Session 1 Slides

Genetic Variation and Economic Behavior

David Cesarini

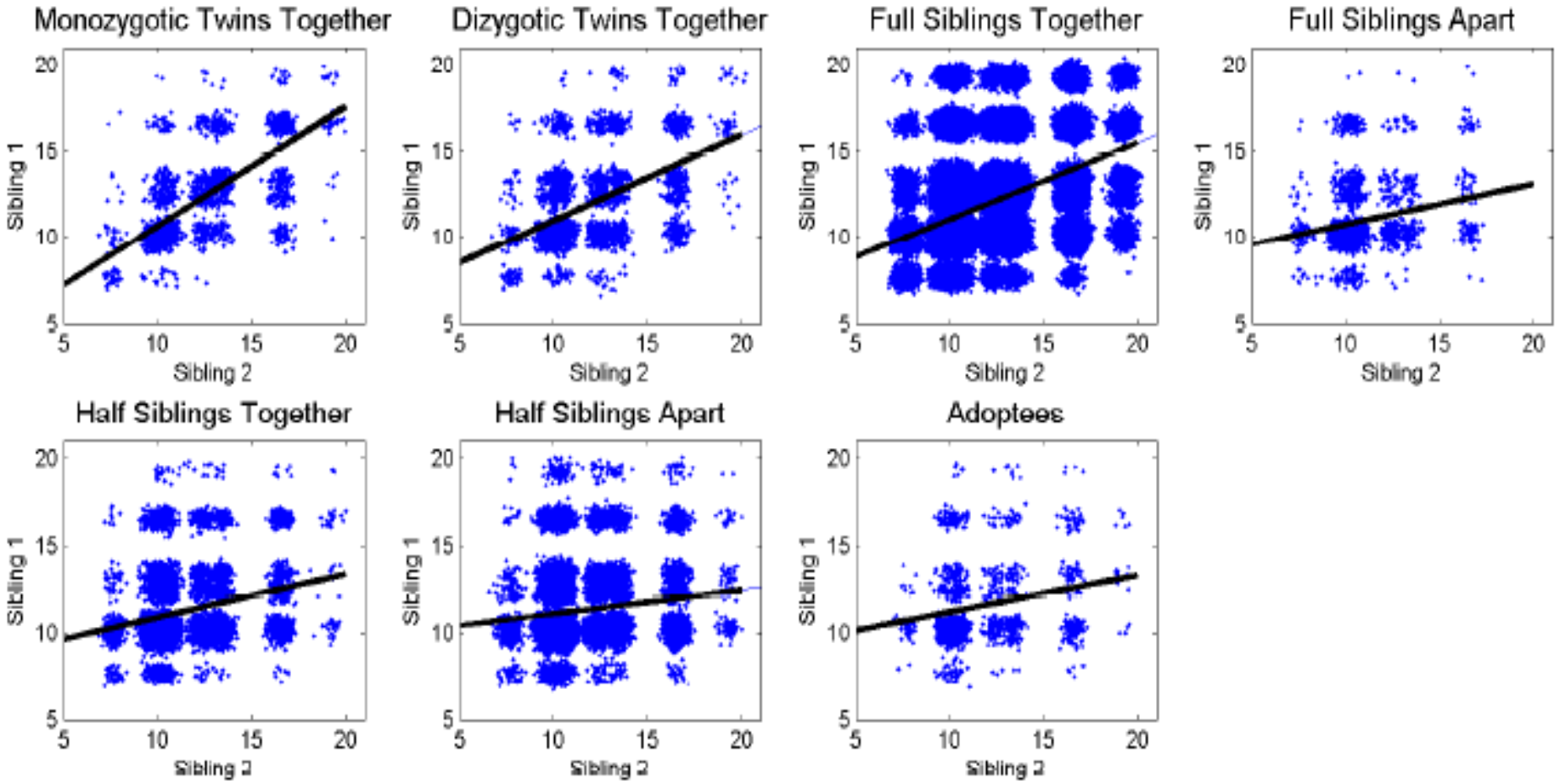
CESS, Economics Department
New York University

Workshop to Explore SSGAC • 12 February 2011

Heritability of Social Science Outcomes

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

An Example: Educational Attainment



Evidence from the SALTY survey

Table III. Polychoric Correlations for Sample

	MZ		DZ		OSDZ		Retest	
	ρ	s.e..	ρ	s.e.	ρ	s.e.	ρ	s.e.
Loss Aversion	0.34	0.05	0.18	0.06	0.04	0.06	0.56	0.07
Discounting	0.25	0.06	0.16	0.05	0.10	0.06	0.61	0.06
Procrastination	0.25	0.04	0.10	0.04	0.14	0.04	0.92	0.01
Risk General	0.35	0.02	0.15	0.02	0.07	0.02	0.64	0.04
Risk Financial	0.31	0.02	0.18	0.02	0.08	0.02	0.67	0.03
Risk HRS	0.44	0.03	0.23	0.04	0.09	0.04	0.62	0.04
Risk Large Gain	0.16	0.08	0.08	0.08	0.10	0.09	0.46	0.10
Risk Large Loss	0.17	0.05	0.04	0.05	0.09	0.05	0.49	0.06
Illusion of Control	0.26	0.05	0.12	0.05	0.12	0.06	0.51	0.07
Ambiguity Aversion	0.21	0.05	0.13	0.05	0.08	0.05	0.41	0.07
Fungibility	0.42	0.09	0.17	0.11	-	-	0.62	0.11
Conjunction Fallacy	0.25	0.06	0.05	0.07	-0.04	0.07	0.54	0.07
Base Rate Neglect	0.29	0.08	0.17	0.07	0.15	0.07	0.64	0.07
Insensitive to Sample Size	0.48	0.04	0.20	0.05	0.03	0.05	0.73	0.04
Fairness	0.31	0.02	0.19	0.02	0.03	0.02	0.57	0.04
Trust	0.35	0.02	0.18	0.02	0.14	0.02	0.63	0.04

Concluding Thoughts

- Variance decomposition subject to a number of important issues of interpretation
  - Environmental mediation of genetic effects (Dickens and Flynn, 2005; Jencks, 1979; Ridley, 2003)
- Suggests a need to understand a need to understand *why* genotype correlates with economic outcomes and behaviors
- Heritable variation in these complex traits likely explained by a heterogeneous collection of mechanisms
  - But many of the precursors of socioeconomic outcomes, for example risk preference, are measured with noise.

The Case for a Social Science Genetic Association Consortium

Daniel J. Benjamin

Economics Department
Cornell University

Workshop to Explore SSGAC • 12 February 2011

Collaborators for Results in the Talk

Craig Atwood (University of Wisconsin-Madison)
Jonathan Beauchamp (Harvard University)
Christopher F. Chabris (Union College)
Jeremy Freese (Northwestern University)
Edward L. Glaeser (Harvard University)
Vilmundur Guðnason (Icelandic Heart Association)
Tamara B. Harris (National Institute on Aging)
Robert M. Hauser (University of Wisconsin-Madison)
Taissa S. Hauser (University of Wisconsin-Madison)
Benjamin M. Hebert (Harvard University)
David I. Laibson (Harvard University)
Lenore J. Launer (National Institute on Aging)
Shaun Purcell (Massachusetts General Hospital, Broad Institute)
Albert Vernon Smith (Icelandic Heart Association)

We gratefully acknowledge NIA for financial support

Some Payoffs from "Genoeconomics"

1. Genes as instrumental variables

2. Understanding market and behavioral mediation of genetic effects
   - Genes are *measures* of (until-now latent) parameters of economic models: abilities and preferences.

3. Biological mechanisms for social behavior
   - Could decompose crude concepts like "risk aversion" and "patience."

4. Policy implications of genetic information
   - Effects of public release on, e.g., market prices and allocations of health insurance.
   - Do the benefits of private release (anticipatory behaviors, reduced uncertainty) outweigh the costs?
   - Targeting social-science interventions
     - E.g., children with dyslexia-susceptibility genotypes could be taught to read differently from an early age.

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 G x E or G x G 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

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

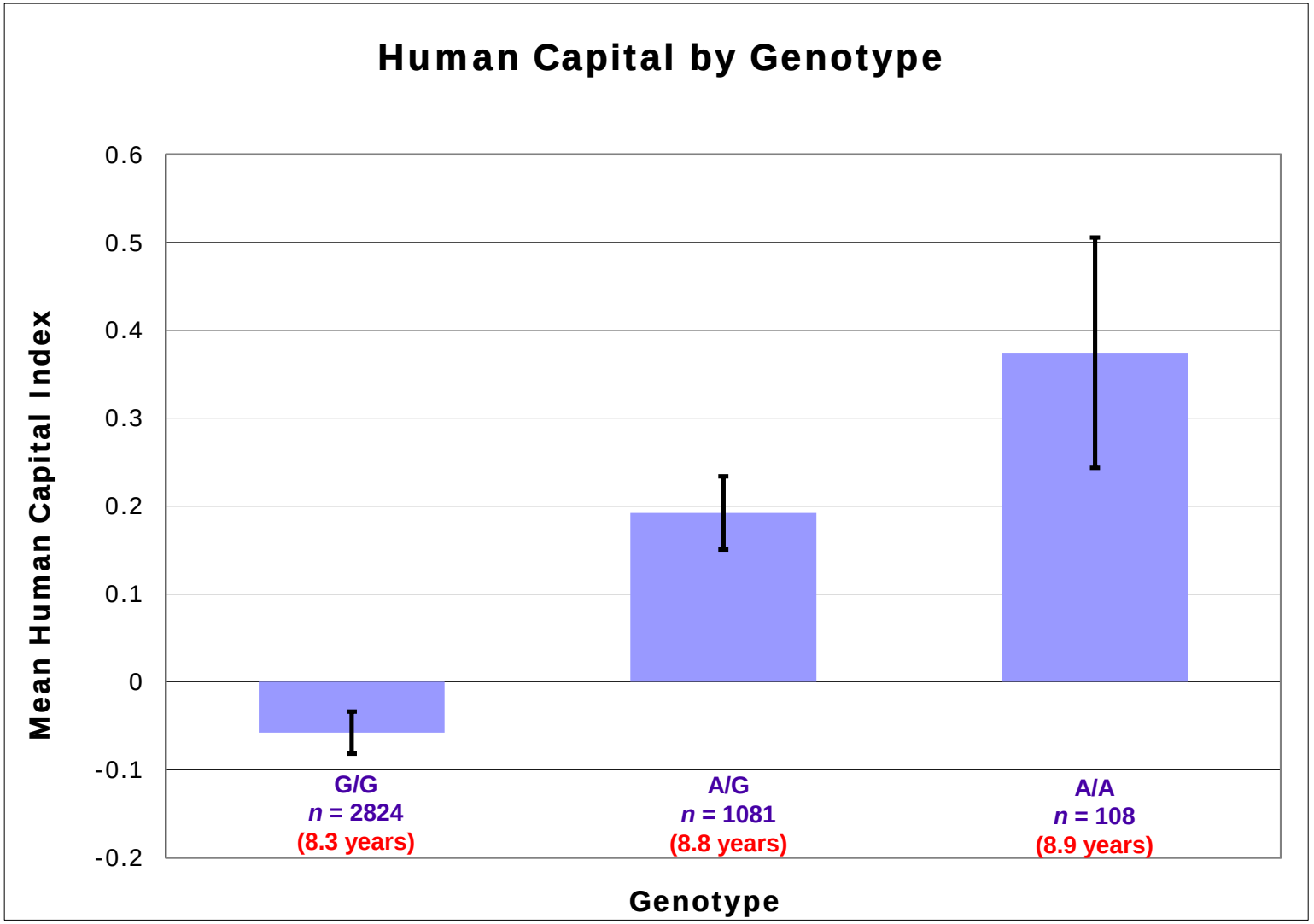
Calculated by Bayes' Rule: $$P(true \mid significant) = \frac{power \cdot prior}{power \cdot prior + 0.05(1 - prior)}$$

Case Study: My Experience

- We developed SNP panel and applied to large, ethnically homogeneous, well-characterized longitudinal dataset: AGES-Reykjavik Study.
- We conducted association analysis with 415 SNPs and 8 “economic” phenotypes. ($N \approx 2300$)

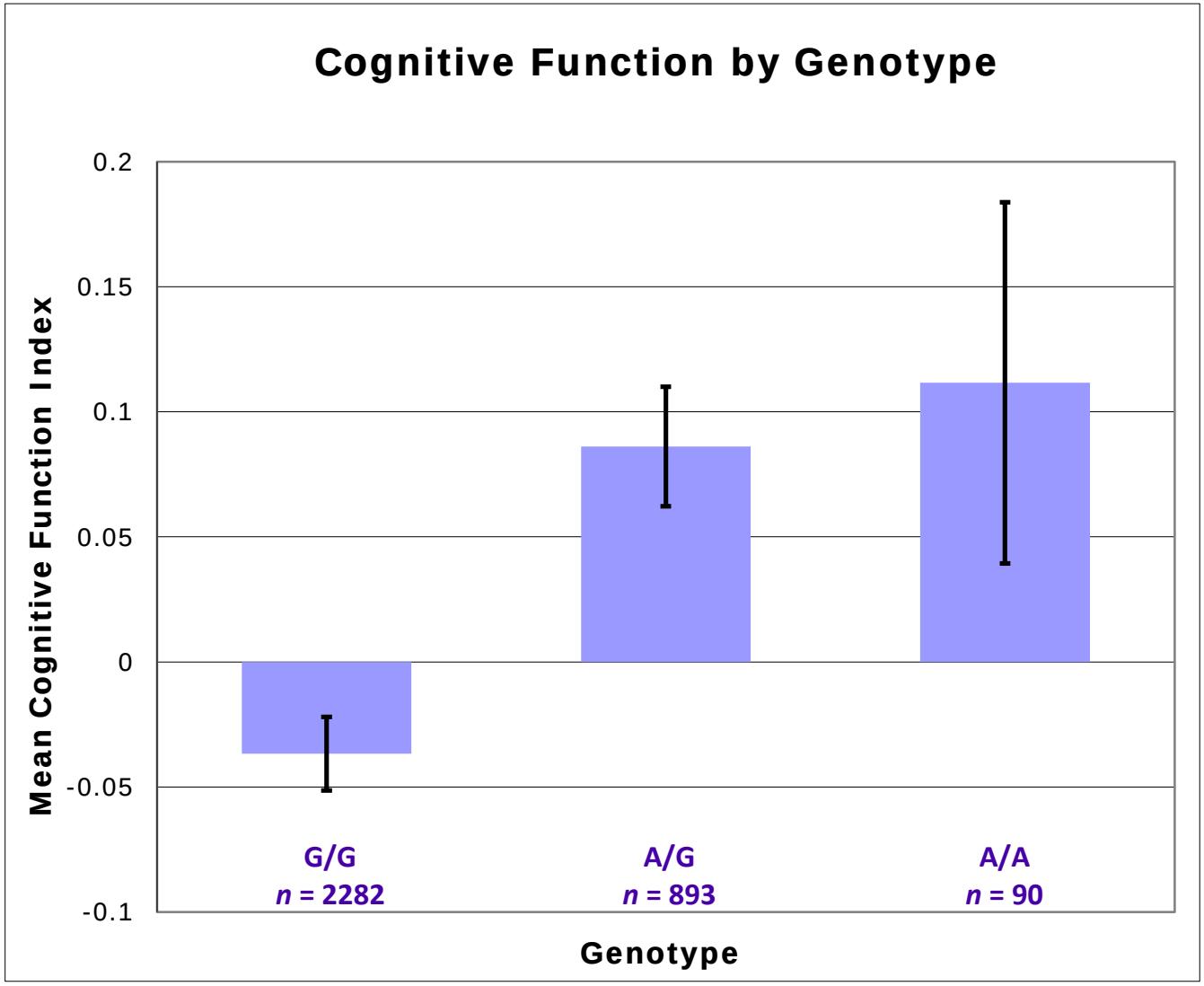
Case Study: My Experience

- We developed SNP panel and applied to large, ethnically homogeneous, well-characterized longitudinal dataset: AGES-Reykjavik Study.
- We conducted association analysis with 415 SNPs and 8 "economic" phenotypes. ($N \approx 2300$)
- We found 3 associations with .001 significance threshold.
- One replicated in a non-overlapping sample from the same dataset: **SSADH rs2267539 associated with "human capital"** (composed of years of schooling and number of languages learned). ($N \approx 1750$)


Human Capital by Genotype
Mean Human Capital Index
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
G/G
n = 2824
(8.3 years)
A/G
n = 1081
(8.8 years)
A/A
n = 108
(8.9 years)
Genotype

Case Study: My Experience

- We developed SNP panel and applied to large, ethnically homogeneous, well-characterized longitudinal dataset: AGES-Reykjavik Study.
- We conducted association analysis with 415 SNPs and 8 "economic" phenotypes. ($N \approx 2300$)
- We found 3 associations with .001 significance threshold.
- One replicated in a non-overlapping sample from the same dataset: **SSADH rs2267539 associated with "human capital"** (composed of years of schooling and number of languages learned). ($N \approx 1750$)
- We found the association was mediated by cognitive function.


Cognitive Function by Genotype
Mean Cognitive Function Index
0.2
0.15
0.1
0.05
0
-0.05
-0.1
G/G
n = 2282
A/G
n = 893
A/A
n = 90
Genotype

Case Study: My Experience

- We developed SNP panel and applied to large, ethnically homogeneous, well-characterized longitudinal dataset: AGES-Reykjavik Study.
- We conducted association analysis with 415 SNPs and 8 "economic" phenotypes. ($N \approx 2300$)
- We found 3 associations with .001 significance threshold.
- One replicated in a non-overlapping sample from the same dataset: **SSADH rs2267539 associated with "human capital"** (composed of years of schooling and number of languages learned). ($N \approx 1750$)
- We found the association was mediated by cognitive function.
- The association failed to replicate in 3 other samples.

Are we alone?

- We could not replicate a promising candidate gene result.
  - Even though the result survived initial replication attempts.
  - Even though there seemed to be a reasonable physiological story connecting the gene to the variable.
- Does the social science genetics literature contain many false positives?
  - Beauchamp et al (forthcoming) find 20 promising, biologically plausible SNPs 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.
- Candidate gene associations with social science variables seem to be especially vulnerable to being false positives.
  - Using WLS data, we could not replicate *any* of 13 SNPs with published *g* associations.
  - We had good power, positive controls (APOE4–parental AD).

Concluding Thoughts

- Why pursue molecular genetics in the social sciences?
  - While high-risk, 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.
- Consortium likely needed for adequate power.
  - Proof-of-concept phenotype: Educational attainment.
  - Can try to harmonize phenotypes and GWAS platform for subsequent analyses of other phenotypes.

Welcome

Philipp Koellinger

Assistant Professor Economics
Erasmus University Rotterdam

Workshop to Explore SSGAC • 12 February 2011

Our team in Rotterdam

- Prof. Patrick Groenen, Econometrics
- Prof. Albert Hofman, Epidemiology
- Dr. Philipp Koellinger, Economics
- Matthijs van der Loos, Economics
- Niels Rietveld, Economics
- Dr. Fernando Rivadeneira, Epidemiology and Internal Medicine
- Frank van Rooij, Epidemiology
- Prof. Roy Thurik, Economics
- Prof. André Uitterlinden, Internal Medicine
- Prof. Cornelia van Duijn, Epidemiology

Our initiative within CHARGE - 1

- CHARGE: Cohorts for Heart and Aging Research in Genomic Epidemiology
  - http://web.chargeconsortium.com/
  - since 2007
  - 75 publications
- Working groups within CHARGE
  - By phenotype
  - Coordinator(s)
  - Analysis plan
  - Every cohort analyses their own data
  - Meta-analysis by one or two teams (typically coordinator)
  - Writing group
  - Different cohorts across working groups

Our initiative within CHARGE - 2

- Educational Attainment as CHARGE working group
  - Infrastructure
    - CHARGE Wiki http://depts.washington.edu/chargeco/wiki/Main_Page
    - Telephone conferences
    - Bi-annual meetings
    - Expertise of steering committee and investigators
  - Well-working ‘code of conduct’
    - Data sharing
    - Publication plans
    - Authorship guidelines

Typical ‘code of conduct’

- Data sharing:
  - Upload descriptive statistics and GWAS results (not the primary data)
  - Once you upload, you are “in”
  - Collaboration agreement
  - Publication and presentation of meta-analysis results as a consortium
    - Once you are “in”, no side-shows and no surprises
- Authorship:
  - “The goal is fair scientific representation from cohort members participating in the WG”
    - First and senior authors (typically from different cohorts)
    - Number of authors reflect contribution of each cohort
    - Ordering reflects individual contribution

Our advisory board

- Dalton Conley
  - New York University, Sociology
- George Davey-Smith
  - University of Bristol, Epidemiology
- Albert Hofman
  - Erasmus University Rotterdam, Epidemiology
- Robert Krueger
  - University of Minnesota, Psychology
- David Laibson
  - Harvard University, Economics
- Peter Visscher
  - Queensland Institute of Medical Research, Statistical Genetics

Session 2 Slides

Experiences from a GWAS on entrepreneurship

Matthijs van der Loos

Erasmus Research Institute of Management (ERIM) and
Department of Applied Economics,
Erasmus School of Economics
Erasmus University Rotterdam

Saturday 12 February 2011

Introduction

Late 2007

Aim Discover genes associated with entrepreneurship

Motivation Mismatch between genetic predisposition and actual outcome may have adverse effects

Now 2011

- What happened in between?
- Where are our results?

Introduction

Late 2007

Aim Discover genes associated with entrepreneurship

Motivation Mismatch between genetic predisposition and actual outcome may have adverse effects

Now 2011

- What happened in between?
- Where are our results?

Some history

- Started in late 2007 with a GWAS of self-employment
- Initially using only data from the Rotterdam Study
- Different model specifications, sex-stratified analyses, different operationalisations
- Replication attempted in TwinsUK and NTR

The Gentrepreneur Consortium

- Goal: identify loci associated with self-employment through meta-analysis of GWAS using imputed SNP data
- Embedded within the CHARGE working group on entrepreneurship
- Concurrently recruited additional studies to increase power
- Combined sample dubbed the Gentrepreneur Consortium (Van der Loos et al. 2010, Eur J Epidemiol)
- This is a lot of (administrative) work!

Likely causes of null results

- Perhaps no genetic influence?
- Noise in phenotype definition
  - Current definition encompasses a very broad spectrum of entrepreneurial activities
  - Control group also not always clearly defined
- Gene-environment interactions are very likely to exist and will be missed by the current meta-analysis design
  - Spatiotemporal differences affect the entrepreneurial environment
  - For example, risk preferences in the US and Japan
- Still underpowered

How heritable is entrepreneurship?

- Twin studies suggest a heritability of $\sim$40% (Nicolaou et al. 2008, Manage Sci)
- New approach: use actual genotype data to estimate variance explained by common SNPs (Yang et al. 2010, Nat Genet)
- Applying this method to RS data suggests a heritability of $\sim$15%
- Twin studies overestimate heritability?
- Consequences for our GWAS efforts?

Future plans

- Focus on high-income entrepreneurs
- GWAS of endophenotypes, such as risk preferences or educational attainment

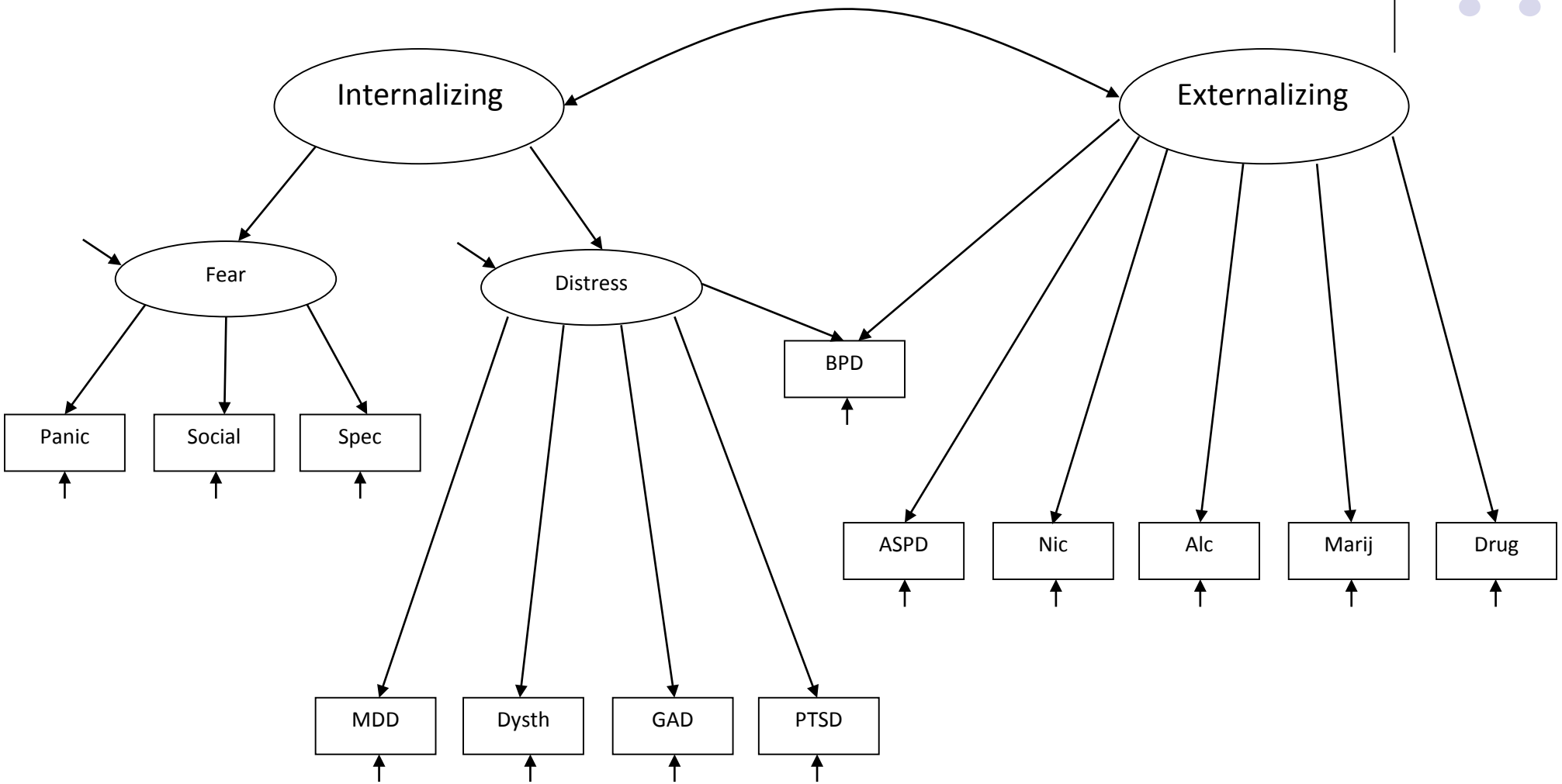
Session 3 Slides

Molecular genetic consortia : Some perspectives from a participant

Bob Krueger
Hathaway Distinguished Professor
University of Minnesota, USA

a dimensional-spectrum model of common forms of psychopathology

Eaton, Krueger, Keyes, Skodol, Markon, Grant, & Hasin, 2010, *Psychol Med*



why pursue molecular genetic inquiry focused on personality?

- personality is at the core of the psychopathology spectrums
  - dispositions function like diagnoses as indicators
    - genetically correlated with diagnoses in our twin research
  - this model is likely to frame major aspects of the DSM-5 meta-structure

- personality dispositions are therefore key variables in behavioral public health
  - understanding the etiology and neurobiology of these dispositions is important

- opportunities emerged for us to become involved in molecular genetic research on personality
  - but it became quickly apparent that progress would require large scale collaborations
    - akin to the vast majority of phenotypes

meta analytic GWAS of personality

Marleen H.M. de Moor 1†, Paul T. Costa 2, Antonio Terracciano 2, Robert F. Krueger 3, Eco J.C. de Geus 1, Tanaka Toshiko 2, Brenda W.J.H. Penninx 4,5,6, Tõnu Esko 7,8,9, Pamela A F Madden 10, Jaime Derringer 3, Najaf Amin 11, Gonneke Willemsen 1, Jouke-Jan Hottenga 1, Marijn A. Distel 1, Manuela Uda 12, Serena Sanna 12, Philip Spinhoven 5, Catharina A. Hartman 4, Patrick Sullivan 13, Anu Realo 14, Jüri Allik 14, Andrew C Heath 10, Michele L Pergadia 10, Arpana Agrawal 10, Peng Lin 10, Richard Grucza 10, Teresa Nutile 15, Marina Ciullo 15, Dan Rujescu 16, Ina Giegling 16, Bettina Konte 16, Elisabeth Widen 17, Diana L Cousminer 17, Johan G. Eriksson 18,19,20, 21,22, Aarno Palotie 17,23,24, 31, Leena Peltonen 17,23,24, 31 **, Michelle Luciano 25, Albert Tenesa 26, Gail Davies 25, Lorna M. Lopez 25, Narelle K. Hansell 27, Sarah E. Medland 27, Luigi Ferrucci 2, David Schlessinger 2, Grant W. Montgomery 27, Margaret J. Wright 27, Yurii S. Aulchenko 11, A.Cecile J.W. Janssens 11, Ben A. Oostra 28, Andres Metspalu 7,8,9, Gonçalo R. Abecasis 29, Ian J. Deary 25, Katri Räikkönen 30, Laura J. Bierut 10, Nicholas G. Martin 27, Cornelia M. van Duijn 11* , and Dorret I. Boomsma , in press, *Mol Psychiatry*

- 17,375 unrelated individuals of European ancestry from Europe, the United States and Australia
- 10 contributing studies
- genotyping platforms rendered commensurate via imputation
  - ~2.5M common SNPs included in HapMap, using the HapMap phase II CEU data as the reference sample
  - ~2,500,000 data points per person
- phenotypes are NEO-FFI (Five Factor Model; FFM) scales
  - Openness ,Conscientiousness, Extraversion, Agreeableness, Neuroticism
  - OCEAN
- Most promising hit: Conscientiousness associated with KATNAL2 gene on 18q21.1 (SNP rs2576037, $P=4.9 \times 10^{-8}$)

Steps taken to build the consortium

- Participation was actually word of mouth
  - We became involved because we saw an abstract for an upcoming meeting and we knew the authors
  - A preferable strategy is to scour the literature for potentially relevant studies
  - This seems especially important for major social science phenotypes
- Creation of a standard operating procedure document

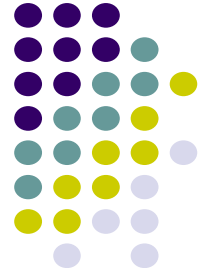
Standard Operating Procedure

- *"The problem in the world today is communication -- too much communication" (H. Simpson)*
  - (communication was **the** critical element)
- An identified leader and timeline
  - Tried to stick to the timeline
- Inclusion criteria
  - Genomewide SNP genotyping
  - EA ethnicity
  - Exact phenotype (NEO-FFI)
    - Currently under expansion

Standard Operating Procedure

- Precise phenotype definition
  - Via uniform scoring scripts
- Precise covariate selection and regression equation
  - Analyses were performed on each sample , then combined , by analysts affiliated with that sample
  - E.g. Neuroticism = b0 + b1 *coded_allele_dose + b2 * sex + b3 * age
  - Meta analysis performed by the organizer
- Expectation that the relevant genetics expertise would exist for each contributing study
  - Pre-imputation QC (e.g., MAF > 1%) and imputation at each site
  - QC was mostly assumed to be standardized already
  - Imputation on HapMap CEU via IMPUTE or MACH
- Sftp server

Uniform datafile format



Column Data

(1) SNP rs-name (if not known than report Affymetrix SNP ID)
(2) Chromosome
(3) Position (Build 35 of 36 depending on HapMap used for imputation)
(4) Coded allele, for which the linear regression effect is reported (A/T/G/C).
(5) Non-coded allele (A/T/G/C).
(6) Beta, the regression coefficient indicating change in personality score per coded allele
(7) Standard error of the effect specified in column 6
(8) P-value, two-sided p-value for the test that Beta=0.
(9) Allele Frequency, of the coded allele specified in column 4
(10) Minor Allele Frequency
(11) HWE p-value
(12) Imputation (whether a SNP was observed or imputed)
Preferably a string variable with values “observed” and “imputed”
(13) Imputation quality for imputed SNPs, set to 1 if the SNP was directly genotyped. R-squared if MACH was used and proper_info if IMPUTE was used.
(14) Effective sample size (number of individuals with genotype (imputed or direct) and phenotype data. Note that this can differ per SNP and per phenotype)

Potential challenges

- Communication
  - One central benign despot was very helpful
  - Requires requisite resources and commitment
- Location of genetics expertise
  - Centralized or study-by-study
- Harmonization
  - Phenotypic
  - Analytic
- Procedural issues beyond the first phenotype
  - A group to consider and coordinate additional analyses and phenotypes

Resources for Biosocial Surveys from DBASSE

Robert M. Hauser
Interim Executive Director
Division of Behavioral and Social Sciences
And Education

Reports *mainly* with support from the
National Institute on Aging

THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine
National Academy of Sciences
National Academy of Engineering
Institute of Medicine
National Research Council

Panel Reports

- Cells to Surveys (2000)
- Biosocial Surveys (2007)
- Genes, Behaviors, and the Social Environment: Moving Beyond the Nature/Nurture Debate (2006)
- Conducting Biosocial Surveys: Collecting, Storing, Accessing, and Protecting Biospecimens and Biodata (2010)
- HRS-GWAS Workshop Summary (forthcoming)

).

How to Get Them

- PDF or hard copy
- National Academies Press
- www.nap.edu

Can Genetics Learn from Social Science?

- **Data pooling vs. meta-analysis**
  - **Tension between**
    - **Data sharing**
    - **Protection**
  - **Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age (2009)**
- **Population heterogeneity**
- **Data harmonization**

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

www.nas.edu

http://www7.nationalacademies.org/dbasse/

Session 4 Slides

Lessons from a GWAS of Educational Attainment (Beauchamp et al., forthcoming)

David Cesarini

CESS, Economics Department
New York University

Workshop to Explore SSGAC • 12 February 2011

Framingham Background

- Of the 14,428 participants, 9,237 have been genotyped.
- Genotyping conducted using the Affymetrix 500k array (Affymetric, 2008)
- Years of education constructed using survey respones.
- Final sample with genetic, educational & demographic data: N=8,496.

Rotterdam Study: Replication Sample

- Prospective, population-based study from the Ommoord district.
- Genotyping was done with the Illumina 550K and 610K arrays
- 9,535 individuals have complete genotypic, education and basic demographic data

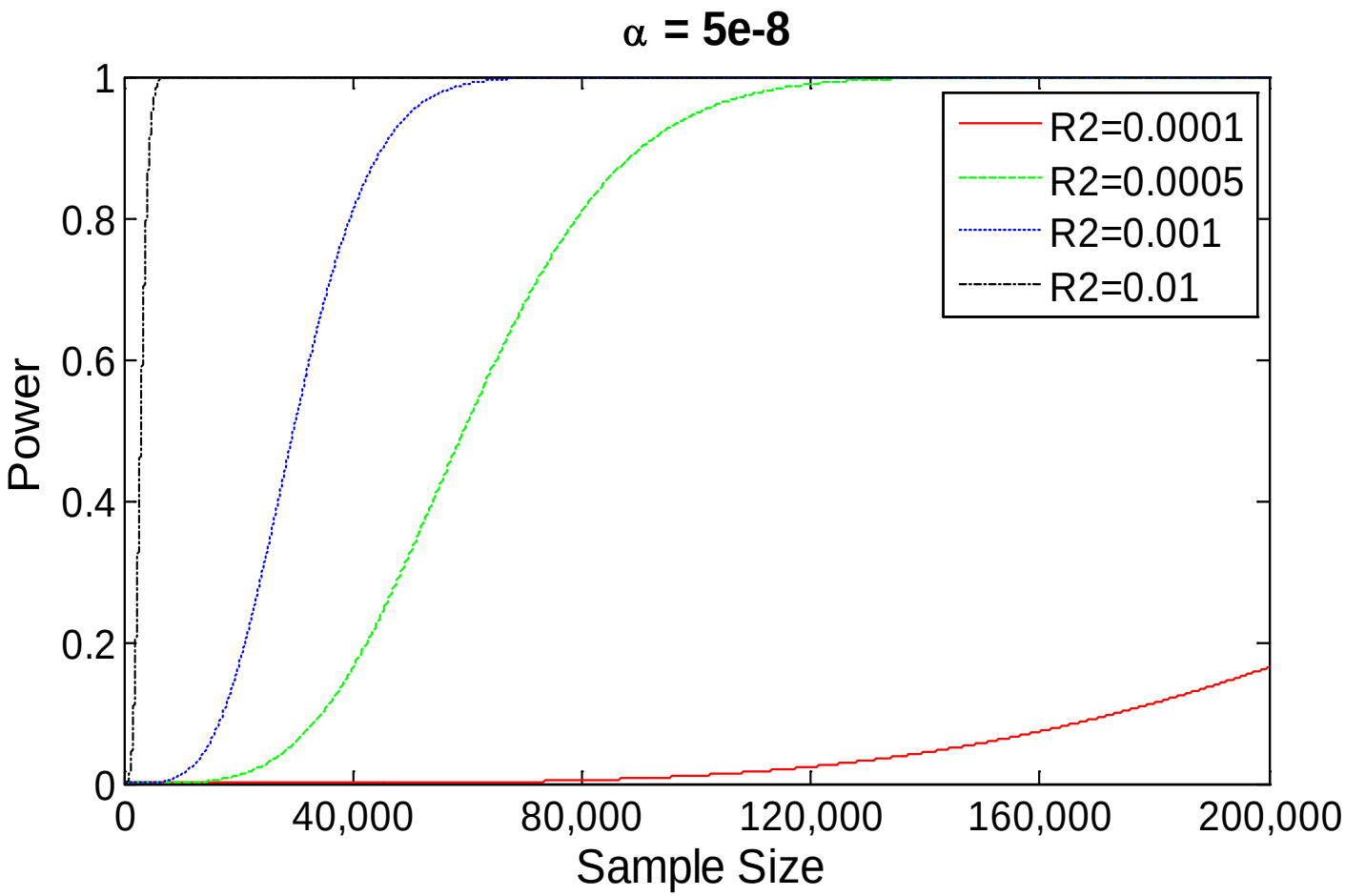
Results

SNP (Chromosome)	$\hat{\beta}$	p-value	Bonf.	Sample	M.A.
rs11758688 (6)	-0.253	$2.97 \cdot 10^{-7}$	0.107	7572	T
rs12527415 (6)	-0.253	$3.03 \cdot 10^{-7}$	0.109	7570	T
rs17365411 (2)	0.260	$3.73 \cdot 10^{-7}$	0.134	7559	C
rs7655595 (4)	-0.266	$3.99 \cdot 10^{-7}$	0.144	7486	G
rs17350845 (1)	-0.291	$6.22 \cdot 10^{-7}$	0.224	7415	C
rs12691894 (2)	-0.246	$6.67 \cdot 10^{-7}$	0.240	7572	G
rs9646799 (2)	0.271	$7.41 \cdot 10^{-7}$	0.267	7478	T
rs11722767 (4)	-0.257	$7.77 \cdot 10^{-7}$	0.280	7574	C
rs10947091 (6)	-0.245	$9.03 \cdot 10^{-7}$	0.325	7574	T
rs6536456 (4)	0.230	$1.32 \cdot 10^{-6}$	0.474	7513	C

Possible Interpretations

- False positive due to multiple hypothesis testing.
- Population stratification.
- True treatment effect local to environmental circumstances in Framingham

Power Graphs



Educational attainment – preliminary analysis plan

Philipp Koellinger
Adriaan Hofman

Workshop to Explore SSGAC • 12 February 2011

We’ve got the power

- 20 studies so far
- 11 countries
- Over 170,000 genotyped observations

Phenotype harmonization

Two measures – two research strategies

1. Educational attainment (EA) according to ISCED classification
2. College degree

1. ISCED classification

Levels	Definition	US years of schooling
0	Pre-primary education	1
1	Primary education or first stage of basic education	7
2	Lower secondary or second stage of basic education	10
3	(Upper) secondary education	13
4	Post-secondary non-tertiary education	15
5	First stage of tertiary education (not leading directly to an advanced research qualification)	19
6	Second stage of tertiary education (leading to an advanced research qualification, e.g. a Ph.D.)	22

How your country fits into this scheme:
http://www.uis.unesco.org/ev.php?ID=7434_201&ID2=DO_TOPIC

1. ISCED classification – USA example

Programme number (prog.<ISCEDlevel>.<number within level>)	ISCED level	Programme destination (A/B/C)	Programme orientation (G/P/V)	National name of the programme	Main diplomas, credentials and certifications awarded	Theoretical starting age	Theoretical cumulative years of education at the end of the programme
1	3	4	5	10	13	15	17
Prog.0.2	0		G	Kindergarten	None	4-6	1
Prog.1.1	1		G	Primary education	None	5-7	7
Prog.2.1	2		G	Middle education (grades 7-9)	None	11-13	10
Prog.3.3	3		G	Secondary education (grades 10-12)	High School Diploma	14-17	13
Prog.4.1	4	C	V	Vocational Certificate (< 1 year)	Occupationally specific vocational certificate	18-30	13
Prog.4.2	4	C	V	Vocational Certificate (1-2 years)	Occupationally specific vocational certificate	18-30	15

2. College degree

- College = 1 if ISCED >= 5
- College = 0 if ISCED <= 4

Genotypes & imputation

- All autosomal SNPs imputed from HapMap Phase II CEU panel
  - *MACH* or *Impute*

Analysis

- Only individuals older than 30 years
- Gender-stratified and pooled models
- Controls:
  - Year of birth *changed*
  - Four principal components of genotypic data, associated with the four largest eigenvalues
  - Sex, Sex * Year of birth *changed*
- Linear regression for ISCED categories
- Logistic regression for college dummy
  - R, Plink, SNPtest, or Mach2QTL
- *No* genomic control

Timeline

- Analysis plan will be distributed next week
  - Please contact us, if you do not hear from us by next Friday
- Meta-analysis conducted by
  - Erasmus U Rotterdam (Niels Rietveld)
  - U Minnesota (Jaime Derringer)
  - QIMR (Sarah Medland) *new*

Session 5 Slides

The way ahead

Philipp Koellinger

Assistant Professor Economics
Erasmus University Rotterdam

Workshop to Explore SSGAC • 12 February 2011

Additional phenotypes

- Working group on additional phenotypes
  - Catalogue of what is currently feasible
  - Suggestions for new data collection
    - Representing interests, input and experiences from various disciplines
  - We are open to additional volunteers and any ideas you have
  - Starting now

Costs and feasibility

- “Phenotyping” vs. “genotyping”
  - Example for N ~ 5000
  - Genotyping costs ~300 EUR per individual
  - 10 additional multiple choice questions ~3 EUR per individual
  - Cost advantage of “phenotyping” 1 : 100
- New genotyped samples appearing constantly

Collection of new phenotypes

- Funding opportunities?
  - NSF
  - NIH / NIA (R21)
  - European Research Council
- If funding would be provided, would you be willing / able to collect additional phenotypes?

Integrating social science datasets

- Are likely to lack analysts and expertise to carry out GWAS themselves
  - We will try to help out
  - U Minnesota and Erasmus U Rotterdam
- Consistent measurement across countries
  - Cross-national equivalency files (PSID)
  - HRS standards

THANK YOU!