

Most Reported Genetic Associations with General Intelligence Are Probably False Positives

Christopher F. Chabris (Union College)*
Benjamin M. Hebert (Harvard University)
Daniel J. Benjamin (Cornell University)
Jonathan Beauchamp (Harvard University)
David Cesarini (New York University)
Matthijs van der Loos (Erasmus University Rotterdam)
Magnus Johannesson (Stockholm School of Economics)
Patrik K.E. Magnusson (Karolinska Institutet)
Paul Lichtenstein (Karolinska Institutet)
Craig S. Atwood (University of Wisconsin-Madison)
Jeremy Freese (Northwestern University)
Taissa S. Hauser (University of Wisconsin-Madison)
Robert M. Hauser (University of Wisconsin-Madison)
Nicholas Christakis (Harvard University)
David Laibson (Harvard University)

Supporting Online Material

Previous Replication Attempts for SNPs Under Study

The SNPs we considered in our studies were the ones mentioned by Payton's review (2009) as having published associations with measures of g that were also available in the WLS dataset (the dataset with the largest number of SNPs discussed by Payton, among the datasets available to us). Tables 1–4 of Payton (2009) list the genes and the published studies. Here, for each of our 12 genotypes, we note whether there were published replications of the original finding associating them with g .

For rs429358 and rs7412 in APOE (which define the e2/e3/e4 haplotype associated with Alzheimer disease), a meta-analysis of 77 studies including 40,942 healthy individuals reported a “small effect” on g (Wisdom et al., 2009).

For rs6265 in BDNF, 9 out of 11 studies with a mean $N = 382$ reported an association with g (Miyajima et al., 2008a, 2008b).

For rs2061174 in CHRM2, there were two replications of the original association, with $N = 762$ and $N = 2,158$.

For rs8191992 in CHRM2/CHRNA4, there was one replication, with $N = 2,158$.

For rs4680 in COMT, a meta-analysis of 46 studies including 9115 individuals reported an association explaining 0.1% of the phenotypic variance in g (Barnett et al., 2008).

For rs17571 in CTSD, there were no replications.

For rs821616 in DISC1, there were no replications.

For rs1800497 in DRD2/ANKK1, there were no replications.

For rs1018381 in DTNBP1, there were no replications.

For rs760761 in DTNBP1, there were no replications.

For rs363050 in SNAP25, there were no replications.

For rs2760118 in SSADH (aka ALDH5A1), there were no replications.

Additional Methods for Study 1

DNA was extracted from saliva samples collected in 2006–2007 using Oragene saliva collection kits. Genotyping was performed by KBioscience (Hoddesdon, UK) using homogeneous Fluorescent Resonance Energy Transfer technology. They used the SNP assay genotyping system KASP for 90 SNPs selected because associations between these SNPs and a variety of phenotypes (including *g* and many others) had been previously published.

Of the initial 15,536 participants enrolled in WLS, 6,908 had data for all the covariates and were missing fewer than 10 of the 90 SNPs that had been genotyped. Of this sample, 4,481 were graduates and 51% of the sample was male. Less than 1% of the sample self-identified as a race other than White/Caucasian, 8% refused to identify their race, and 91% of the sample self-identified as White/Caucasian.

Additional Methods for Study 2

The 40–100 year age range at the time of testing is approximate, as the birth year was inferred from age at each FHS exam and approximate date of each FHS exam. Very few subjects were close to the upper end of this range.

Many of the FHS subjects came from the same families because the Offspring cohort is made up of the descendants of the Initial cohort and the spouses of the descendants. The Framingham population was overwhelmingly White/Caucasian at the time these cohorts were

enlisted, and 99.6% of the Third Generation cohort (the descendants of the Offspring cohort) self-identified as White/Caucasian.

Genomic data imputation had been conducted at the Broad Institute and was made available to other users of the FHS data. Genotypic data from the Affymetrix 500K and the MIPS 50K genotyping platforms were combined for the imputation; after filtering out 156,819 SNPs that were likely to have been incorrectly genotyped, 378,163 SNPs were left for the imputation. (SNPs were considered problematic and not used if they failed one of several standard quality control tests, including being out of Hardy-Weinberg equilibrium—at $p < .000001$, a stringent threshold to account for multiple hypothesis testing—being missing in more than 3% of the sample, being absent from the HapMap, having frequency less than 1%, and others.) MACH (version 1.0.15) was used to impute all autosomal SNPs on HapMap, using the publicly available phased haplotypes from HapMap (release 22, build 36, CEU population) as a reference panel. All 10 SNP genotypes analyzed here were in Hardy-Weinberg equilibrium.

Tests for association used the following covariates as control variables: a cubic of age, a cubic of age interacted with sex, a dummy for FHS cohort membership, and the first ten principal components of the genetic data (to control for population stratification). The non-independence of standard errors for individuals in the same family is accounted for by clustering (Liang & Zieger, 1986) at the level of the extended family.

Additional Methods for Study 3

Between December 2010 and May 2011, 10,946 SALT respondents were genotyped by the SNP&SEQ Technology Platform, Uppsala, using the Illumina HumanOmniExpress BeadChip genotyping platform. A total of 79,893 SNPs were omitted because their minor allele frequency

was lower than 0.01, 3,071 markers were excluded because they failed a test of Hardy-Weinberg equilibrium at $p \leq 10^{-7}$, and 3,922 SNPs were missing in more than 3% of the sample.

IMPUTE Version 2 (Howie et al., 2009) was used to impute all autosomal SNPs on HapMap, using the publicly available phased haplotypes from HapMap2 (release 22, build 36, CEU population) as a reference panel. The principal components of the genotypic data were constructed using the same method as in Study 2. All 10 SNP genotypes analyzed here were in Hardy-Weinberg equilibrium.

Cognitive ability test data were manually retrieved from archives for all monozygotic (MZ) and same-sex dizygotic (DZ) twins born between 1936 and 1950. For later cohorts, the information has been digitized, so data on all male twins born after 1950, including men from opposite-sex pairs, was obtained from the Swedish National Service Administration. With the exception of males in opposite-sex pairs born before 1951, we successfully recovered the test scores of over 95% of the males born between 1936 and 1958.

According to Cesarini (2010), the quality of the cognitive data is also supported by high sibling correlations in performance on the test: $r = .822$ in monozygotic twins and $r = .534$ in dizygotic twins. The correlations for other sibling types (adoptees, full and half siblings reared together or apart) are also in line with consensus estimates from the literature (Bouchard, 1998). To account for non-independence within families, we used the same clustering technique as in the analysis of the FHS data.

Additional References

- Cesarini, D. (2010). Family influences on productive skills, human capital and lifecycle income. In *Essays on genetic variation and economic behavior* (Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA). [<http://dspace.mit.edu/handle/1721.1/57897>]
- Howie, B.N., Donnelly, P., & Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genetics*, 5(6), e1000529.
- Liang, K.-Y., & Zeger, S.L. 1986. Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika*, 73, 13–22.
- Miyajima, F., Ollier, W., Mayes, A., Jackson, A., Thacker, N., Rabbitt, P., et al. (2008a). Brain-derived neurotrophic factor polymorphism Val66Met influences cognitive abilities in the elderly. *Genes, Brain, and Behavior*, 7, 411–417.
- Miyajima, F., Quinn, J. P., Horan, M., Pickles, A., Ollier, W.E., Pendleton, N., et al. (2008b). Additive effect of BDNF and REST polymorphisms is associated with improved general cognitive ability. *Genes, Brain, and Behavior*, 7, 714–719.
- Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. et al. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, 38(8), 904–909.
- Wisdom, N. M., Callahan, J. L., & Hawkins, K. A. (2009). The effects of apolipoprotein E on non-impaired cognitive functioning: A meta-analysis. *Neurobiology of Aging*, 32, 63–74.