CVFS Genetic Study Design and Applications

Colter Mitchell
CVFS Webinar Series
December 1, 2021

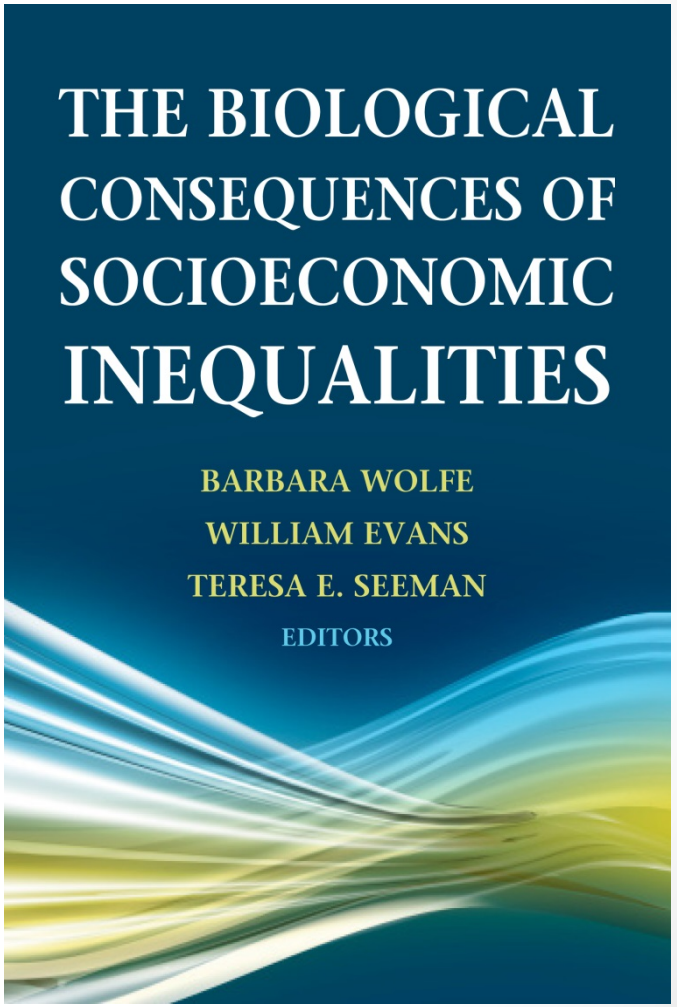
UNIVERSITY OF MICHIGAN

Outline

- Motivation
- Basic Genetics Refresher
- CVFS Study Design
- Applications
  - Polygenic Scores
  - GxE and rGE
  - Family Models

The Biological Consequences of Social Disadvantage

- Last 15 years has a seen a massive increase in biosocial work
- Initially this work focused on large systems: blood pressure, cholesterol, glucose levels, etc.
- Rapidly moving to the molecular level and brain(focus of this talk)
- Not all populations equally represented



Why would social scientists care about biology?

- **Mechanisms**-most social contextual effects have at least one biological mechanism to health/behavior
- **Indicators**-can we see a biomarker before the outcome?
- **Controls**-better measurement of the social contextual effects
- **Moderation**- differential response to the environment
- **Social Justice**-not all populations have been examined equally

Whose biological data do we have

What is a representative brain? Neuroscience meets population science

Emily B. Falk[a,b,c,1], Luke W. Hyde[d,e,f,1], Colter Mitchell[e,g,1,2], Jessica Faul[e,3], Richard Gonzalez[b,d,h,3], Mary M. Heitzeg[i,3], Daniel P. Keating[d,e,i,j,3], Kenneth M. Langa[e,k,l,3], Meghan E. Martz[d,3], Julie Maslowsky[m,3], Frederick J. Morrison[d,3], Douglas C. Noll[n,3], Megan E. Patrick[e,3], Fabian T. Pfeffer[e,g,3], Patricia A. Reuter-Lorenz[d,e,o,3], Moriah E. Thomason[p,q,r,3], Pamela Davis-Kean[b,d,e,f,4], Christopher S. Monk[d,e,f,i,o,4], and John Schulenberg[d,e,f,4]

PNAS | October 29, 2013 | vol. 110 | no. 44 | 17615–17622

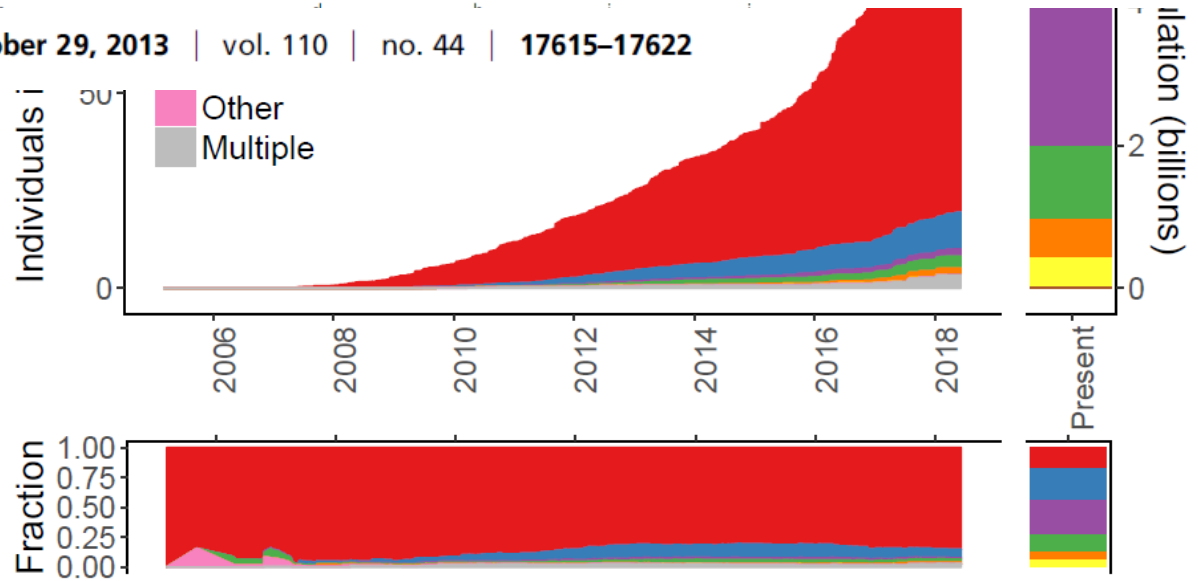
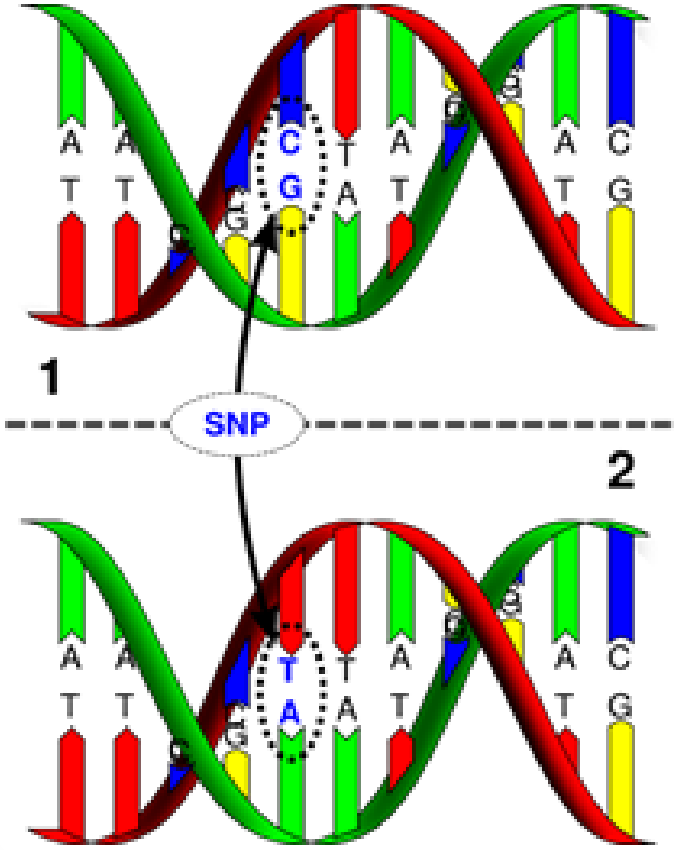


Figure 1 – Ancestry of GWAS participants over time compared to the global population. Cumulative data as reported by the GWAS catalog [23]. A notable caveat is

Falk et al 2013, Mitchell 2018, Martin et al 2019

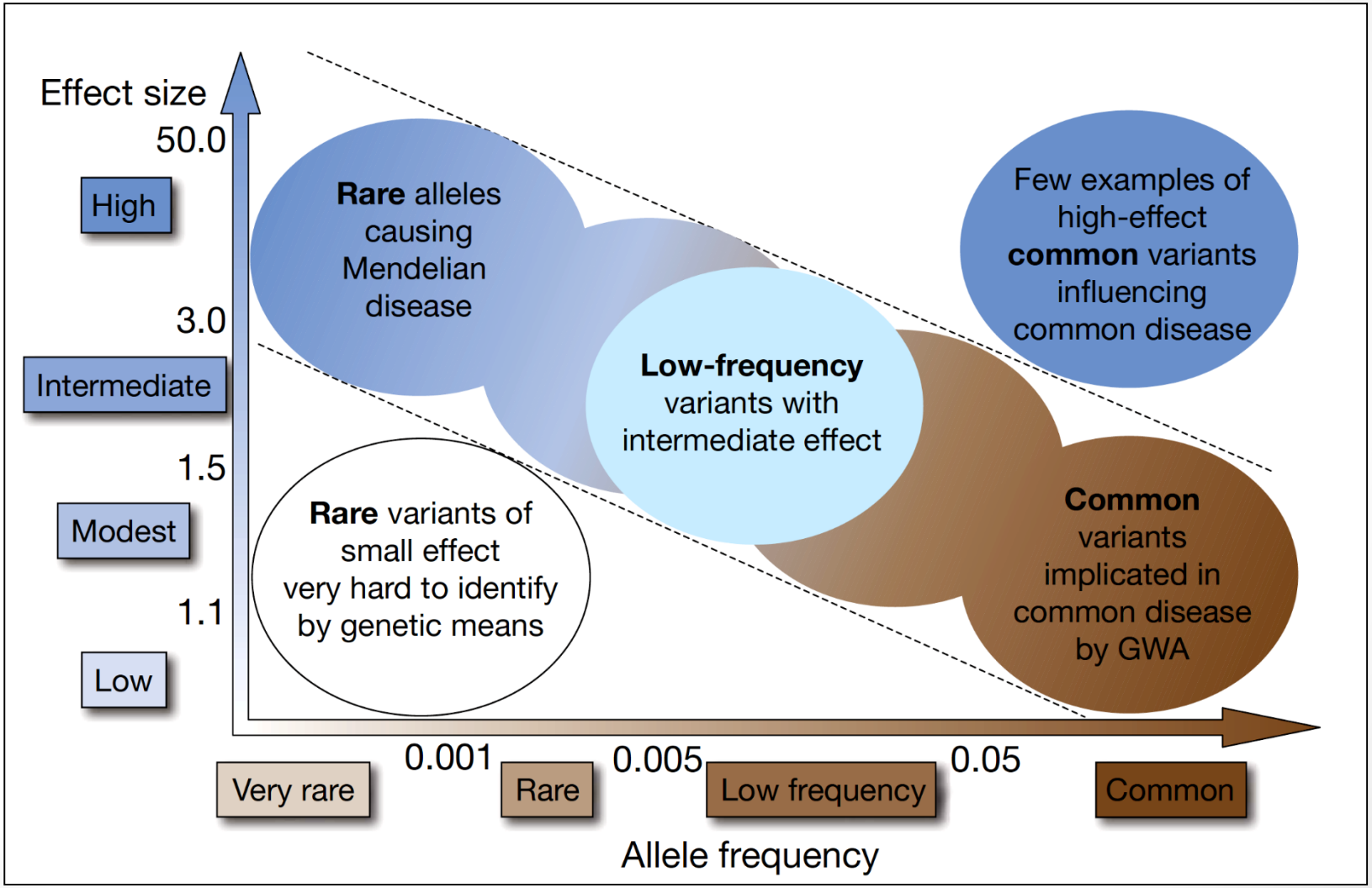
Genetics Refresher

Single Nucleotide Polymorphisms (SNPs)

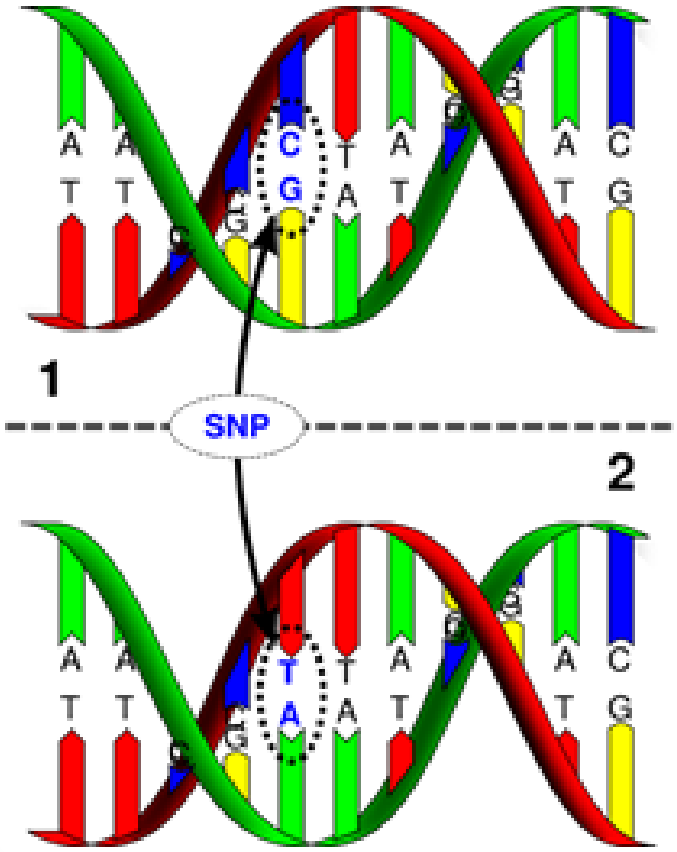


- Simplest form of genetic variation.
  - Others exist (e.g., CNVs)
- SNPs occur ~1/300 base pairs
- In >1% of the population.

GWAS generally identify common SNPs with modest effect sizes



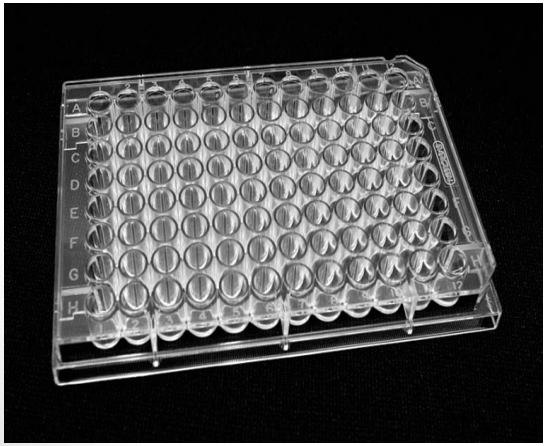
Single Nucleotide Polymorphisms (SNPs)



- Two flavors of a SNP are "alleles".
  - Major and minor alleles.
  - Reference Allelle.
- We inherit two pieces of genetic information, one from each parent.
- Focus is on number (0,1,2) of reference alleles at a SNP.

Genotyping

- Early work was very intensive (and still is for some genetic markers)
- Now more automated, but now data intensive
- Assay 100'sK to millions of SNPs



SNP data

Genome-wide data

	SNP 1	SNP 2	$\cdots$	SNP 1,000,000
P1	0	1	$\cdots$	2
P2	1	0	$\cdots$	0
P3	1	2	$\cdots$	1
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
P1000	2	1	$\cdots$	2

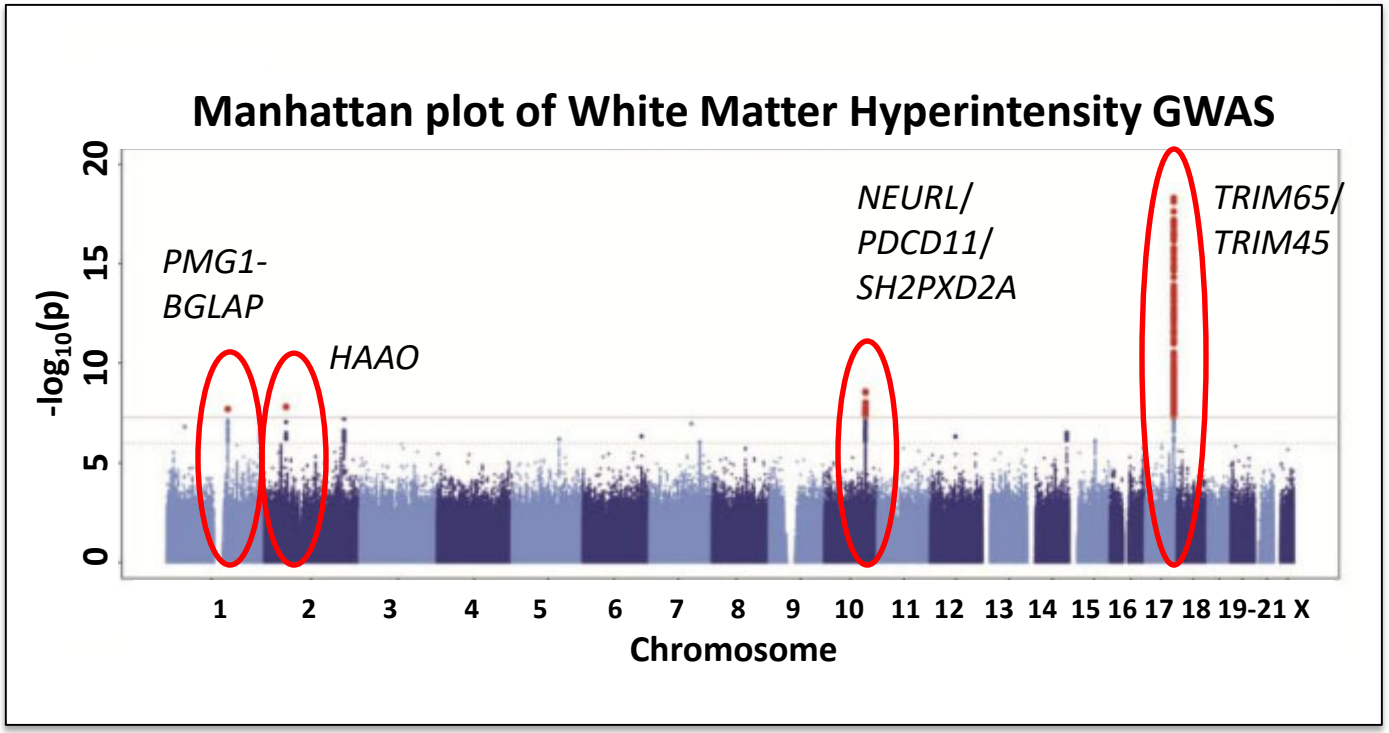
$1,000 \times 1,000,000$ matrix; each cell $\in \{0, 1, 2\}$.

What to do with such data?

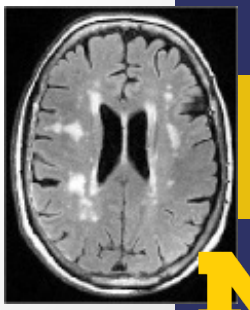
How a Genome-wide Analysis Typically Works

- Discovery sample
  - Simple OLS or Logistic Regression done millions of times
  - Find Genome-wide (p-value < $5x10^{-8}$) significant hit
  - Verify clustering or correlated markers
  - Often done in meta-analysis of many studies
- Replication Sample
  - Smaller n, but often still large
  - Sometimes still a meta-analysis

Visualizing GWAS results: Manhattan plots



- Each **dot** is the p-value from a single SNP's association
- X-axis = chromosomal location
- Y-axis = the **observed** p-values
- Height = strength of association



Verhaaren, et al. (2015) Circulation: Cardiovascular Genetics 8(2):398-409.

Meta-analysis: strengths and limitations

- **Strengths**
  - Significantly boosts **power** and **reduces false positives**
  - All cohorts measure the same SNPs (due to **imputation**)
  - **Only summary statistics** (beta, p) need to be shared with collaborators, not raw data

- **Limitations**
  - **Measurement** of trait often varies across cohorts (need for harmonization)
  - **Assumes common genetic effect** regardless of differences in non-genetic factors across cohorts (ex. SES, age)
  - **Limits complexity** of analysis
    - Number of models
    - Adjustment covariates

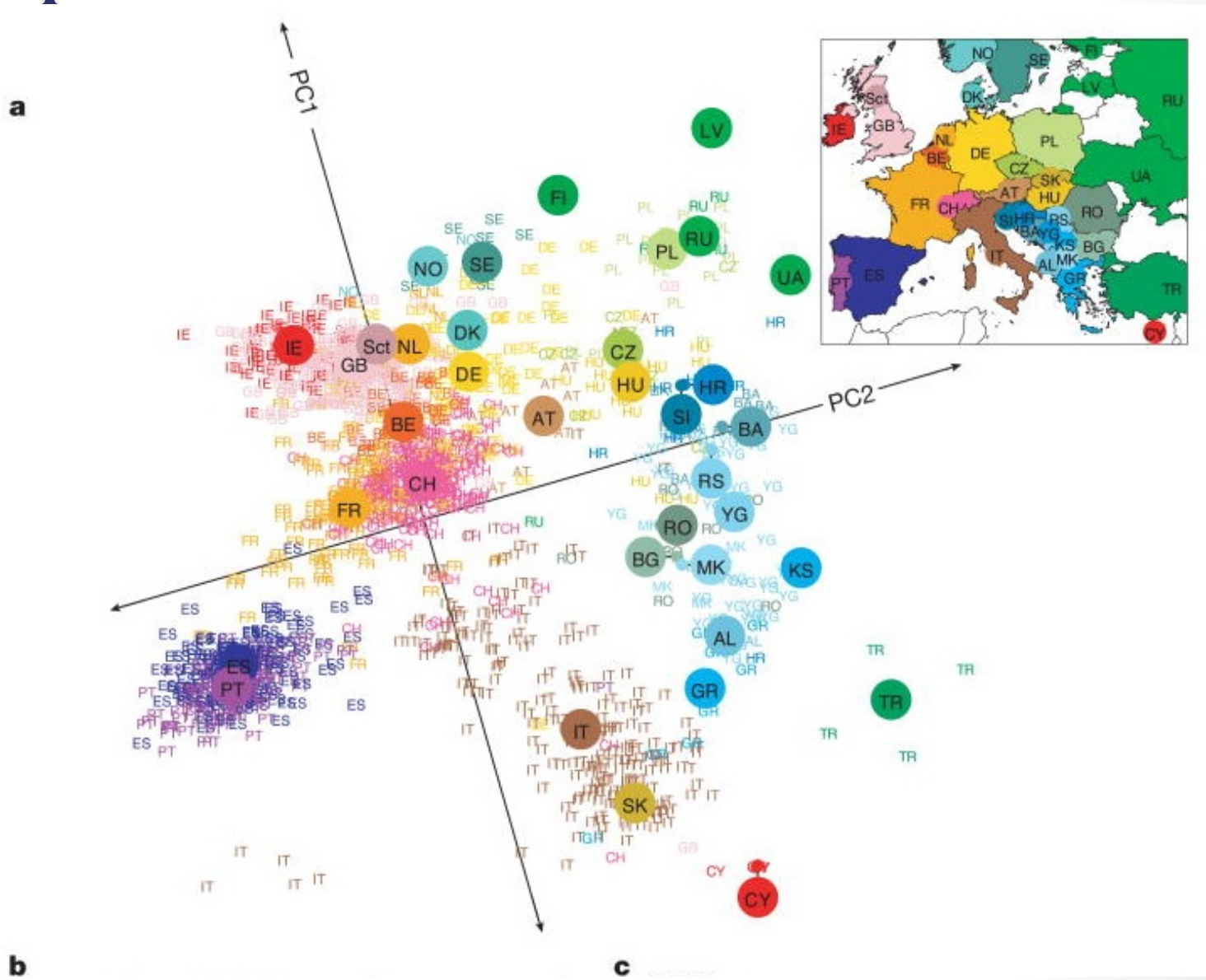
Race/Ethnicity/Caste and Ancestry

- **Race, Ethnicity, and Caste**-socially constructed and a key measure in examining social and health inequities

- **Ancestry**—variation in genetic architecture between populations
  - Used in genomic analyses
  - Results from many generations of demographic processes: migration, fertility, and mortality (which have been/are under social control)
  - Genetic variation ≠ causal effect on biology/health

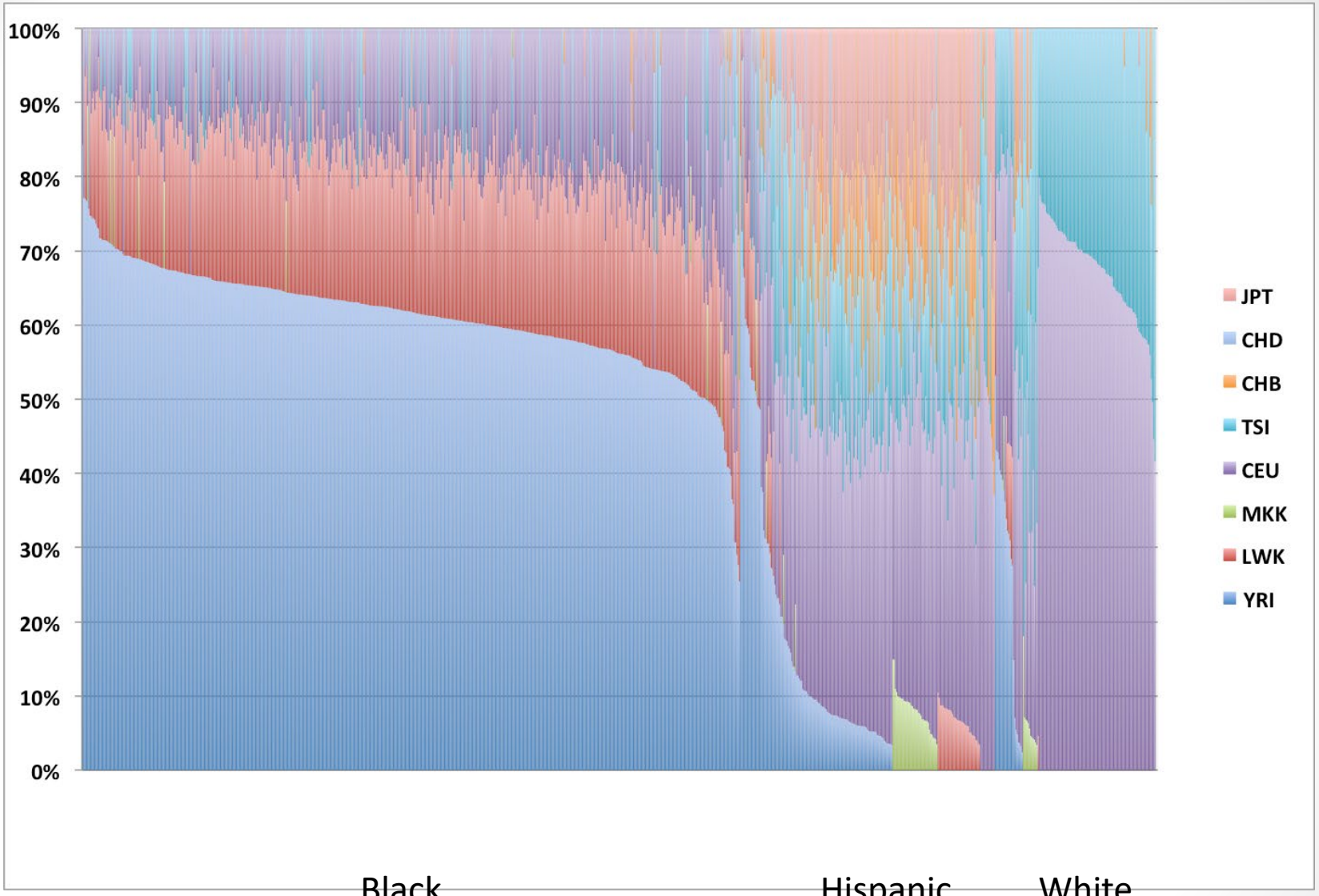
- The number and homogeneity of ancestries is very different by race and ethnicity

Population Stratification



Novembre, 2008, *Nature*

Population Stratification & Admixture of the Fragile Families and Child Wellbeing Study



JPT-Japanese (Tokyo),CHD-Chinese(Denver),CHB-Han(Beijing),TSI-Tuscan(Italy),CEU-European(Utah),MKK-Maasai(Kenya),LWK-Luhya(Kenya),YRI-Yoruba(Nigeria)

CVFS Genetic Data

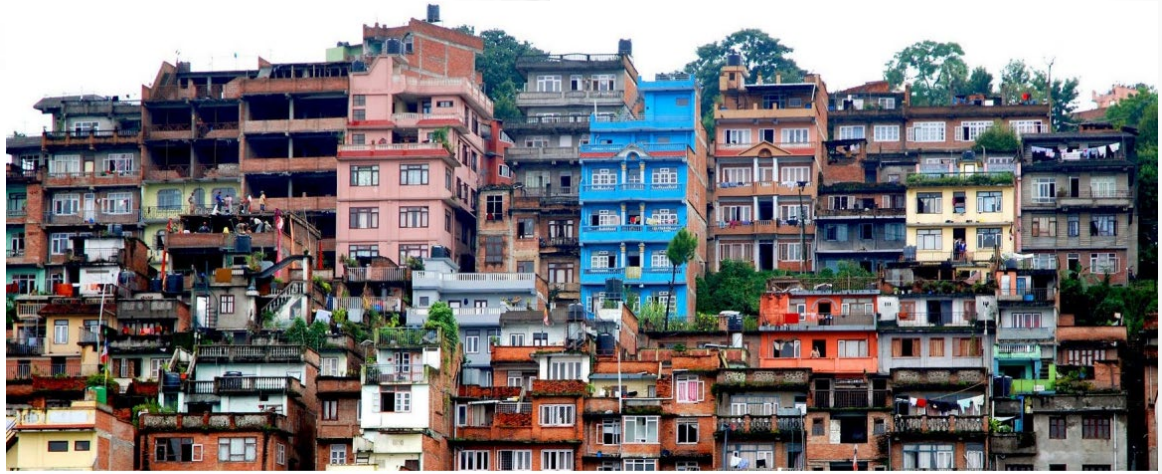
Genetic Data from CIDI Data

INSTITUTE FOR SOCIAL RESEARCH • SURVEY RESEARCH CENTER
SURVEY RESEARCH OPERATIONS
UNIVERSITY OF MICHIGAN

Sept. 29, 2021

Using A Life History Calendar to Improve Measurement of Lifetime Experience with Psychiatric Disorders

William Axinn and Stephanie Chardoul



LINKING DATA IN THE CVFS AND BEYOND

Emily Treleaven
Adrienne Epstein

October 13, 2021

CVFS Genetic Data

- Saliva-easy to collect, room temperature storage for months
- 96% of CIDI participants
- Separate consents for DNA collection and sharing Genetic Data

CVFS CIDI Data

Table 1. Sample description

	CVFS sample[a] 2016–2018 (*N* = 10 714)		Chitwan district 2011 (*N* = 579 984)		Nepal census 2011 (*N* = 26 494 504)	
Gender	Number	%	Number	%	Number	%
Male	4923	46.0 (45.0, 46.9)	279 087	48.1	12 849 041	48.5
Female	5791	54.1 (53.1, 55.0)	300 897	51.9	13 645 463	51.5
Age						
15–24	3935	36.7 (35.8, 37,6)	127 870	35.6	5 290 051	35.1
25–34	3008	28.1 (27.2, 28.9)	90 545	25.2	3 814 659	25.3
35–44	2080	19.4 (18.7, 20.2)	70 718	19.7	2 990 440	19.8
45–59	1691	15.8 (15.1, 16.5)	69 919	19.5	2 996 698	19.9
Ethnicity						
Brahmin/Chhetri	4634	43.3 (42.3, 44.2)	239 466	41.3	8 499 061	32.1
Hill Janajati	2106	19.7 (18.9, 20.4)	176 875	30.5	5 886 260	22.2
Dalit	1301	12.1 (11.5, 12.8)	50 655	8.7	3 474 767	13.1
Newar	640	6.0 (5.5, 6,4)	30 256	5.2	1 321 933	5.0
Terai Janajati	1942	18.1 (17.4, 18.9)	63 592	11.0	2 257 951	8.5
Others	91	0.9 (0.7, 1.0)	19 140	3.3	5 054 532	19.1
S.L.C. or more						
Yes	4098	38.3 (37.3, 39.2)	3 288 783	12.4	102 483	17.7
No	6615	61.8 (60.8, 62.7)	23 205 721	87.6	477 501	82.3

SLC = school leaving certificate.
[a]The age range in the CVFS sample is between 15 years and 59 years. For comparison, the calculated age distribution for the Chitwan district and Nepal census excludes people younger than 15 years and older than 59 years.

K. M. Scott et al. *Psychological Medicine*(2020)

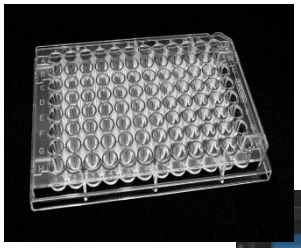
CVFS Genetic Data

- DNA was extracted in Nepal
- Shipped to Broad Institute for Genotyping
- Jordan Smoller (and PGC) leading quality control and analyses

PGC
Psychiatric Genomics Consortium

Jordan Smoller

BROAD INSTITUTE



Expected CVFS Genetic Data Family Relatedness

Family Relatedness	Number of Individuals
Trios (Mother, Father, Child)	5,398*
One parent, >1 sib-pair	1,435
Parent-child duo	618
No parents, >1 sib-pair	939
Unrelated	1,918
Total	**10,308**

*Families can have more than one set of trios if multiple children with both parents participated

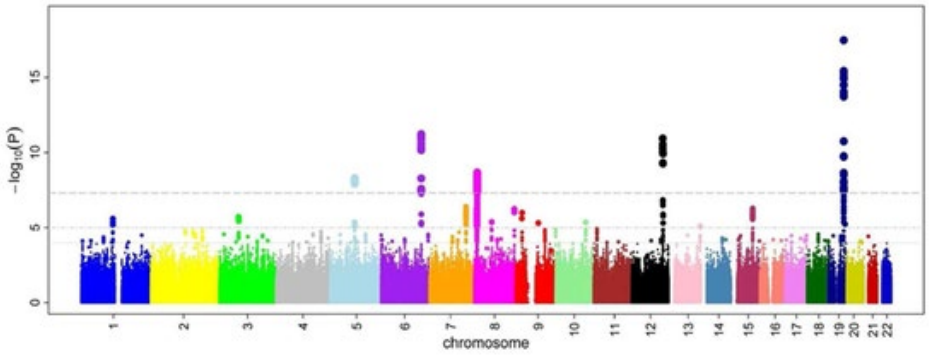
Likely close to 2,000 trios (very large from this type of study)
Over 5,100 mother-child pairs
Over 4,500 father-child pairs
Over 2,000 sibling pairs

Application: PGS

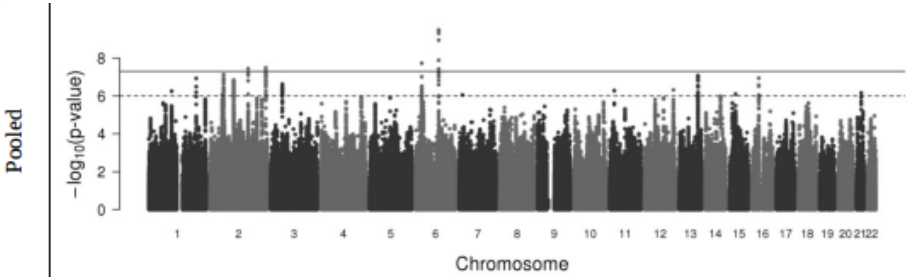
Typical approach for score construction

$$PGS_i = \sum_{j=1}^{J} W_j G_{ij}$$

IDNO	Polygenic_Score
100001	51374.52
100002	57506.1
100003	54567.35
100004	50922.69
100005	51467.5
100006	56791.58
100007	53955.28
100008	58652.57
100009	58987.74
100010	56127.94



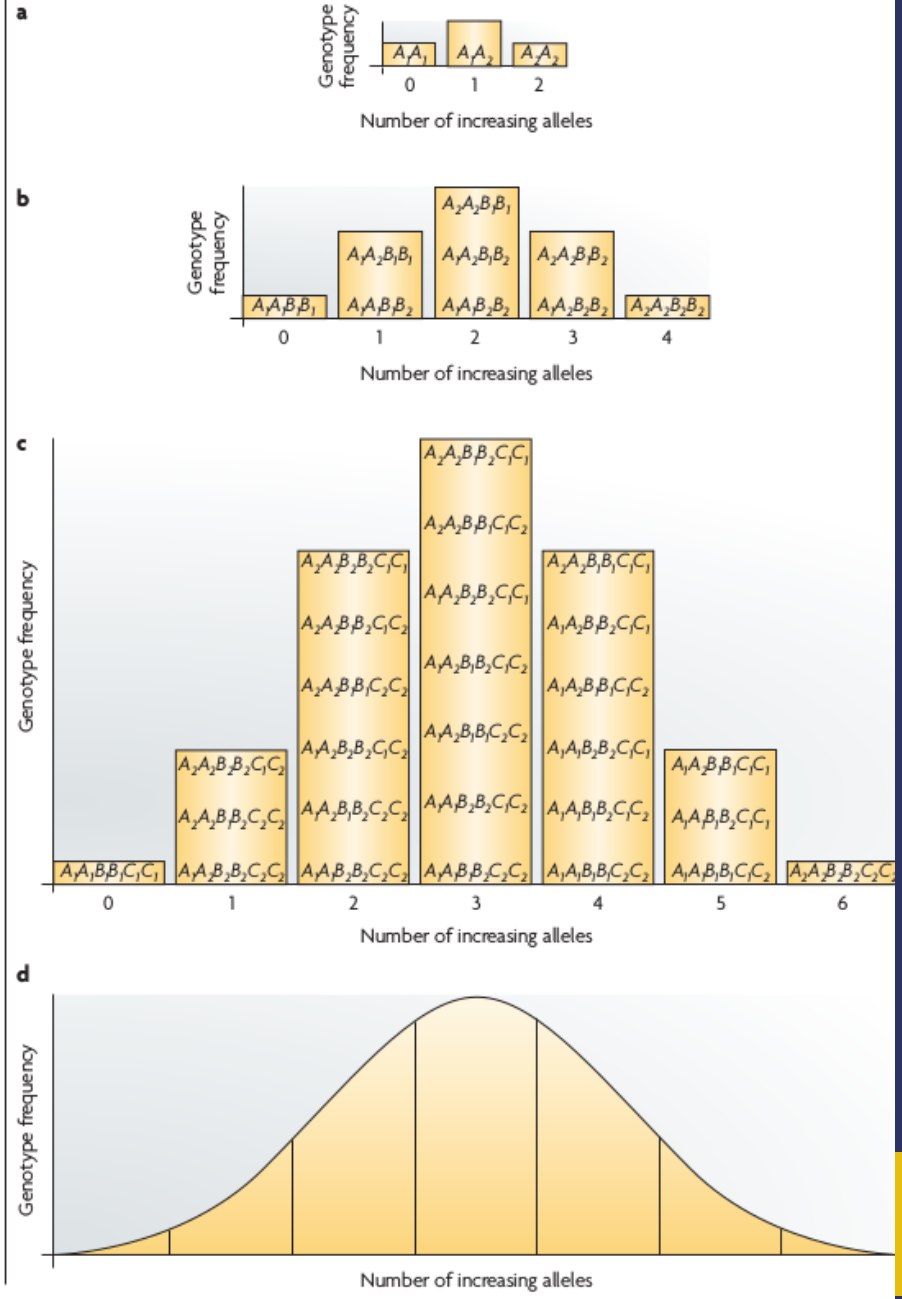
Score Construction



+

	SNP 1	SNP 2	$\cdots$	SNP 1,000,000
P1	0	1	$\cdots$	2
P2	1	0	$\cdots$	0
P3	1	2	$\cdots$	1
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
P1000	2	1	$\cdots$	2

$1,000 \times 1,000,000$ matrix; each cell $\in \{0, 1, 2\}$.



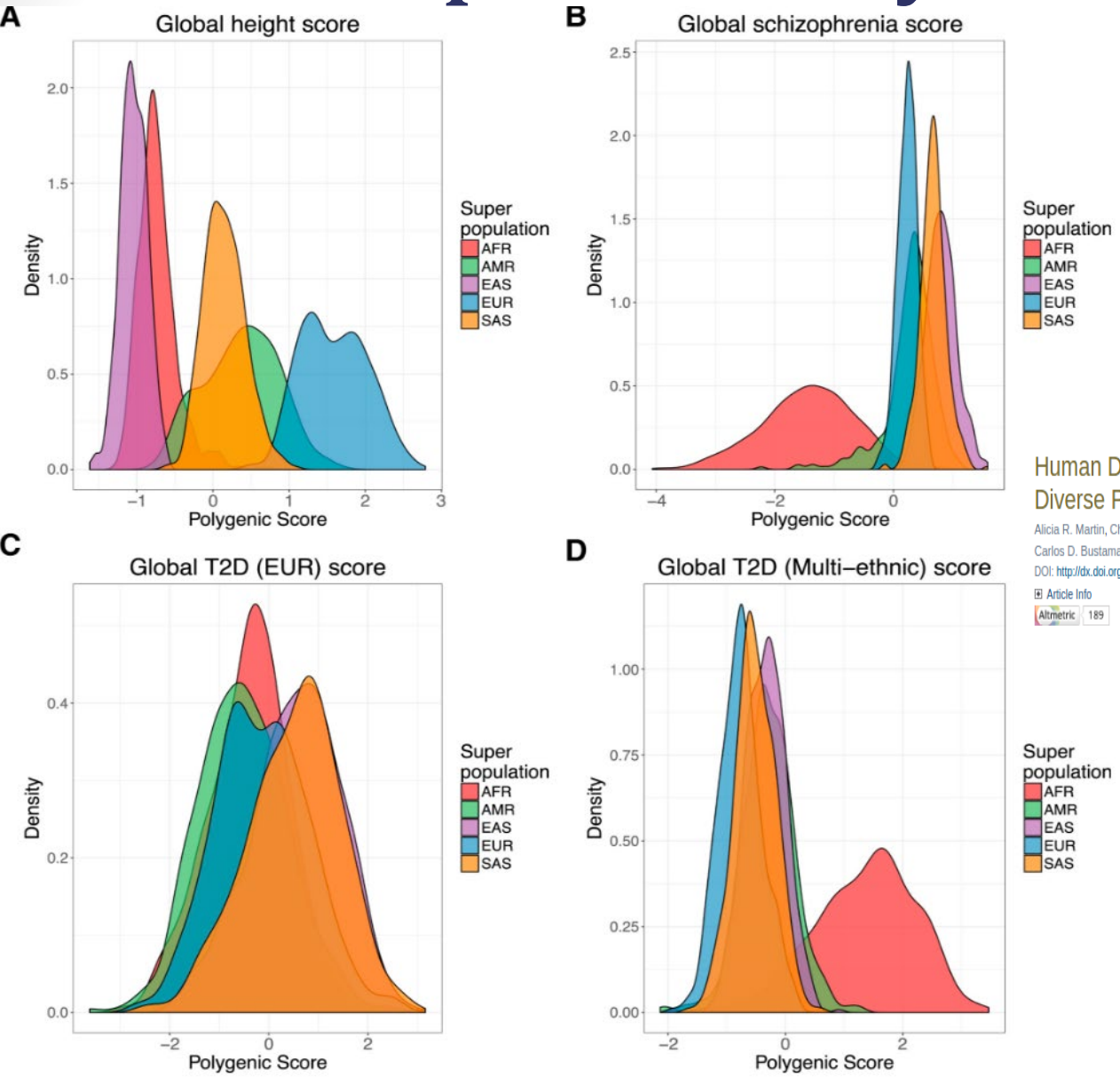
Polygenic Score Construction

Heterogeneity in polygenic scores for common human traits

Erin B Ware MPH PhD[*1], Lauren L Schmitz PhD[1], Jessica Faul MPH PhD[1], Arianna Gard MA[2], Colter Mitchell PhD[1], Jennifer A Smith MPH PhD[1,3], Wei Zhao PhD[3], David Weir PhD[1], Sharon LR Kardia PhD[3]

1- Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI 48104
2- Department of Psychology, College of Literature, Science, and the Arts, University of Michigan, Ann Arbor, MI 48109
3- Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI 48109
*Corresponding author ebakshis@umich.edu

PGS not portable by ancestry



Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations

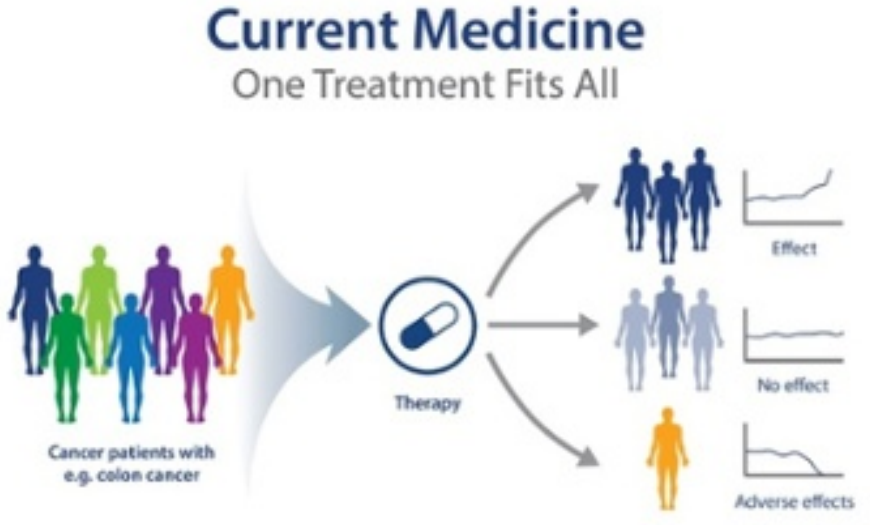
Alicia R. Martin, Christopher R. Gignoux, Raymond K. Walters, Genevieve L. Wojcik, Benjamin M. Neale, Simon Gravel, Mark J. Daly, Carlos D. Bustamante, Eimear E. Kenny

DOI: http://dx.doi.org/10.1016/j.ajhg.2017.03.004 | CrossMark

Article Info

Altmetric 189

Personalized/Precision Medicine



Typical trade off

more prediction=less biological insight

Selected Recent Reviews and Commentaries

- R. Roberts. Genetic Risk Stratification- Tipping Point for Global Primary Prevention of Coronary Artery Disease , Circulation. 2018;137:2554-2556.
- A. Torkamani, et al. The personal and clinical utility of polygenic risk scores. Nature Reviews Genetics May 2018
- L. Hercher. Genome Culture: A Personal Risk Score May Be the Next Big Thing in Genetic Medicine , Genome Magazine, April 2018
- J.W. Knowles, et al. Cardiovascular disease: The rise of the genetic risk score. PLoS Medicine 2018 Mar 15(3) e1002546
- K. Beaney, et al. How close are we to implementing a genetic risk score for coronary heart disease? Expert review of molecular diagnostics 2017 Oct 17(10) 905-915
- S. Mistry, et al. The use of polygenic risk scores to identify phenotypes associated with genetic risk of bipolar disorder and depression: A systematic review. Journal of affective disorders. 2018 Jul;234:148-155.
- Martin, A.R., Kanai, M., Kamatani, Y. *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet* **51,** 584–591 (2019)

Application: Gene by Environment Interaction (GxE) and Correlation (rGE)

GxE Reviews

American Journal of PUBLIC HEALTH

Special Issue: Integration of Behavioral, Social Science and Genetics Research,

Vol. 103, S1 (October 2013)

Genetic Differential Sensitivity to Social Environments: Implications for Research

Colter Mitchell, PhD, Sara McLanahan, PhD, Jeanne Brooks-Gunn, PhD, Irwin Garfinkel, PhD, John Hobcraft, BSc, and Daniel Notterman, MD

Defining the Environment in Gene–Environment Research: Lessons From Social Epidemiology

Jason D. Boardman, PhD, Jonathan Daw, PhD, and Jeremy Freese, PhD

Annual Review of Psychology
2014. 65:41–70

Gene-Environment Interaction

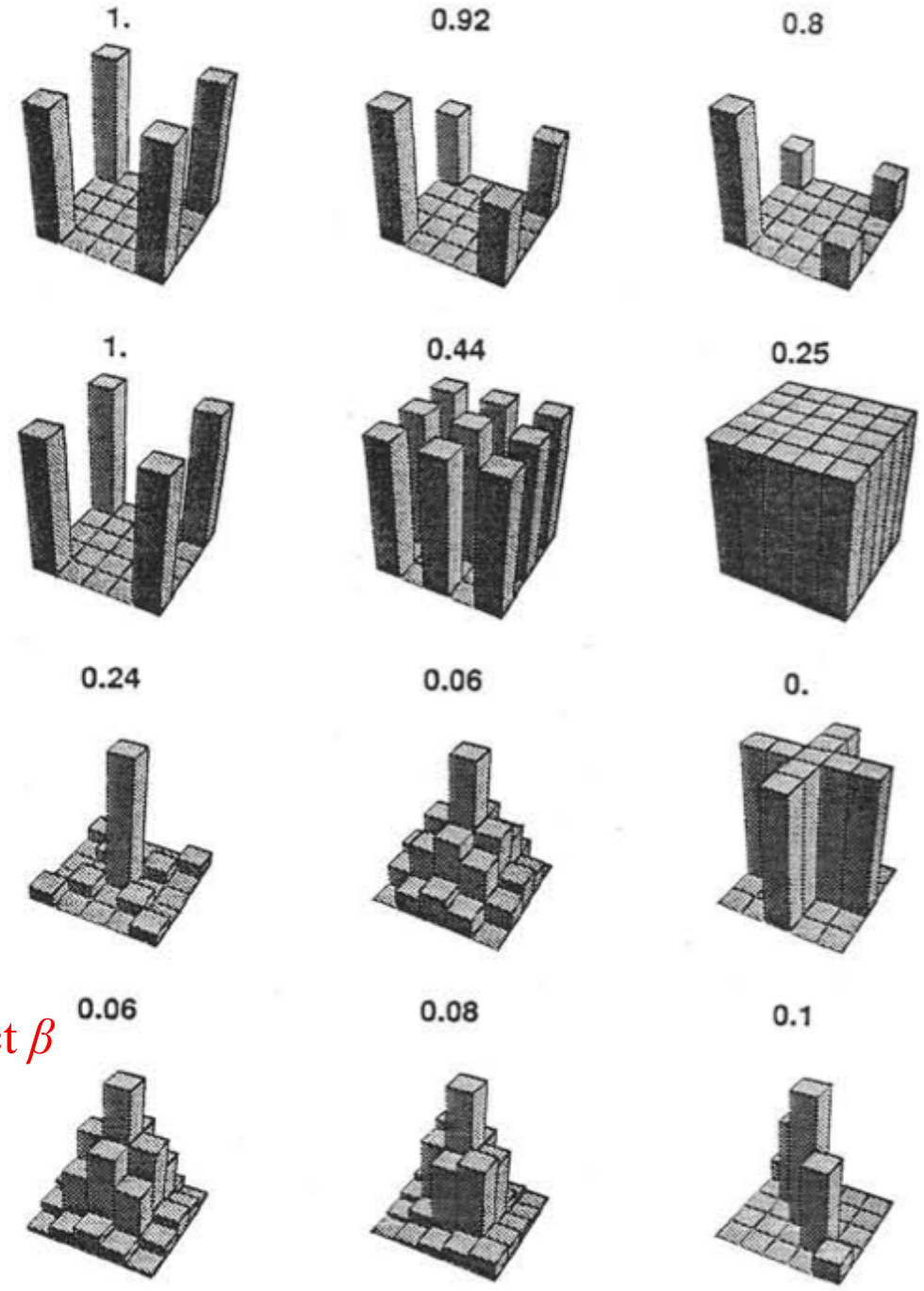
Stephen B. Manuck[1] and Jeanne M. McCaffery[2]

[1]Department of Psychology, University of Pittsburgh, Pittsburgh, Pennsylvania 15260; email: manuck@pitt.edu

[2]Department of Psychiatry and Human Behavior, The Miriam Hospital, and Warren Alpert School of Medicine at Brown University, Providence, Rhode Island 02903; email: jeanne_mccaffery@brown.edu

Power and Interactions:

Intuitions from experiments do not carry to field studies



4x the sample if interaction β= main effect β

16x the sample if interaction β= ½ main effect β

McClelland & Judd, 1993 *Psych Bull.*

Gene-Environment Correlation

- Active-genes choose environment
- Evocative-environment responds to genes
- Passive-genes and environment are passed together

Mitchell, et al. 2013. *AJPH*

Education genetics shape environments

Psychological Science OnlineFirst, published on June 1, 2016 as doi:10.1177/0956797616643070

Research Article

The Genetics of Success: How Single-Nucleotide Polymorphisms Associated With Educational Attainment Relate to Life-Course Development

Daniel W. Belsky[1,2], Terrie E. Moffitt[3,4,5,6], David L. Corcoran[7], Benjamin Domingue[7], HonaLee Harrington[3], Sean Hogan[8], Renate Houts[3], Sandhya Ramrakha[9], Karen Sugden[3], Benjamin S. Williams[3], Richie Poulton[9], and Avshalom Caspi[3,4,5,6]

[1]Department of Medicine, Duke University School of Medicine; [2]Social Science Research Institute, Duke University; [3]Department of Psychology & Neuroscience, Duke University; [4]Department of Psychiatry and Behavioral Sciences, Duke University School of Medicine; [5]Center for Genomic and Computational Biology, Duke University; [6]MRC Social, Genetic & Developmental Psychiatry Research Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London; [7]Graduate School of Education, Stanford University; and [8]Dunedin Multidisciplinary Health & Development Research Unit, Department of Psychology, University of Otago

Abstract

A previous genome-wide association study (GWAS) of more than 100,000 individuals identified molecular-genetic predictors of educational attainment. We undertook in-depth life-course investigation of the polygenic score derived from this GWAS using the four-decade Dunedin Study ($N = 918$). There were five main findings. First, polygenic scores predicted adult economic outcomes even after accounting for educational attainments. Second, genes and environments were correlated: Children with higher polygenic scores were born into better-off homes. Third, children's polygenic scores predicted their adult outcomes even when analyses accounted for their social-class origins; social mobility analysis showed that children with higher polygenic scores were more upwardly mobile than children with lower scores. Fourth, polygenic scores predicted behavior across the life course, from early acquisition of speech and reading skills through geographic mobility and mate choice and on to financial planning for retirement. Fifth, polygenic-score associations were mediated by psychological characteristics, including intelligence, self-control, and interpersonal skill. Effect sizes were small. Factors connecting DNA sequence with life outcomes may provide targets for interventions to promote population-wide positive development.

Keywords

genetics, behavior genetics, intelligence, personality, adult development

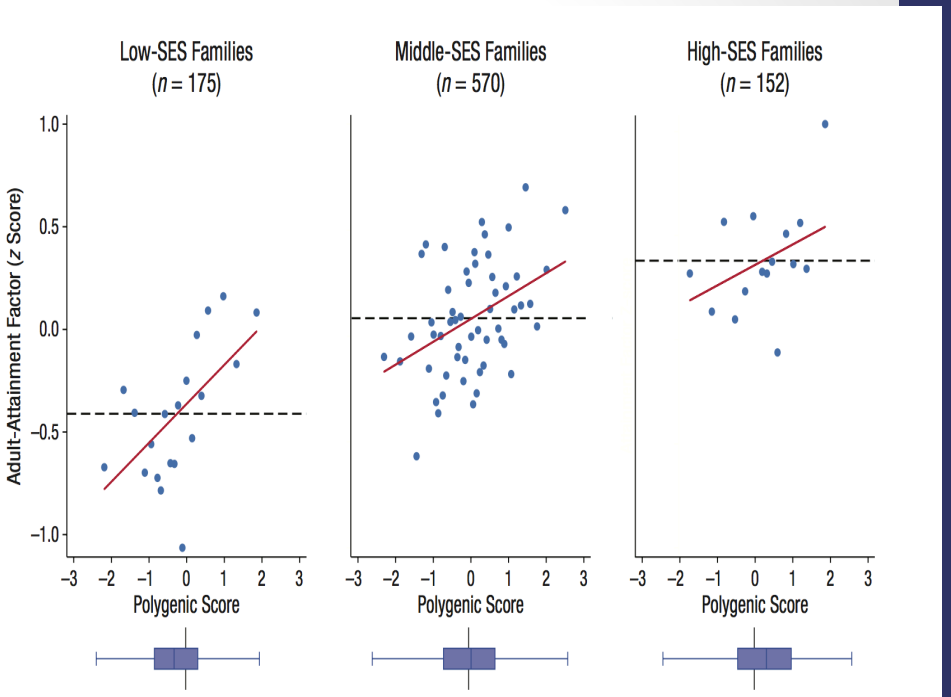
Received 12/28/15; Revision accepted 5/14/16

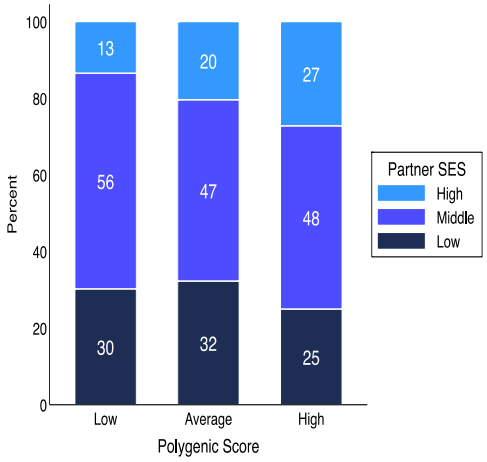
In 2013, scientists reported the first successful genome-wide association study (GWAS) of a social-science outcome, educational attainment (Rietveld et al., 2013). Their analysis of millions of genetic variants in more than 100,000 individuals hinted at the existence of a molecular map to success in schooling written in the alphabet of DNA. As anticipated, rather than finding a so-called gene for education, this study revealed a genetic continuum: Some individuals carry very few alleles associated with educational attainment, the bulk of the population carries

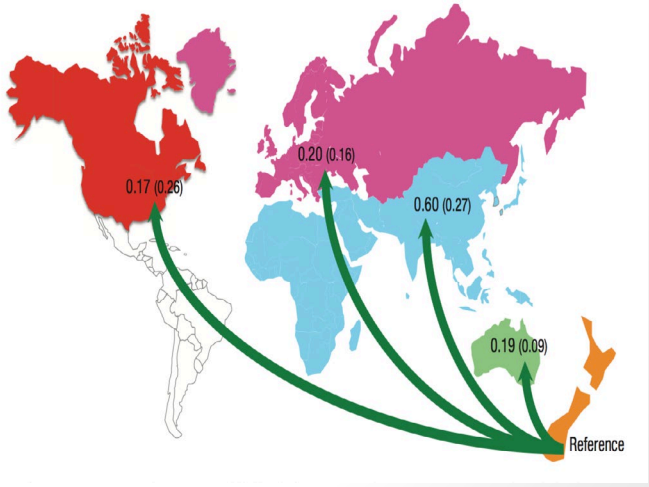
Corresponding Author:

Daniel W. Belsky, 2020 W. Main St., Suite 201, Durham, NC 27708

E-mail: dbelsky@duke.edu



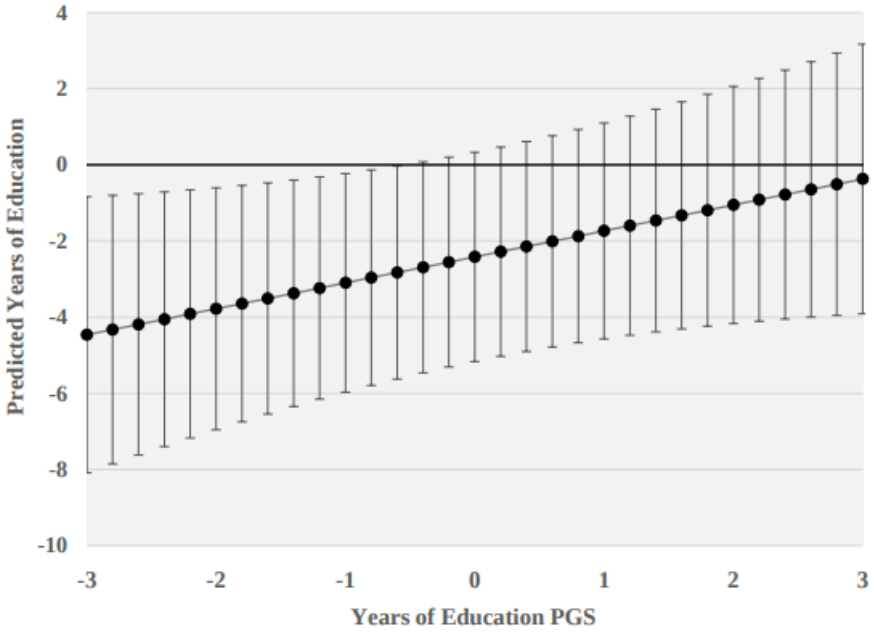




Slide courtesy of Daniel Belsky

The role of macro environments

Figure 2. Difference in Predicted Years of Education: Veterans versus Non-Veterans



THE EFFECT OF VIETNAM-ERA CONSCRIPTION AND GENETIC POTENTIAL FOR EDUCATIONAL ATTAINMENT ON SCHOOLING OUTCOMES

Lauren L. Schmitz
Dalton Conley

Working Paper 22393
http://www.nber.org/papers/w22393

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2016

NATURE HUMAN BEHAVIOUR LETTERS

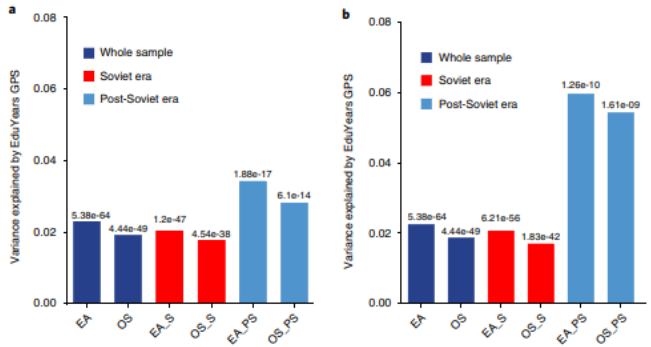


Fig. 1 | Variance explained by EduYears GPS in the post-Soviet and Soviet groups. a,**b**, The GPS was calculated using a 0.1 GWA study *P* value threshold for educational attainment (EA) and occupational status (OS) for the whole EGCUT sample (N(EA) = 12,483; N(OS) = 11,419) and when divided into historical eras using two cutoffs: the post-Soviet (PS) group included participants 15 years or younger when independence was regained, and the Soviet (S) group included the rest of the participants (N(EA_S) = 10,381; N(OS_S) = 9,417; N(EA_PS) = 2,102; N(OS_PS) = 2,002) (**a**); the post-Soviet (PS) group included participants 10 years or younger when independence was regained and the Soviet (S) group included the rest of the participants (N(EA_S) = 11,808; N(OS_S) = 10,767; N(EA_PS) = 675; N(OS_PS) = 652) (**b**).

nature human behaviour
LETTERS
https://doi.org/10.1038/s41562-018-0332-5

Genetic influence on social outcomes during and after the Soviet era in Estonia

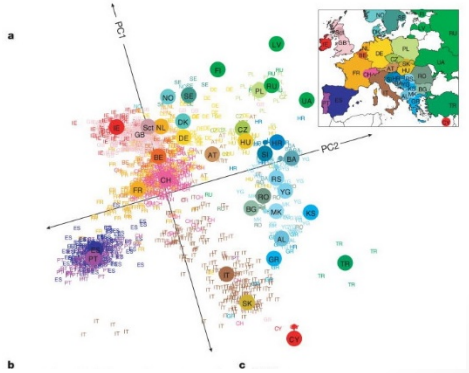
Kaili Rimfeld[1*], Eva Krapohl[1], Maciej Trzaskowski[2], Jonathan R. I. Coleman[1,3], Saskia Selzam[1], Philip S. Dale[4], Tonu Esko[5], Andres Metspalu[5] and Robert Plomin[1]

Application: Family Models

Family Models

Family Relatedness	Number of Individuals
Trios (Mother, Father, Child)	5,398*
One parent, >1 sib-pair	1,435
Parent-child duo	618
No parents, >1 sib-pair	939
Unrelated	1,918
Total	**10,308**

- Trio (Mother-Father-Child) analyses-break the effect of population stratification



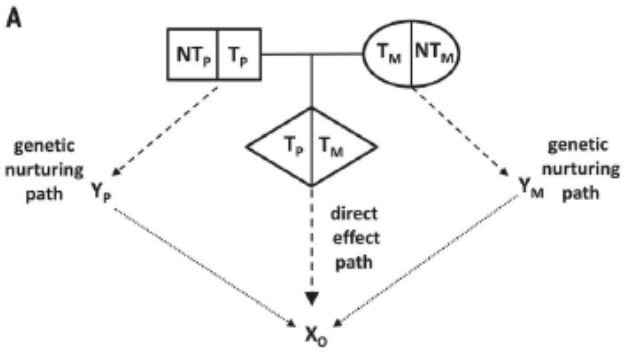
Genetic Nurture

GWAS of socially influenced outcomes will pick up (a lot?) more than just biological influences.

The social genome of friends and schoolmates in the National Longitudinal Study of Adolescent to Adult Health

Benjamin W. Domingue, Daniel W. Belsky, Jason M. Fletcher, Dalton Conley, Jason D. Boardman, and Kathleen Mullan Harris

PNAS January 9, 2018. 201711803; published ahead of print January 9, 2018. https://doi.org/10.1073/pnas.1711803115



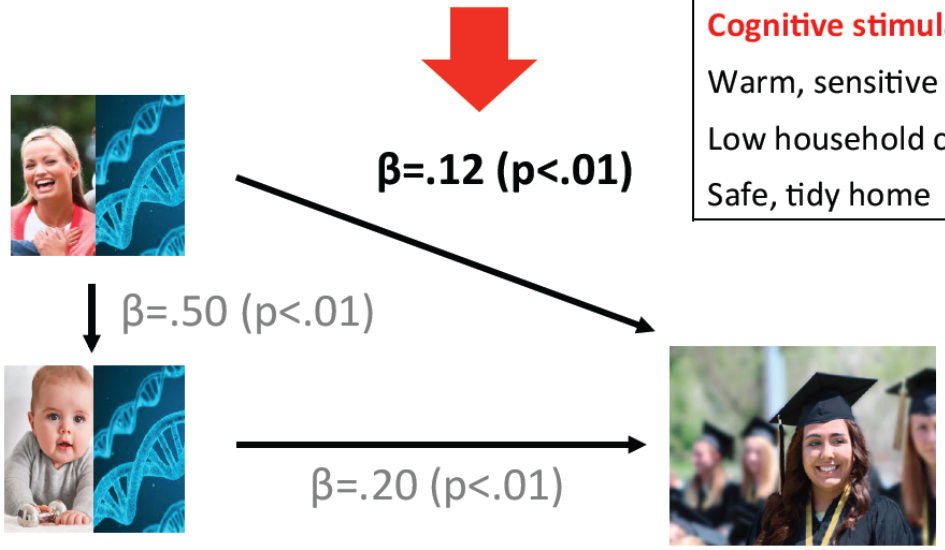
HUMAN GENOMICS

The nature of nurture: Effects of parental genotypes

Augustine Kong,[1,2,3*] Gudmar Thorleifsson,[1] Michael L. Frigge,[1] Bjarni J. Vilhjalmsson,[4,5] Alexander I. Young,[1,2,6] Thorgeir E. Thorgeirsson,[1] Stefania Benonisdottir,[1] Asmundur Oddsson,[1] Bjarni V. Halldorsson,[1] Gisli Masson,[1] Daniel F. Gudbjartsson,[1,3] Agnar Helgason,[1,7] Gyda Bjornsdottir,[1] Unnur Thorsteinsdottir,[1,8] Kari Stefansson[1,8*]

Passive rGe and Genetic Nurture

Evidence for genetic nurture

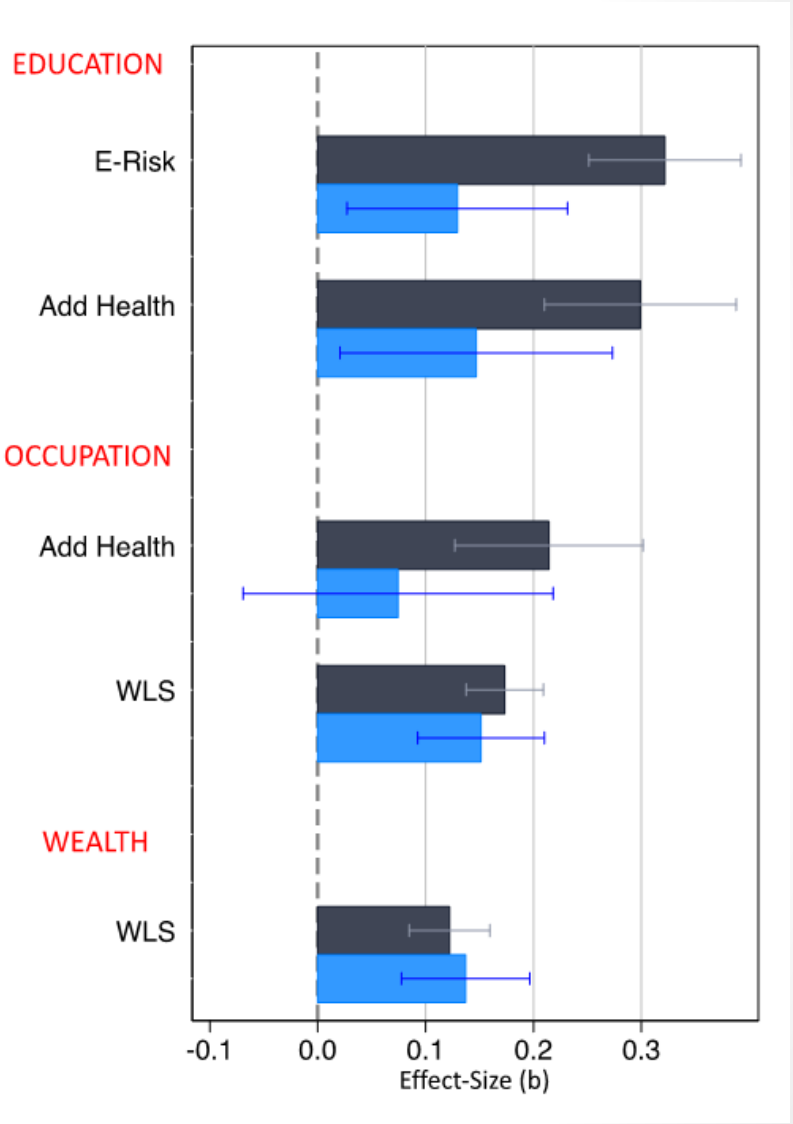


MEDIATION BY MOMS PARENTING	% mediated	P-val.
Cognitive stimulation	**75%**	**<.01**
Warm, sensitive parenting	17%	ns
Low household chaos	42%	<.01
Safe, tidy home	25%	<.01

Wertz et al 2019 *Dev Psych*

Sibling comparisons

- Controls for any influences shared by siblings growing up in the same household
- Slight variations in genetic differences is random—more causal estimate
- Typically attenuates effects (see right)



Belsky et al, 2018 *PNAS*

Social Scientists and Genetics

- NIA Biomarker Network https://gero.usc.edu/cbph/network
- Social Science Genetic Association Consortium- https://www.thessgac.org/
- Psychiatric Genomics Consortium- https://www.med.unc.edu/pgc/

- Genomics for Social Scientists- https://hrs.isr.umich.edu/genomics-workshop