An Intellectual Journey from Candidate Gene Studies to GWAS

David Cesarini
Economics Department
New York University

Conference on Genetics and Behavior, University of Chicago, April 2014

Background

- We are at the beginning of an explosion in social-science genetics research.
- Two fundamental approaches to gene discovery:
  - Candidate gene studies (specify *ex ante* hypotheses about small set of SNPs)
  - GWAS (*atheoretica*l testing of a large number of SNPs)

Outline

- My talk: existing candidate approaches, as practiced to date, are not working.
- Dan's talk: there is a single principle that can help us understand why.
- Phil's talk: the findings from properly designed studies replicate consistently.

Candidate-Gene Study

- Set significance threshold α = .05 / *#Hypotheses*.
- Eminently reasonable, and has worked when hypotheses are direct. (e.g., *APOE* and Alzheimer's)
- But social-science results often fail to replicate.
  - **Population stratification.**
  - Uncorrected multiple hypothesis testing / publication bias.
  - Low power and weak hypotheses (in the small samples typically used).

Genome-Wide Association Studies (GWAS)

- Atheoretical testing of all SNPs measured on the chip (typically 0.5-2.5 million).
- Set significance threshold $\alpha = 5 \times 10^{-8}$ (since ≈1 million independent SNPs in genome).
- Some advantages of GWAS:
  - Hypothesis-free design makes need to correct for multiple hypothesis testing transparent.
  - Possible and easy to control for principal components of the GWAS data—helps deal with the major confound of population stratification.

Most Reported Genetic Associations With General Intelligence Are Probably False Positives

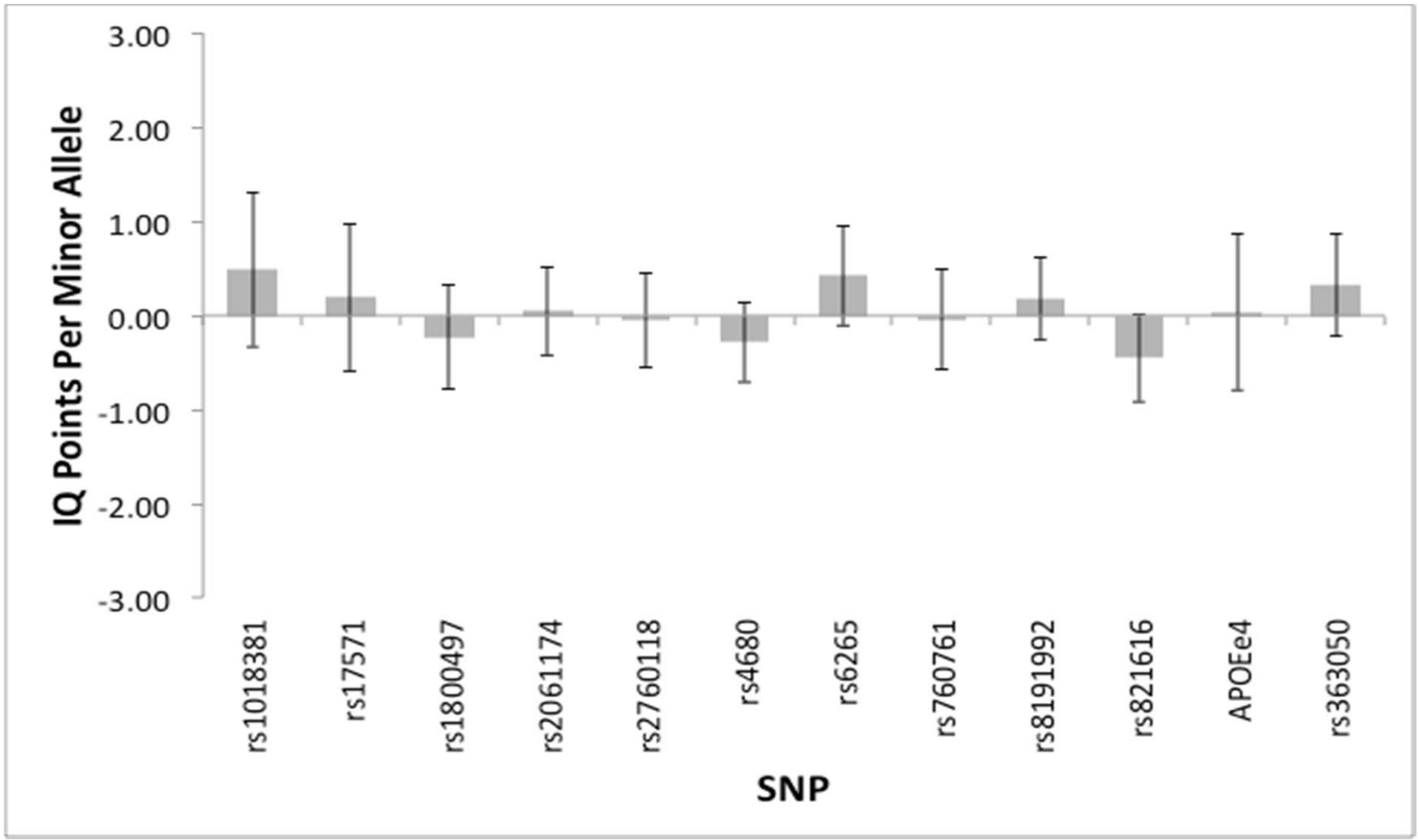
Psychological Science
23(11) 1314–1323
© The Author(s) 2012
Reprints and permissi
sagepub.com/journalsP
DOI: 10.1177/095679
http://pss.sagepub.com
SAGE

Christopher F. Chabris[1], Benjamin M. Hebert[2], Daniel J. Benjamin[3], Jonathan Beauchamp[2], David Cesarini[4], Matthijs van der Loos[5], Magnus Johannesson[6], Patrik K. E. Magnusson[7], Paul Lichtenstein[7], Craig S. Atwood[8], Jeremy Freese[9], Taissa S. Hauser[10], Robert M. Hauser[10], Nicholas Christakis[11,12], and David Laibson[2]

Methodology

- Selected the candidate SNPs from a review (Payton, 2008).
- Tried to replicate the associations between 11 SNPs + *APOE* with published *g* associations in three different samples with a combined sample size of 10,000.
- In none of the samples were we able to replicate any of the associations reported in the literature.
- We cannot reject the null hypothesis that the SNPs jointly have no explanatory power for *g*.

Pooled Estimates (11 SNPs + *APOE*)



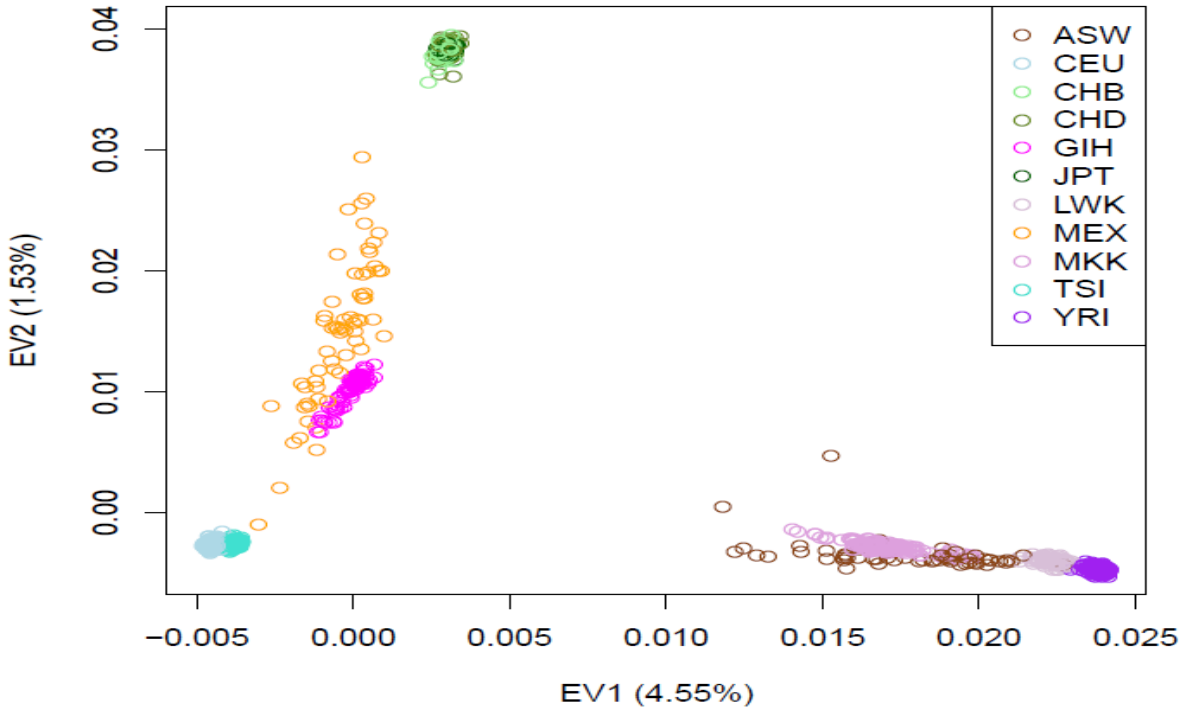
Editorial Policy on Candidate Gene Association and Candidate Gene-by-Environment Interaction Studies of Complex Traits

John K. Hewitt

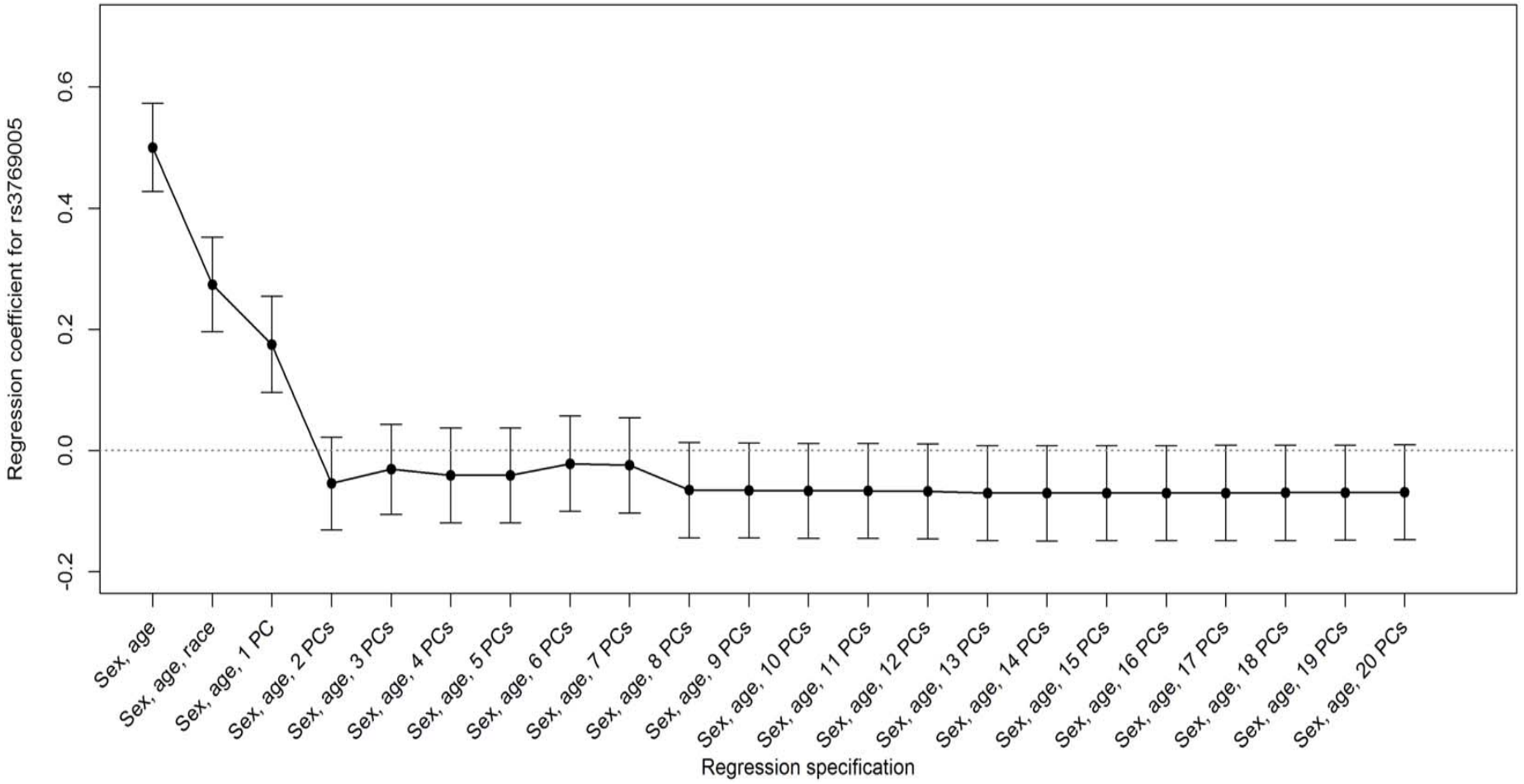
The literature on candidate gene associations is full of reports that have not stood up to rigorous replication. This is the case both for straightforward main effects and for candidate gene-by-environment interactions (Duncan and Keller 2011). As a result, the psychiatric and behavior genetics literature has become confusing and it now seems likely that many of the published findings of the last decade are wrong or misleading and have not contributed to real advances in knowledge. The reasons for this are complex, but include the likelihood that effect sizes of individual polymorphisms are small, that studies have therefore been underpowered, and that multiple hypotheses and methods of analysis have been explored; these conditions will result in an unacceptably high proportion of false findings (Ioannidis 2005).

Population Stratification

- Allele frequencies are correlated with unobserved environmental confounds.
- Principal component analysis used to model ancestry differences (Price *et al.* 2006)



A Spurious Association: *LCT* and Educational Attainment



Restricting Sample to “Whites” Insufficient