The Case for a Social Science Genetic Association Consortium

Daniel J. Benjamin

Economics Department
Cornell University

2nd SSGAC Workshop • 29 October 2011

Heritability of Social Science Outcomes

- Socioeconomic Outcomes
  - Educational attainment: ~50% (Behrman et al., 1975; Miller et al., 2001; Scarr and Weinberg, 1994; Lichtenstein et al., 1992)
  - Income: ~40% (Björklund, Jäntti and Solon, 2005; Sacerdote, 2007; Taubman, 1976)
- Economic Preferences
  - Risk preferences: ~30% (Cesarini et al., 2009; Zhong et al. 2009; Zyphur et al. 2009)
  - Bargaining behavior, altruism and trust: ~30% (Wallace et al., 2007; Cesarini et al., 2008)
- Economic Behaviors
  - Financial decision-making: ~30% (Barnea et al., 2010; Cesarini et al, 2010)
  - Susceptibility to decision-making anomalies: ~25% (Cesarini et al., 2011)

Comparison to Other Traits

- Compared to other traits (e.g., height, personality), the heritabilities of economic phenotypes are lower, often ~30-40%.

- These differences are diminished when measurement error/transitory variance is accounted for.
  - MZ correlation in income rises to 0.55 when smoothing out transitory fluctuations by taking a 20 year average (Benjamin et al. 2011).
  - MZ correlation in a measure of risk aversion rises to 0.70 when adjusting for low reliability (Beauchamp et al., 2011).

Some Payoffs from “Genoeconomics”

1. Direct measures of previously latent parameters
   - Abilities and preferences are latent in most econ models.

2. Biological mechanisms for social behavior
   - Could decompose crude concepts like “risk aversion” and “patience.”

3. Genes as instrumental variables

4. Prediction using genetic information
   - Would facilitate targeting social-science interventions.
     - E.g., children with dyslexia-susceptibility genotypes could be taught to read differently from an early age.
   - Parents could expose children at a young age to activities that appeal to child’s preferences and abilities.

Challenge #1: Phenotype selection

- Want high-reliability phenotypes, consistently measured across many datasets.
  - E.g., height, *g*, years of education.

- Want proximate biological pathway for effect.
  - If pathway too distal, effect will likely be small, so low power.
  - If different pathways in different local environments, few datasets available to replicate.
  - Proximate pathway more likely for phenotypes shared with animal models.
  - E.g., aggression? Risk aversion? Impulsivity?

Challenge #2: Causal inference

- Confounds, e.g.:
  - Ethnicity
  - Gene-environment correlation
  - Gene-gene correlation
- Need convergent evidence from:
  - Large family samples
  - Modeling and estimation of environmental effects
  - Knock-out experiments with animal models
  - Biological evidence on protein products of genes
- Will take a long time to accumulate evidence.

Challenge #3: Statistical power

- Low power is due to small effect sizes.
  - *COMT* has R^2 = .1% for cognitive ability.
  - Largest height association is R^2 = .3%.
- Low power exacerbated by:
  - Multiple hypothesis testing + publication bias.
  - Inconsistent or low-reliability phenotypes.
  - Search for GxE or GxG interaction.
- Evidence for low power:
  - Many published associations not reproducible.

Calibration: Power Analysis

- Two alleles: High and Low.
- Equal frequency of High and Low.
- Phenotype distributed normally.
- Either there is a true association or not.
- If associated, R^2 = .1% (large for behavior).

- Sample size for 80% power: 7,845.

- Now suppose significant association at α = .05.

Posterior probability of a true association

Conditional on significant at $\alpha = .05$…

		Sample size		
		$N = 100$ (power = .06)	$N = 5{,}000$ (power = .61)	$N = 30{,}000$ (power = .99)
Prior	.01%	.01%	.12%	.20%
prob-	1%	1%	11%	17%
ability	10%	12%	58%	69%

Calculated by Bayes' Rule: $$P(true \mid significant) = \frac{power \cdot prior}{power \cdot prior + 0.05(1 - prior)}$$

Posterior probability of a true association

Conditional on significant at $\alpha = 5 \times 10^{-8}$…

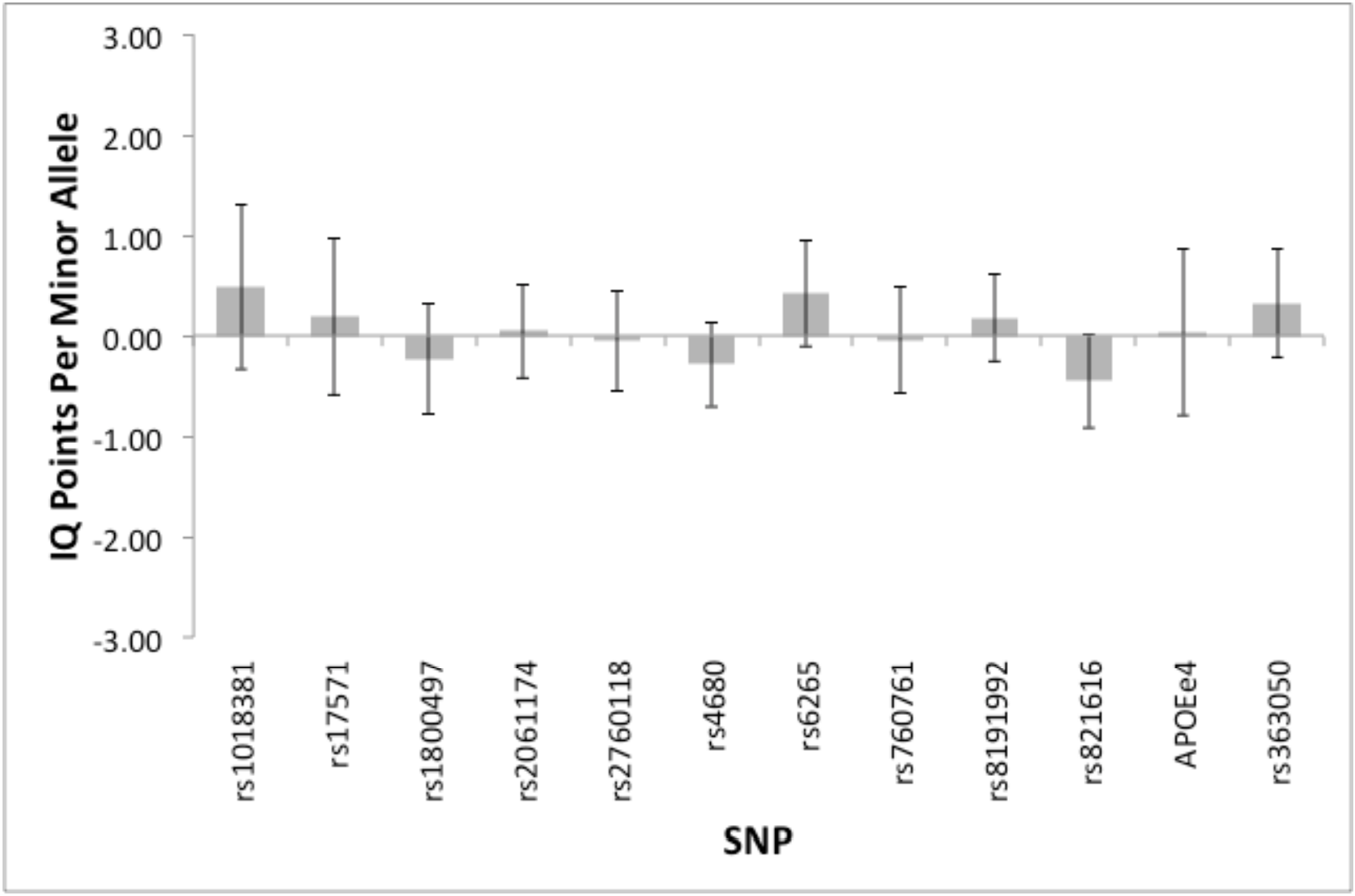
		Sample size		
		N = 100 (power = .00)	N = 5,000 (power = .00)	N = 30,000 (power = .51)
Prior prob-ability	.01%	.03%	57%	100%
	1%	3%	99%	100%
	10%	25%	100%	100%

Calculated by Bayes' Rule: $$P(true \mid significant) = \frac{power \cdot prior}{power \cdot prior + 0.05(1 - prior)}$$

My own experiences

- We could not replicate a promising candidate gene result (Benjamin et al., 2011).
  - Using AGES-RS data, we found a an association between educational attainment and a candidate gene previously associated with brain development (*N* = 2,349).
  - The association was found to be mediated by cognitive function.
  - The result survived a replication attempt from a non-overlapping sample from AGES-RS that had been genotyped subsequently (*N* = 1,759).
  - The association completely failed to replicate in the Framingham Heart Study (*N* = 7,357), the Wisconsin Longitudinal Study (*N* = 3,408), and the controls in the NIMH Swedish Schizophrenia Study (*N* = 1,235).
- Our experience is not unique.
  - Beauchamp et al (forthcoming) find 20 suggestive SNP associations in an education GWAS in Framingham (*N* = 7,574).
  - In replication attempt with Rotterdam Study (*N* = 9,535), ***none*** significant at .05 level, and only 9 of 20 had same sign.

Pooled estimates (11 SNPs + APOE)



Concluding Thoughts

- Why pursue molecular genetics in the social sciences?
  - It may be transformative for the social sciences.
  - Effects may be too small…but if so, better to find out sooner.
  - There is no way to know whether it will succeed without trying!
- In any event, it will be hot in the near future because there are major potential payoffs, and the data are there.
  - As genotyping costs plummet, GWAS data will be collected in many major social surveys.
- As we pursue it, it is urgent that we stop recapitulating the mistakes of medical genetics and set high standards.
- For this reason, we formed the SSGAC in Feb, 2011.
  - Proof-of-concept phenotype: Educational attainment.
  - This workshop: Discuss and launch GWAS of other phenotypes, harmonize collection of new phenotypes, and discuss other analyses SSGAC could help push forward.