FAQs about "Common Genetic Variants Associated with Cognitive Performance Identified Using Proxy-Phenotype Method"

The document was prepared by Daniel J. Benjamin, David Cesarini, Christopher F. Chabris, Philipp D. Koellinger, David Laibson, Michelle N. Meyer, and Niels Rietveld.

For clarifications or additional questions, please contact: Daniel Benjamin (daniel.benjamin@gmail.com).

I. This study is the outcome of a collaboration between the Social Science Genetic Association Consortium (SSGAC) and the Childhood Intelligence Consortium (CHIC). What is the SSGAC, and what is the CHIC?

The SSGAC is a research infrastructure designed to stimulate dialogue and cooperation between medical researchers and social scientists. The SSGAC facilitates collaborative research that seeks to identify associations between specific genetic markers (segments of DNA) and social science variables, such as behavior, preferences, and personality. One major impetus for the formation of the SSGAC was the growing recognition that most effects of individual genetic markers on behavioral traits are extremely small, and that, consequently, very large samples are required to accurately measure them. Several years ago medical researchers responded to a similar recognition—that most effects of individual genetic markers on complex diseases are very small—by forming research consortia in which groups collaborate by pooling results across many datasets. These efforts have borne considerable fruit, including recent findings on the genetics of autism (Gaugler et al., 2014) and schizophrenia (Ripke et al., 2014). The SSGAC is an attempt to encourage analogous pooling among social-science geneticists and is organized under the auspices of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE), a successful medical consortium. The SSGAC was founded by three social scientists (Daniel Benjamin, David Cesarini, and Philipp Koellinger) who are excited about the potentially transformative impact that genetic data could have on the social sciences. The Advisory Board for the SSGAC is composed of prominent researchers representing various disciplines: Dalton Conley (Sociology, New York University), George Davey Smith (Epidemiology, University of Bristol), Tõnu Esko (Molecular Genetics, Broad Institute and Estonian Genome Center), Albert Hofman (Epidemiology, Erasmus University), Robert Krueger (Psychology, University of Minnesota), David Laibson (Economics, Harvard), Sarah Medland (Statistical Genetics, Queensland Institute of Medical Research), Michelle Meyer (Bioethics, Union Graduate College and Icahn School of Medicine at Mount Sinai), and Peter Visscher (Statistical Genetics, University of Queensland). The pilot project of the SSGAC was a large-scale GWAS on educational attainment whose results were published in *Science* (Rietveld et al., 2013).

CHIC was established to combine the efforts of several research centers to elucidate the genetic and environmental bases of differences in cognitive performance among children. The results of the first project of CHIC were published in *Molecular Psychiatry* (Benyamin et al., 2014) and are based on data from six discovery (N = 12,441) and three replication (N = 5,548) cohorts with a total sample size of 17,989 children of European ancestry for whom genome-wide single nucleotide polymorphism (SNP) genotypes and cognitive performance scores were available.

Prior to the current study published in *Proceedings of the National Academy of Sciences*, the CHIC initiative was the largest study that aimed to identify genetic variants associated with cognitive performance.

The current project is a collaboration between the SSGAC and the CHIC. We combined our complementary resources and expertise, making the current project the largest effort to investigate genetic associations with cognitive performance to date.

II. What is "cognitive performance"?

We use the term cognitive performance to refer to the fact that, even though different types of cognitive tests sometimes differ from each other in important ways, people who perform well on one type of cognitive test also tend on average to perform well on other types of cognitive tests. To measure cognitive performance, it is standard practice to summarize how well participants did on a range of cognitive tests with a statistical technique called "factor analysis." This technique attempts to transform the observed performance on many tasks into a single score such that the tasks are uncorrelated in any subpopulation where all individuals have the same score value. In certain cases the results of factor analysis are numerically similar to those of "principal component analysis," which finds a weighted sum of the scores that accounts for as much of the variability in the data as possible. Scores such as these (or similar ones) serve as our measure of cognitive performance.

We use the term "cognitive performance" rather than the more traditional terms "cognitive ability" or "general intelligence" because it does not prejudge the sources or causes of differences among people, which could be explained by genetic factors, environmental factors, their interactions, or all of these. Performance (as opposed to ability or intelligence) is a neutral term intended to capture only the fact that there are stable, systematic differences in how people perform on cognitive tests, without making any claims about whether they are (a) innate or acquired and (b) fixed or malleable. As the distinctness of these two sets of questions suggests, it is entirely possible that traits that are significantly innate, because they are under genetic influence, are also quite malleable. For example, eyesight is very strongly influenced by genes, but through the environmental interventions of eyeglasses, contact lenses, and surgery, that trait is also highly malleable.

Furthermore, although "cognitive performance" is a key construct for the social sciences and medicine, it is not the only aspect of cognition that matters. There are many other cognitive skills, some of which may be independent of "cognitive performance" as our study measures it, that have considerable importance in different settings and for different outcomes. For example, in some fields creativity and social skills may contribute to success to an equal or even greater extent than cognitive performance.

III. What was already known about the genetics of cognitive performance prior to this study?

Decades of twin and family studies suggest that cognitive performance is influenced by genes (1). For example, these studies have consistently found that identical twins are far more similar to each other in their cognitive performance than fraternal twins are (2).

However, despite considerable interest and effort, research to date has largely failed to identify specific common genetic variants (DNA sequence variations that occur commonly within a population as opposed to "rare variants"[1]) associated with cognitive performance in the normal range. "Candidate gene studies," which investigate a small set of genes that are believed to have well-understood functions and potential relevance for cognition, have led to many published association results that have not consistently replicated (3). Other research approaches which investigated a large set of common genetic variants using "genome-wide association studies (GWAS)" on cognitive performance have not found replicable results either (4, 5), with the exception of variants in the gene *APOE,* which significantly predicts cognitive decline in older individuals (6–8).

IV. What did you do in this particular study on cognitive performance?

Our current paper builds on the insights gained by previous research efforts which suggested that the effect sizes of any common genetic variant on cognitive performance are likely to be so small that they cannot be detected using existing data and research approaches (4, 5). In contrast to these earlier efforts, we applied an alternative, two-stage research strategy, which we call the proxy-phenotype method. In the first stage, we conducted a genome-wide association study on a "proxy phenotype" (in this case, educational attainment) that is correlated with the phenotype of interest (in this case, cognitive performance) and is well measured in a much larger sample of respondents. This first stage identified a relatively small set of SNPs (a type of common genetic variant) that are associated with educational attainment. In the second stage, these SNPs serve as empirically plausible candidates that we tested in independent samples for association with cognitive performance. This two-stage approach has more statistical power than previously employed research methods, allowing us to identify robust results that are likely to replicate. (For further details on the proxy-phenotype method, see "How is your approach—the "proxy-phenotype method"—different from prior work?" below.)

V. What did you find?

First, we found three common genetic variants that are robustly associated with cognitive performance. These associations remained statistically significant even when we took account of the fact that we tested many more SNPs than these three. The effect of these variants was extremely small, accounting for approximately 0.3 points per copy of each variant on the familiar IQ scale (which has a mean of 100 and standard deviation of 15). According to our analysis of these three genetic variants, the *largest* possible difference between two individuals would occur if the individuals were identical except that one of them possessed two copies of each of the

[1] "Rare variants" refer to DNA sequences that are identical for the vast majority of individuals, with only a small fraction of individuals having a different sequence. Very recently, (12) used an approach similar to ours to identify rare genetic variants associated with cognition.

three variants associated with greater cognitive performance and the other possessed no copies of any of the three variants. In this extreme—and rare—case, the former individual would be expected to score about 1.8 points higher on an IQ test. At least 98% of people will not possess the full six copies of the "positive" variants; indeed the typical person would possess three, with others differing by no more than 0.9 IQ points above or below the average—*as a result of these three SNPs*.

Second, in an independent sample of older Americans (N = 8,652), we found that a polygenic score derived from 60 of the education-associated SNPs from stage 1 is associated with memory and absence of dementia, which illustrates the potential health relevance of our findings. (A polygenic score is an index that combines many different SNPs to create an overall measure that predicts variation in some behavior or trait of interest. In this case, we found that when considered as a group, the education-associated SNPs were also able to predict memory and cognitive health.)

Third, we conducted a range of bioinformatics analyses to gain further insights into the biological mechanisms of cognitive performance that are implicated by our genetic association results. We stress that the accuracy of our conclusions is limited by current knowledge of the neurological functions of genes. As understanding of the underlying biological function of different genes develops over time, so will our ability to interpret findings arising from studies such as this one. Thus, strong conclusions about underlying biological mechanisms would be premature. With this caveat in mind, we nevertheless believe that our findings regarding biological pathways are intriguing. The convergence of all of these analyses suggests that four specific genes (*KNCMA1*, *NRXN1*, *POU2F3*, *SCRT*) are involved in differences between people in cognitive performance. All four of these genes happen to be associated with a single neurotransmitter pathway involved in synaptic plasticity, which is the brain's main mechanism for learning and memory.

Cognitive performance is a complex phenomenon that is influenced by a large number of both genetic and environmental factors, and our study focuses on only a tiny piece of the puzzle. We only examine a small set of common genetic variants, we consider only one of many forms of genetic difference among individuals, and we do not include specific measures of environmental factors. There are additional sources of genetic variation that remain to be discovered. It is also likely that the effects of variants—both those we found and those remaining to be discovered—will differ based on environmental conditions (for example, a variant associated with greater cognitive performance may only have that effect in the context of schooling of sufficient quality). These other genetic effects, environmental effects, and interactions between genetic and environmental factors are important topics of active research.

VI. How is your approach—the "proxy-phenotype method"—different from prior work?

Prior work has typically relied on one of two research strategies. The first is the candidate-gene study, 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). This suggests that the candidate-gene approach is susceptible to false positive results that can mislead researchers and misdirect research efforts. The second strategy is the genome-wide association study (GWAS), in which researchers test millions of single-nucleotide polymorphisms (SNPs) for association with the phenotype, without regard for whether these SNPs might be biologically related to the phenotype, and then apply an extremely stringent threshold for determining statistical significance in order to reduce the likelihood of false positives. For physical and medical phenotypes, GWAS methods have identified many novel associations that do in fact replicate (10). GWAS methods on cognitive performance, however, have not yet identified any genome-wide-significant associations (4, 5).

The proxy-phenotype method that we applied in our study is an alternative that combines features of both of these previous approaches. In the first stage, we conduct a GWAS on a "proxy phenotype" to identify a relatively small set of 69 SNPs that are associated with the proxy phenotype. In the second stage, these 69 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. Hence, no additional data collection was required for our study (we leveraged previous GWAS efforts on educational attainment and cognitive performance).

As far as we aware, prior work (with the exception of our own previous study on educational attainment (11)) has not laid out the statistical framework and conducted calculations to evaluate whether and when identifying candidates via a proxy phenotype gives a study more statistical power (i.e., makes a study more likely to find true positive results) than would a direct GWAS-style analysis of the phenotype of interest. We do this in the online supplement to the present paper. Conducting such calculations is crucial for understanding why our analysis in this paper succeeds in identifying novel associations with cognitive performance even though the previous GWAS papers have not.

Although related ideas have been discussed before (12), the proxy-phenotype method has not generally been applied as an alternative to the GWAS and candidate-gene approaches. We think its application could prove fruitful when the main phenotype of interest is not available (yet) in large enough samples for GWAS, but a correlated phenotype is available in large samples. For example, in addition to educational attainment, GWAS results from the widely-available phenotypes of height and body-mass index (BMI) could provide proxies for discovering genes associated with conditions that are correlated with those phenotypes, thereby speeding progress on significant behavioral and medical conditions.

We caution, however, that the proxy-phenotype method (like theory-based candidate-SNP approaches) could generate an unacceptably high rate of false positives if it were applied without high statistical power (i.e., large enough samples) and if results were reported selectively (i.e., publishing successful discoveries of association with the ultimate phenotype and not failures). To avoid this, we propose a set of "best practices" that proxy-phenotype studies should follow: researchers should (a) conduct power calculations in advance to justify the use of the proxy-phenotype method for a particular phenotype of interest and report these calculations in their papers; (b) circulate a data analysis plan to all participating cohorts prior to conducting any

analysis and register the plan in a public repository; (c) commit to publishing all findings from the study, including null results; and (d) conduct Bayesian calculations that provide an estimate of how likely the results are to be false positives.

VII. Why is it important to know that genetic effect sizes for traits like cognitive function are extremely small?

The effect size of an individual genetic variant on a behavioral trait determines which research strategies will succeed and which will fail. The tiny effect sizes for genetic variants identified in our study suggest that studies seeking to identify genetic influences on behavioral traits should include at least tens of thousands of research participants in order to generate accurate results. The effect sizes here are so tiny that, given the size of our sample, we could only identify them with the two-stage approach we employed.

The small effects of individual genetic variants also imply that the genetic influences on cognitive performance result from the cumulative effects of hundreds or thousands of different genetic variants, not just a few. When a variant has a small effect size, therefore, its presence or absence has practically no relevance at all for predicting any particular individual's overall cognitive performance.

VIII. If effect sizes are so small, why bother studying them?

Despite their tiny effect sizes, identifying genetic variants related to behavioral traits could be useful for a number of reasons. Here are two examples:

First, even if a genetic variant has a very small effect, identifying it may lead to insights regarding the biological pathways involved in the phenotype. For example, one very successful cholesterol-lowering drug works by targeting a biological mechanism that was identified in a GWAS on circulating lipid levels, even though the effect sizes in the GWAS were small (13, 14). It is possible that studies on the genetics of cognitive performance may eventually shed light on biological mechanisms underlying memory loss and dementia.

Second, even if an *individual* genetic variant has a very small effect, many genetic variants taken together may explain a large share of the variance in a trait. For example, estimates from previous research suggest that, if extremely large samples were available to conduct genome-wide association studies on cognitive performance, as much as 29% of the differences among people in cognitive performance might be captured by all currently known common genetic variants, considered as a group (15). This amount of statistical accuracy is still too low to be relevant for predicting any one person's cognitive performance, but it would be useful for taking account of genetic factors when studying the effect of an educational or health policy. When social scientists study expensive policy interventions, such as providing preschool to disadvantaged children, controlling for as many factors as possible can help generate more accurate estimates of the effectiveness of the policy. Just as one would not want to analyze the effects of education without considering background factors like family income, one would like to consider genetic differences as well; indeed, not doing so could lead to inaccurate conclusions and ineffective policies.

IX. What are the contributions of your study?

In addition to the most direct contribution—identifying three SNPs that are robustly associated with cognitive performance—we believe that our study also makes a number of other contributions. Here are two:

First, our paper describes and demonstrates a method that genetics researchers can use in future work. If a trait of interest is not available (yet) in a sample that is large enough to conduct a GWAS directly, but a strongly correlated trait is available, the proxy-phenotype method can be used to increase the chance of finding statistically meaningful results for the trait of interest. Thus, this two-stage approach can make more efficient use of available data resources, enabling discoveries and insights that might otherwise have to wait for the creation of larger genetic databases suitable for GWAS.

Second, our study provides additional evidence about the effect sizes that can be expected for associations between individual genetic variants and complex behavioral traits. These effect-size estimates—which are roughly one-twentieth as large as those found for human height, a complex *physical* trait—will help researchers determine the sample sizes they should use in social-science genetics research. The estimates are also useful for assessing how much to trust existing reports of genetic associations from small samples that lack the statistical power to measure small genetic effects.

These and other contributions—such as our exploration of potential biological pathways that might underlie the associations we observe—are likely to trigger follow-up research in various disciplines (such as psychology, economics, and other social sciences, as well as epidemiology, medicine, and biology).

X. What policy lessons concerning cognitive performance do you draw from this study?

None. *Any* practical response—genetic or environmental, individual or policy-level—to this or similar research would be extremely premature. In this respect, our study is no different from genome-wide association studies (GWAS) of complex medical outcomes. In medical GWAS research, it is well understood that the known genetic variants are not yet predictive enough of complex diseases to have significant value for assessing the risk to any given individual. Our current paper, like our previous ones, shows that most genetic effects on cognitive performance and other behavioral outcomes are probably even smaller and more diffuse.

However, studies like ours may provide scientists with tools that enable them to better understand the effects of government policies, which may eventually result in better-designed policies. For example, consider a program that aims at improving the education of disadvantaged children. Such a program could easily be very expensive, so a government may decide to run a small, randomized field experiment to test the effectiveness of the intervention before implementing it on a larger scale. Research that aims to evaluate the effectiveness of this intervention may benefit from including variables derived from genetic data (e.g., polygenic

scores) that can improve the accuracy of the estimated effect of the policy (for more detail, see "If effect sizes are so small, why bother studying them?" above).

XI. Did you find "the gene" for cognitive performance?

No. We did not find "the gene" for cognitive performance—or anything else. Cognitive performance, like most complex behaviors and outcomes, is influenced by a very large number of genes, each with effects that are likely to be tiny—as well as a huge host of environmental factors, and probably also interactions between genetic and environmental factors. If anything, our results reinforce the claim that there is no such thing as "the gene for cognitive performance," just as there is no such thing as "the gene for athleticism" or "the gene for social skill," and so on.

XII. Did you find the most relevant genes for cognitive performance?

Probably not. Our study only looked at a small number of common genetic variants. Many other genetic variants exist that our study did not investigate, and some of them may be even more relevant for cognitive performance than the ones we identified.

XIII. Does this study show that an individual's level of cognitive performance is determined at birth?

No, and this is probably one of the most common misconceptions about genetic research. Even if it were true—and it is certainly not—that genetic factors accounted for *all* of the differences among individuals in cognitive performance, it would not follow that the ability to perform well on cognitive tests is "determined" at birth (or at conception). There are two reasons for this:

First, some genetic effects may operate through environmental channels. For example, suppose that the genetic variants we identified help students to focus longer in school and, as a result, to become more efficient in completing tests of cognitive performance. In this case, changes to the intermediate environmental channels—the structure and organization of schooling—could have drastic effects on the outcome. For example, the genetic association might not be found at all among children without access to school education, and it might not apply as strongly if the education system were organized differently than it is at present.

Second, even if the genetic effects operated entirely through non-environmental mechanisms that are difficult to modify, such as direct influences on the formation of neurons and the chemical interactions among them, there could still exist powerful environmental interventions that do not contribute to differences across individuals in the current population. In a famous example attributable to the economist Arthur Goldberger, even if all the variation in eyesight were due to genes, there could still be enormous benefits from introducing eyeglasses. Similarly, policies such as a required minimum number of years of education and help for individuals with learning disabilities can increase cognitive performance in the entire population and/or reduce differences among individuals.

In sum, we found a few common genetic variants associated with cognitive performance, each of

which we estimate to have only a tiny effect on that outcome. But even if we had found that a single gene has a very large effect on cognitive performance, that finding would be perfectly consistent with the possibility that environmental factors or interventions can modify or even cancel out this influence; traits that have even a large genetic component are not necessarily immutable. For instance, the metabolic disorder phenylketonuria (PKU) is caused by gene mutations that prevent the carrier from metabolizing the amino acid phenylalanine. Without environmental intervention, PKU leads to medical problems and mental retardation. But through early genetic detection and an environmental intervention—namely, maintaining a special diet free of phenylalanine, monitoring protein levels, and taking daily medication—individuals with PKU can lead lives with normal cognitive function and life expectancy.

XIV. How are your findings relevant for health?

There is a well-known relationship between cognitive performance and health outcomes, and this connection was one important motivation for our study (16, 17). Indeed, some of the genetic variants we identify may be associated with cognitive performance because of their effects on health (which, in turn, could impact performance in cognitive tests).

The "polygenic score" that we have estimated predicts absence of dementia, suggesting a genetic link between cognitive performance in children, educational attainment through early adulthood, and cognitive health among the elderly. This is an interesting insight from a medical perspective because research on Alzheimer's disease and dementia has clearly shown that highly educated individuals are less likely to develop these mental disorders and, if they do, the onset of the disease is typically delayed. Exactly why this is so is still unknown (18). Our findings suggest that at least part of the protective effect of education seems to come from genetic factors that influence cognitive health and performance across the entire lifespan. Furthermore, it is possible that the biological mechanisms underlying our genetic association results could someday have direct relevance for diagnosing and treating dementia and related disorders such as Alzheimer's disease. Obviously, more research will be needed to confirm these connections.

XV. What are the potential benefits of the type of work you are doing?

We believe that our type of work can eventually prove beneficial in several ways:

1. It can clarify popular misunderstandings. For example, properly conducted genetics research in the social sciences has made clear the limits of deterministic views of complex traits by establishing accurate upper bounds for effect sizes and prediction accuracy.

2. It can decrease the risk of discrimination and stigmatization. Accurate estimates of the upper bounds of genetic effect sizes and prediction accuracy help confirm that claims about "a gene for X" are typically not warranted for genetically complex traits such as behavior or cognition.

3. It can compensate for disadvantage. For example, new insights into the biology underlying cognitive health can point to new therapeutic leads, or help understand how genes translate into good or poor outcomes through modifiable environmental channels.

4. It can help individuals to project their own risks. For example, an individual planning retirement may want to know about his or her risk of cognitive decline. In many cases, knowledge about risks could help people take preemptive actions that reduce the disease risk or mitigate its effects on themselves, their families, their finances, and so on.

5. It can improve the quality of policy research in the social sciences. For example, polygenic scores can be used to improve our understanding of the effectiveness of interventions designed to improve educational outcomes and cognitive performance for individuals from disadvantaged backgrounds.

6. In the longer-term, learning more about the genetic influences on phenotypes related to cognitive health may lead to effective environmental interventions (as in the case of PKU). In order to study effective policies to reduce gaps in educational attainment or similar disparities, it is helpful to account for genetic effects on those outcomes. In addition, identifying genetic variants that contribute to differences in cognitive performance may lead to insights regarding the biological pathways underlying that outcome, and may provide a firm foundation for research on interactions between genetic and environmental influences.

Of course, our study is only a tiny step toward achieving these ultimate potential benefits.

XVI. Could this kind of research lead to discrimination against, or stigmatization of, people with the relevant genotypes?

There is a risk that behavioral genetic research findings may, whether wilfully or not, be misinterpreted and used to stigmatize or discriminate. One response to this risk is to abstain from conducting behavioral genetics research. In general, however, we do not think that the best response to the possibility that useful knowledge might be misused is to refrain from producing that knowledge. Indeed, there are at least two major ethical problems with abstention.

First, behavioral genetics research, including studies of the relationships between genes and a variety of cognitive traits, is already being conducted and will continue to be conducted. Unfortunately, not all of this work involves sound methodologies or transparent communication of results. In this context, researchers who are committed to developing, implementing, and spreading best practices for conducting and communicating potentially controversial research, including behavioral genetics research, arguably have an ethical responsibility to *participate* in the development of this body of knowledge, rather than abstain from it and hope for the best.

For instance, an important theme in our earlier work has been to point out that most existing studies in social-science genetics that report genetic associations with behavioral traits have serious methodological limitations, fail to replicate, and have likely resulted in false positives (Benjamin et al. 2012; Chabris et al. 2012). A leading journal for the genetics of behavioral traits similarly concluded that "it now seems likely that many of the published [behavior genetics] findings of the last decade are wrong or misleading and have not contributed to real advances in knowledge" (Hewitt 2012). One of the most important reasons why existing work has generated unreliable results is that their sample sizes were far too small, given that the true effects of individual genetic markers on behavioral traits are tiny (Hewitt 2012).

Even when well conducted, research results may not be well communicated. Researchers working in this area should be careful to avoid language that is easily misinterpreted, and to communicate clearly not only the potential benefits of the research, which motivate us to do the work, but also the work's limitations.

Second, the results of behavioral genetics research can be used to discourage stigmitization. One benefit of recent behavioral genetics research is that it has made clear the *limits* of deterministic views of complex traits by establishing accurate upper bounds for the amount of variance attributable to common genetic variants—thus perhaps making discrimination and stigmatization *less* likely in the future. Existing claims of genetic associations with complex social-science traits have reported widely varying effect sizes in terms of variance account for (i.e., R^2)—many of them purporting to explain ten to one hundred times as much variance as did the genetic variants we found in this study. Our evidence, from a much larger sample, suggests instead that individual SNPs associated with cognitive performance have about *one-twentieth* as large effects as do SNPs for complex physical traits, such as height, that are also influenced by many separate genes.

References

1. Plomin R, DeFries J, Knopik V, Neiderhiser J (2013) *Behavioral Genetics* (Worth Publishers).

2. Bouchard TJ, McGue M (2003) Genetic and environmental influences on human psychological differences. *J Neurobiol* 54:4–45.

3. Chabris CF et al. (2012) Most reported genetic associations with general intelligence are probably false positives. *Psychol Sci* 23:1314–1323.

4. Benyamin B et al. (2014) Childhood intelligence is heritable, highly polygenic and associated with FNBP1L. *Mol Psychiatry* 19:253–258.

5. Davies G et al. (2011) Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Mol Psychiatry* 16:996–1005.

6. Wisdom NM, Callahan JL, Hawkins KA (2011) The effects of apolipoprotein E on non-impaired cognitive functioning: A meta-analysis. *Neurobiol Aging* 32:63–74.

7. Lambert J-C et al. (2013) Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 45:1–9.

8. Davies G et al. (2014) A genome-wide association study implicates the APOE locus in nonpathological cognitive ageing. *Mol Psychiatry* 19:76–87.

9. Payton A (2009) The impact of genetic research on our understanding of normal cognitive ageing: 1995 to 2009. *Neuropsychol Rev* 19:451–77.

10. Visscher PMM, Brown MAA, McCarthy MII, Yang J (2012) Five years of GWAS discovery. *Am J Hum Genet* 90:7–24.

11. Rietveld CA et al. (2013) GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science (80-)* 340:1467–1471.

12. Stefansson H et al. (2014) CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature* 505:361–366.

13. Edwards SL, Beesley J, French JD, Dunning AM (2013) Beyond GWASs: illuminating the dark road from association to function. *Am J Hum Genet* 93:779–97.

14. Teslovich TM et al. (2010) Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 466:707–713.

15. Marioni RE et al. (2014) Molecular genetic contributions to socioeconomic status and intelligence. *Intelligence* 44:26–32.

16. Batty GD, Deary IJ, Gottfredson LS (2007) Premorbid (early life) IQ and later mortality risk: systematic review. *Ann Epidemiol* 17:278–88.

17. Auld MC, Sidhu N (2005) Schooling, cognitive ability and health. *Health Econ* 14:1019–34.

18. Gatz M et al. (2006) Potentially modifiable risk factors for dementia in identical twins. *Alzheimer's Dement* 2:110–117.

XVII. Technical FAQs (September 10, 2014)

Our paper has generated several on-line conversations. As we read these discussions, we felt that it might be helpful to supplement our previously circulated FAQs (above) with some additional observations.

1. You only found 3 SNPs associated with cognitive performance, each with very small effect sizes. Does this mean that common genetic variants contribute very little to variation among individuals in cognitive performance?

No. For cognitive performance (CP), several papers, beginning with (5) have found that the common variants measured on existing genotyping platforms account for a substantial share of the heritable variation in CP across individuals. The point estimates tend to hover in the 40-50% range. For the CP phenotype, we therefore anticipate that with large enough samples, we will be able to identify common variants that can account for a substantial fraction of the heritable variation. Our findings can be viewed as a small first step in that process.

2. Are your study's findings likely to be false positives because the effects are small?

We don't think so. We have great sympathy for concerns about the inferential challenges that arise in these kinds of studies. Indeed, we have published numerous papers arguing that most candidate gene associations with behavioral traits are probably spurious because the studies lacked adequate power and used *p*-value thresholds that were too liberal (what matters for the Bayesian calculations is not just statistical power, but also the nominal significance threshold). Because individual-SNP effects are now known to be very small and because we worry about false positives, we gathered a large sample and conducted power calculations in order to design our analysis plan to maximize the likelihood of true positives.

Our research design was motivated by what we believe to be a correct diagnosis of the causes of the replication problems that plague candidate gene approaches. The SI of our paper contains a Bayesian analysis that explains why our methodology has sufficient statistical power to generate credible results. (We note that the results included one SNP that both reached genome-wide significance and replicated at a Bonferroni-corrected 5% level; see FAQ #4, below.)

We agree with the skeptics that one should be critical of conventional candidate gene studies (with small samples), and one should worry that effect sizes inferred from small samples won't replicate. But it is possible to conduct a study that has sufficient power to generate credible results with sufficiently large samples and corrections for multiple testing, as we show in our Bayesian analysis (see SI). Moreover, it is important to understand that ours is not a conventional candidate gene study: we explicitly contrast our findings to those of conventional candidate gene studies (and GWA studies), and then we carry out statistical calculations to explain why our methodology had some successes when others have failed to date.

Finally, some critics have expressed the intuition that our findings are likely to be false positives because the SNPs' effect sizes are smaller than the measurement error in cognitive performance. Interpreted as a statistical claim, that intuition is incorrect; measurement error makes it harder to

detect effects, not easier. We also note that if "smaller than measurement error" were the criterion for spotting a likely false positive, then many other well-regarded papers would be similarly suspect, including all published height GWA studies. The height SNP with greatest explanatory power explains a percentage of variance (0.4%) that is an order of magnitude smaller than the percentage of transitory variance in measured human height (4%, since the retest reliability of height is ~0.96). Alternatively, one might argue that, as scientists, we should not be interested in finding SNPs whose effect sizes are smaller than the measurement error. We disagree for a number of reasons—one reason being that it is important to establish the magnitude of SNPs' effect sizes so that it is possible to evaluate the credibility of reported findings and to design well-powered studies.

3. Is this study just a failed GWAS? Did you data-mine until you found something?

No and no. The study is not a failed GWAS. We published the GWAS findings on educational attainment last year (11). In the current study, relative to our previous GWAS, we removed from the Education Sample approximately 20,000 individuals from seven datasets that had high-quality measures of cognitive performance. We did this in order to maximize the Cognitive Performance Sample size. It was clear from our power calculations that our results were much more likely to be true positives if the 20,000 individuals were included in the Cognitive Performance Sample rather than the Education Sample. We also chose our *p*-value threshold for the education analysis to be more liberal than genome-wide significance in order to maximize the number of true positives likely to result from the overall, two-stage procedure (rather than just from the first-stage GWAS).

We did not data-mine until we found something. We agree that data-mining is a danger and believe that unfortunately such practices are widespread. It was precisely to minimize such concerns that we posted our planned analyses in an online repository prior to accessing the data. In the paper, we say:

> "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 the 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 the findings. We followed these procedures in this paper."

In other words, it was our ex-ante plan to conduct the study in two stages, and we tied our hands by specifying the *p*-value threshold (and many other aspects of the analyses) prior to accessing the data.

4. Did you have any genome-wide significant hits in the first-stage GWAS?

Yes. We had one SNP with a genome-wide significant association with educational attainment (see Table 1: rs1487441) which subsequently replicated at $p = 1.24 \times 10^{-4}$ with cognitive performance. However, we did not focus on genome-wide significance because the study design was not a GWAS. As noted above, our calculations indicated that applying a *p*-value threshold less strict than genome-wide significance would improve our power for the overall, two-stage design to detect SNPs for which there is credible evidence of association. Again, these calculations were made prior to accessing the data.

We hope this helps bring some more clarity to the paper and its findings. We want to reiterate that we find the skeptical reactions understandable given the methodological track record of many studies in this area.

SSGAC