Schools & the Genetics of Educational Attainment: Evidence from Add Health

BEN DOMINGUE
What do we mean by “genetics”?

• Genome-wide data

<table>
<thead>
<tr>
<th>SNP 1</th>
<th>SNP 2</th>
<th>…</th>
<th>SNP 1,000,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>0</td>
<td>1</td>
<td>…</td>
</tr>
<tr>
<td>P2</td>
<td>1</td>
<td>0</td>
<td>…</td>
</tr>
<tr>
<td>P3</td>
<td>1</td>
<td>2</td>
<td>…</td>
</tr>
<tr>
<td></td>
<td>:</td>
<td>:</td>
<td></td>
</tr>
<tr>
<td>P1000</td>
<td>2</td>
<td>1</td>
<td>…</td>
</tr>
</tbody>
</table>

1,000 × 1,000,000 matrix; each cell ∈ \{0, 1, 2\}. 
What do we mean by “genetics”?

• Genome-wide data

<table>
<thead>
<tr>
<th></th>
<th>SNP 1</th>
<th>SNP 2</th>
<th>⋮</th>
<th>SNP 1,000,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>0</td>
<td>1</td>
<td>⋮</td>
<td>2</td>
</tr>
<tr>
<td>P2</td>
<td>1</td>
<td>0</td>
<td>⋮</td>
<td>0</td>
</tr>
<tr>
<td>P3</td>
<td>1</td>
<td>2</td>
<td>⋮</td>
<td>1</td>
</tr>
<tr>
<td>⋮</td>
<td>⋮</td>
<td>⋮</td>
<td>⋮</td>
<td>⋮</td>
</tr>
<tr>
<td>P1000</td>
<td>2</td>
<td>1</td>
<td>⋮</td>
<td>2</td>
</tr>
</tbody>
</table>

1,000 × 1,000,000 matrix; each cell ∈ {0, 1, 2}.

• How do we tie this large set of genetic variants to specific traits?
GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment

All authors with their affiliations appear at the end of this paper.
GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment

All authors with their affiliations appear at the end of this paper.

- Second generation educational attainment GWAS now available.
- Genome-wide association studies produce information about small associations between millions of variants and outcomes of interest.
  - Can we simplify?

Genome-wide association study identifies 74 loci associated with educational attainment


Affiliations | Contributions | Corresponding authors
---|---|---
Nature 533, 539–542 (26 May 2016) | doi:10.1038/nature17671
Received 24 June 2015 | Accepted 16 March 2016 | Published online 11 May 2016
Polygenic Score (PGS)

<table>
<thead>
<tr>
<th>SNP 1</th>
<th>SNP 2</th>
<th>...</th>
<th>SNP 1,000,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>P2</td>
<td>1</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>P3</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td>:</td>
<td>:</td>
<td>...</td>
</tr>
<tr>
<td>P1000</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

$1,000 \times 1,000,000$ matrix; each cell $\in \{0, 1, 2\}$.
Polygenic Score (PGS)

For technical reasons, we focus on non-Hispanic white respondents.
Key facts about this PGS

• Predicts out of sample.
• Predicts net of maternal education & maternal genetics.
• Predicts amongst siblings.

Replicability and Robustness of Genome-Wide-Association Studies for Behavioral Traits

Cornelius A. Rietveld1,2, Dalton Conley3, Nicholas Eriksson4, Tõnu Esko5, Sarah E. Medland6, Anna A. E. Vinkhuyzen7, Jian Yang7, Jason D. Boardman8,9, Christopher F. Chabris10, Christopher T. Dawes11, Benjamin W. Domingue8, David A. Hinds4, Magnus Johannesson12, Amy K. Kiefer4, David Laibson13, Patrik K. E. Magnusson14, Joanna L. Mountain4, Sven Oskarsson15, Olga Rostapshova13, Alexander Teumer16, Joyce Y. Tung4, Peter M. Visscher7,17, Daniel J. Benjamin18, David Cesaroni19,20, Philipp D. Koellinger1,2,21, and the Social Science Genetics Association Consortium

Is the Effect of Parental Education on Offspring Biased or Moderated by Genotype?

Dalton Conley, Benjamin W. Domingue, David Cesaroni, Christopher Dawes, Cornelius A. Rietveld, Jason D. Boardman

DOI: 10.15195/v2.a8

Polygenic Influence on Educational Attainment

New Evidence From the National Longitudinal Study of Adolescent to Adult Health
Benjamin W. Domingue, Daniel W. Belsky, Dalton Conley, Kathleen Mullan Harris, Jason D. Boardman
DOI: 10.1177/2332858415599972, Aug 2015
Polygenic Score

Years of Edu

Genetics
What is role of schools?

Questions
1. How are relevant genetics distributed across schools?
2. Any indication that schools are moderating genetic effects?
• Probability sample drawn from >100 schools.
• Subsample (siblings) first genotyped.
• Full cohort now genotyped.
  ○ 10,577 total
  ○ 6,524 non-Hispanic White
Clustering of PGS into schools

- ICC showing proportion of variation in outcomes that is between schools.
  - e.g., ~20% of the variation in maternal education is between schools.
Clustering of PGS into schools

[Bar chart showing clustering of PGS into schools with categories such as EA, AA, and HISP for variables like PGS, W4 Edu, W1 PVT, W4 IQ, Mom Edu, and GPA.]
Clustering of PGS into schools

• What is effect of the clustering we do observe?
  – Genetics of EA student’s peers are more predictive than their own.
How much variation is there in the PGS effect between schools?

- First computed correlations between educational attainment and PGS for each school.
- Then considered the distribution.
How much variation is there in the PGS effect between schools?
In closing

• Preliminary Findings:
  – It seems clear that there is interesting genetic clustering into schools.
  – Little evidence for school-level moderation of genetic effect.
In closing

• Preliminary Findings:
  – It seems clear that there is interesting genetic clustering into schools.
  – Little evidence for school-level moderation of genetic effect.

• Consistent with a model in which genetic risks are not evenly distributed but where there is little gene-environment interaction (GxE).
  – Key role for gene-environment correlation (rGE).

• Also of interest: upstream (developmental pathways) and downstream (later-life outcomes) associations with PGS.
Thanks!

bdomingue@stanford.edu

Collaborators: Dan Belsky, Kathleen Mullan Harris, Jason Boardman