Provisional Paper Title: Do genetics discovered in GWAS of educational attainment predict brain volume?

Proposing Author: M Elliott & DW Belsky

Author's Email: dbelsky@duke.edu, maxwell.elliott@duke.edu

Objective of the study: To test if genetics discovered in GWAS of educational attainment predict total brain volume and, if so, to test if this association might mediate a portion of the association between education-linked genetics and cognitive functioning.

Data analysis methods:

Genome-wide association studies of educational attainment have identified genetics that predict years of completed schooling and related phenotypes in datasets from different continents and historical epochs [1–7]. These genetics also predict cognitive functioning, including in children who have not yet entered school [6]. These findings suggest that education-linked genetics influence educational attainment partly through their influence on cognitive function. But how education-linked genetics contribute to cognitive functioning is not known. Bioinformatics analysis of GWAS results suggest that education-linked genetics are implicated in brain development [2]. One brain development phenotype linked with cognitive function is brain volume. People with larger brains tend to score higher on tests of intelligence [8–11]. And biometric (family-based) genetic analyses find that the genetics of brain volume and the genetics of intelligence share substantial overlap [12–14]. Collectively, these findings suggest the hypothesis that one pathway connecting education-linked genetics and intelligence is increased brain volume.

Samples. To test this hypothesis, we will analyze four datasets with measurements of education-linked genetics, total brain volume, and intelligence. Two of these datasets come from population-based studies, UK Biobank (UKB, N=5,701) and the Dunedin Longitudinal Study (Dunedin, N=422). Two of these datasets come from primarily college-student samples, the Genome Superstruct Project (GSP, N=1,183) and the Duke Neurogenetics Study (DNS, N=515).

Analysis samples will include only participants of European descent because this is the population in which education GWAS were conducted. (GWAS measure only a subset of common variation in the human genome. The method depends on patterns of correlation among spatially proximate genotypes, called linkage disequilibrium, to infer genotypes of unmeasured causal genetic variation from the subset of variation that is measured. Patterns of linkage disequilibrium vary between populations with different ancestry, e.g. Africans and

Europeans. As a result, the SNPs measured in GWAS may reflect different unmeasured genotypes in different populations. Thus, GWAS findings for education made in European samples may not provide valid information in other populations [15].)

Genetics. We will measure education-linked genetics using the polygenic score method [16]. Polygenic scores are summaries of genome-wide genetic influence on a phenotype derived from results of GWAS. To calculate a person's polygenic score, each single-nucleotide polymorphism (SNP) is assigned a weight based on the effect-size estimated in the GWAS. Then, the weighted count of phenotype-associated alleles is averaged across the person's genome to calculate their polygenic score. We will calculate polygenic scores based on the most recent GWAS of education [2], available from the Social Science Genetic Association Consortium (https://www.thessgac.org/data). Following the methods used in our previous analyses, we will compute polygenic scores using all SNPs, i.e. we will not apply any p-value threshold for inclusion of a SNP in the score and we will not restrict the set of SNPs to be in linkage equilibrium [5,6,17,18]. Both simulation and empirical evidence find this approach produces maximally predictive scores [16,19–21].

To correct for ancestry differences within the European-descent UKB, Dunedin, GSP, and DNS samples, we will adjust polygenic score analyses for the first 10 principal components estimated from the individual genome-wide SNP datasets [17,22,23].

Total-brain-volume will be measured from high-resolution, T1-weighted MRI images. In the UK Biobank TBV (total gray matter plus total white matter) was estimated using SIENAX[24]. In DNS, GSP and Dunedin, images were processed using the Freesurfer recon-all image processing pipeline[25]. Specifically, the BrainSegNotVent value was used for each subject which represents the volume of all gray matter and white matter structures in the cortex and cerebellum, excluding the ventricles and brain stem.

Cognitive Function will be measured using a different test in each sample. In the UK biobank cognitive function was measured using 13 reason and logic puzzles [26]. In Dunedin, cognitive function was measured using the Wechsler Adult Intelligence Scales (WAIS) [27]. In GSP, cognitive function was measured using the Shipley Institute of Living Scale [28]. In DNS, cognitive function was measured using the Wechsler Abbreviated Scale of Intelligence (WASI) [29].

ANALYSIS

There are four analyses:

1) Test if total brain volume predicts intelligence
2) Test if education-linked genetics predict intelligence
3) Test if education-linked genetics predict total-brain-volume
4) Test if total-brain-volume mediates the association between education-linked genetics and intelligence.

In addition to these four analyses, we will conduct a sensitivity analysis to explore potential

heterogeneity between the population-based samples and the college-student samples. We will repeat our analysis in the subset of UK Biobank participants who had completed a college degree (N=2,808). This analysis will test if characteristics or experiences related to attending college might affect the relationships under study.

All models will include covariate adjustment for sex. Samples with participants of mixed chronological age will include covariate adjustment for age. As described above, all polygenic score analyses will be adjusted for the first ten principal components estimated from the genome-wide SNP dataset of the study being analyzed.

Results will be reported for individual samples. We will also meta-analyze results from all samples.

Variables needed:

Educational-attainment polygenic score
10 principal components estimated from genome-wide SNP data
Total Brain Volume
Cognitive Function
Age
Sex

Dunedin, DNS, and UK Biobank analyses will be performed at Duke by the Moffitt-Caspi and Hariri Labs. GSP analyses will be performed by the Holmes Lab.

Significance of the Study (for theory, research methods or clinical practice):

Genome-wide association studies of educational attainment have identified genetics that predict years of completed schooling and related phenotypes in datasets from different continents and historical epochs [1–7]. These genetics also predict cognitive functioning, including in children who have not yet entered school [6]. These findings suggest that education-linked genetics influence educational attainment partly through their influence on cognitive function. But how education-linked genetics contribute to cognitive functioning is not known. Bioinformatics analysis of GWAS results suggest that education-linked genetics are implicated in brain development [2]. One brain development phenotype linked with cognitive function is brain volume. People with larger brains tend to score higher on tests of intelligence [8–11]. And biometric (family-based) genetic analyses find that the genetics of brain volume and the genetics of intelligence share substantial overlap [12–14]. Collectively, these findings suggest the hypothesis that one pathway connecting education-linked genetics and intelligence is increased brain volume. If brain volume does mediate the association between education-linked genetics and cognitive functioning, this finding will advance understanding of mechanisms of genetic influence on educational attainment. If not, it will suggest need for more refined neural phenotypes for neurogenetic analysis.

References cited:

1. Rietveld, C. A. *et al.* GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* **340,** 1467–71 (2013).
2. Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533,** 539–542 (2016).
3. de Zeeuw, E. L. *et al.* Polygenic scores associated with educational attainment in adults predict educational achievement and ADHD symptoms in children. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet. Off. Publ. Int. Soc. Psychiatr. Genet.* **165B,** 510–520 (2014).
4. Beauchamp, J. P. Genetic evidence for natural selection in humans in the contemporary United States. *Proc. Natl. Acad. Sci.* **113,** 7774–7779 (2016).
5. Domingue, B. W., Belsky, D. W., Conley, D., Harris, K. M. & Boardman, J. D. Polygenic Influence on Educational Attainment. *AERA Open* **1,** 2332858415599972 (2015).
6. Belsky, D. W. *et al.* The Genetics of Success: How Single-Nucleotide Polymorphisms Associated With Educational Attainment Relate to Life-Course Development. *Psychol. Sci.* **27,** 957–972 (2016).
7. Selzam, S. *et al.* Predicting educational achievement from DNA. *Mol. Psychiatry* (2016). doi:10.1038/mp.2016.107
8. van Valen, L. Brain size and intelligence in man. *Am. J. Phys. Anthropol.* **40,** 417–423 (1974).
9. Pietschnig, J., Penke, L., Wicherts, J. M., Zeiler, M. & Voracek, M. Meta-analysis of associations between human brain volume and intelligence differences: How strong are they and what do they mean? *Neurosci. Biobehav. Rev.* **57,** 411–432 (2015).

10. Haier, R. J., Jung, R. E., Yeo, R. A., Head, K. & Alkire, M. T. Structural brain variation and general intelligence. *NeuroImage* **23,** 425–433 (2004).
11. McDaniel, M. A. Big-brained people are smarter: A meta-analysis of the relationship between in vivo brain volume and intelligence. *Intelligence* **33,** 337–346 (2005).
12. Posthuma, D. *et al.* Genetic Correlations Between Brain Volumes and the WAIS-III Dimensions of Verbal Comprehension, Working Memory, Perceptual Organization, and Processing Speed. *Twin Res. Hum. Genet.* **6,** 131–139 (2003).
13. Toga, A. W. & Thompson, P. M. Genetics of Brain Structure and InTELligence. *Annu. Rev. Neurosci.* **28,** 1–23 (2005).
14. Deary, I. J., Penke, L. & Johnson, W. The neuroscience of human intelligence differences. *Nat. Rev. Neurosci.* **11,** (2010).
15. Martin, A. R. *et al.* Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *Am. J. Hum. Genet.* **100,** 635–649 (2017).
16. Dudbridge, F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet.* **9,** e1003348 (2013).
17. Conley, D. *et al.* Assortative mating and differential fertility by phenotype and genotype across the 20th century. *Proc. Natl. Acad. Sci.* **113,** 6647–6652 (2016).
18. Domingue, B. W., Liu, H., Okbay, A. & Belsky, D. W. Genetic Heterogeneity in Depressive Symptoms Following the Death of a Spouse: Polygenic Score Analysis of the U.S. Health and Retirement Study. *Am. J. Psychiatry* **174,** 963–970 (2017).
19. Wray, N. R., Goddard, M. E. & Visscher, P. M. Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res.* **17,** 1520–1528 (2007).

20. Dudbridge, F. & Newcombe, P. J. Accuracy of Gene Scores when Pruning Markers by Linkage Disequilibrium. *Hum. Hered.* **80,** 178–86 (2016).
21. Ware, E. B. *et al.* Heterogeneity in polygenic scores for common human traits. *bioRxiv* 106062 (2017). doi:10.1101/106062
22. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38,** 904–909 (2006).
23. Price, A. L., Zaitlen, N. A., Reich, D. & Patterson, N. New approaches to population stratification in genome-wide association studies. *Nat. Rev. Genet.* **11,** 459–463 (2010).
24. Smith, S. M. *et al.* Accurate, Robust, and Automated Longitudinal and Cross-Sectional Brain Change Analysis. *NeuroImage* **17,** 479–489 (2002).
25. Fischl, B. FreeSurfer. *NeuroImage* **62,** 774–781 (2012).
26. Lyall, D. M. *et al.* Cognitive test scores in UK biobank: Data reduction in 480,416 participants and longitudinal stability in 20,346 participants. *PLoS ONE* **11,** (2016).
27. Wechsler, D. *WAIS-III administration and scoring manual*. *The Psychological Corporation, San Antonio, TX* (1997). doi:10.1177/1073191102009001003
28. Zachary, R. A. Shipley Institute of Living Scale: Revised Manual. *Los Angel. East. Psychol. Serv.* (1986).
29. Wechsler, D. WASI -II: Wechsler abbreviated scale of intelligence - second edition. *J. Psychoeduc. Assess.* **31,** 337–41 (2013).