis of these 69 SNPs on high-quality measures of cognitive performance in the independent Cognitive Performance Sample, which included 11 cohorts (n = 24,189). We constructed linear polygenic scores from the meta-analysis results of the second stage and tested them for association with cognitive performance in four family-based cohorts (pooled n = 4,463), with the meta-analysis sample excluding the respective validation sample. Analyses on cognitive health phenotypes were conducted in an independent cohort of older Americans, the HRS, using the two measures that are available in more than one wave in that sample: total word recall and total mental status (n = 8,652 and n = 8,539, respectively). We tested the association between these two phenotypes and a linear polygenic score that was constructed using the coefficient estimates from the GWAS meta-analysis of years of schooling (as in the first stage, excluding only HRS; n = 98,110). *SI Appendix* includes all of the details on the samples and methods.

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 a Material Transfer Agreement (*SI Appendix*). Results from the meta-analysis, the complete biological annotation, and a frequently asked questions document describing the findings of this paper are available at the website of the consortium: www.ssgac.org. The formation of the SSGAC was made possible by a National Science Foundation EArly-concept Grant for Exploratory Research (EAGER) grant and National Institutes of Health (NIH)/Office of Behavioral and Social Science Research (OBSSR) Supplemental Grant SES-1064089. This research was funded, in part, by Ragnar Söderberg Foundation Grant E9/11, Swedish Research Council Grant 412-2013-1061, and National Institute on Aging/NIH Grants P01AG005842, P01AG005842-20S2, P30AG012810, and T32AG000186-23. A full list of acknowledgments is in *SI Appendix*.

1. Bouchard TJ, Jr, McGue M (2003) Genetic and environmental influences on human psychological differences. *J Neurobiol* 54(1):4–45.
2. Plomin R, DeFries J, Knopik V, Neiderhiser J (2013) *Behavioral Genetics* (Worth Publishers, New York).
3. Chabris CF, et al. (2012) Most reported genetic associations with general intelligence are probably false positives. *Psychol Sci* 23(11):1314–1323.
4. Benyamin B, et al.; Wellcome Trust Case Control Consortium 2 (WTCCC2) (2014) Childhood intelligence is heritable, highly polygenic and associated with FNBP1L. *Mol Psychiatry* 19(2):253–258.
5. Davies G, et al. (2011) Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Mol Psychiatry* 16(10):996–1005.
6. Wisdom NM, Callahan JL, Hawkins KA (2011) The effects of apolipoprotein E on non-impaired cognitive functioning: A meta-analysis. *Neurobiol Aging* 32(1):63–74.
7. Lambert J-C, et al.; European Alzheimer's Disease Initiative (EADI); Genetic and Environmental Risk in Alzheimer's Disease; Alzheimer's Disease Genetic Consortium; Cohorts for Heart and Aging Research in Genomic Epidemiology (2013) Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 45(12):1452–1458.
8. Davies G, et al. (2014) A genome-wide association study implicates the APOE locus in nonpathological cognitive ageing. *Mol Psychiatry* 19(1):76–87.
9. Payton A (2009) The impact of genetic research on our understanding of normal cognitive ageing: 1995 to 2009. *Neuropsychol Rev* 19(4):451–477.
10. Visscher PM, Brown MA, McCarthy MI, Yang J (2012) Five years of GWAS discovery. *Am J Hum Genet* 90(1):7–24.
11. Rietveld CA, et al. (2013) GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* 340(6139):1467–1471.
12. Wainwright MA, Wright MJ, Geffen GM, Luciano M, Martin NG (2005) The genetic basis of academic achievement on the Queensland Core Skills Test and its shared genetic variance with IQ. *Behav Genet* 35(2):133–145.
13. Calvin CM, et al. (2012) Multivariate genetic analyses of cognition and academic achievement from two population samples of 174,000 and 166,000 school children. *Behav Genet* 42(5):699–710.
14. Marioni RE, et al. (2014) Molecular genetic contributions to socioeconomic status and intelligence. *Intelligence* 44(100):26–32.
15. Johnson W, Bouchard TJ, Krueger RF, McGue M, Gottesman II (2004) Just one g: Consistent results from three test batteries. *Intelligence* 32(1):95–107.
16. Ree MJ, Earles JA (1991) The stability of g across different methods of estimation. *Intelligence* 15(3):271–278.
17. Chabris CF (2007) *Integrating the Mind: Domain General Versus Domain Specific Processes in Higher Cognition*, ed Roberts MJ (Psychology Press, Hove, United Kingdom), pp 449–491.
18. Lango Allen H, et al. (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 467(7317):832–838.
19. Speliotes EK, et al.; MAGIC; Procardis Consortium (2010) Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 42(11):937–948.
20. Price AL, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38(8):904–909.
21. Hill WD, et al. (2014) Human cognitive ability is influenced by genetic variation in components of postsynaptic signalling complexes assembled by NMDA receptors and MAGUK proteins. *Transl Psychiatr* 4:e341.
22. Stefansson H, et al. (2014) CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature* 505(7483):361–366.
23. Ferreira MAR, Purcell SM (2009) A multivariate test of association. *Bioinformatics* 25(1):132–133.
24. Galesloot TE, van Steen K, Kiemeney LALM, Janss LL, Vermeulen SH (2014) A comparison of multivariate genome-wide association methods. *PLoS ONE* 9(4):e95923.