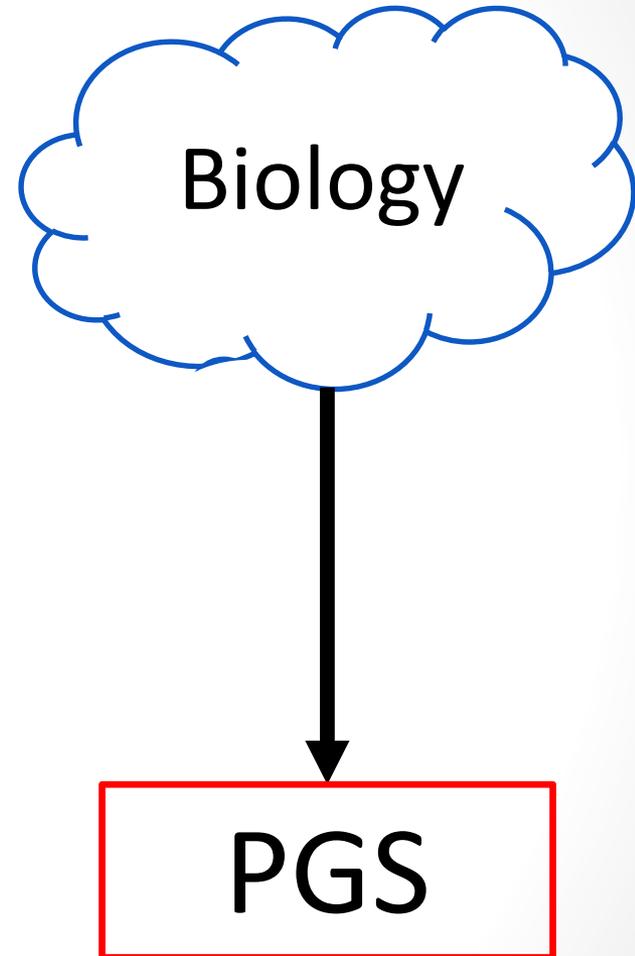


Potential and Pitfalls of Polygenetic Scores for Social and Demographic Research

Colter Mitchell
CVFS Webinar Series
March 9, 2022

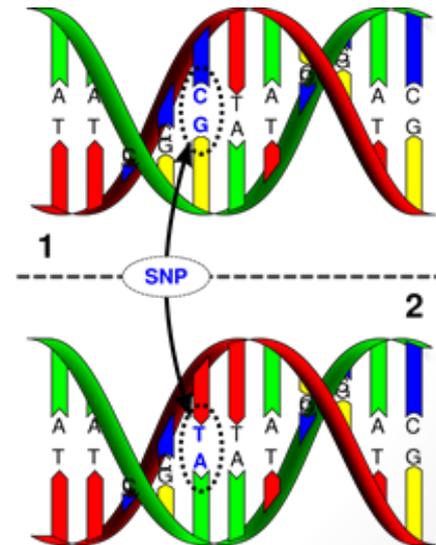
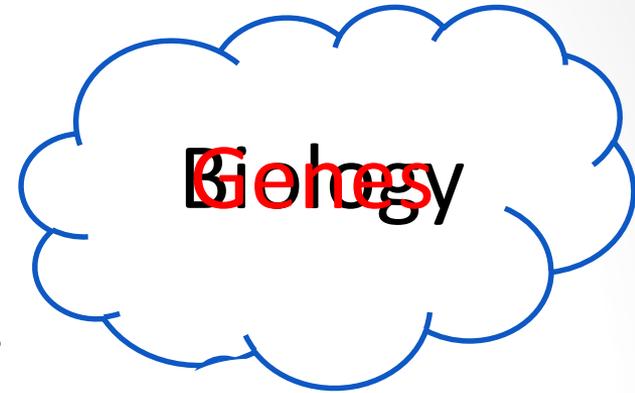
35-minute primer

- **Potential**—why the idea of a genetic measure is important for social & demographic research
- **Pitfalls**—How PGS operationalize genetic measures and concerns about them
- **Practicalities**—What is possible now and likely in the future



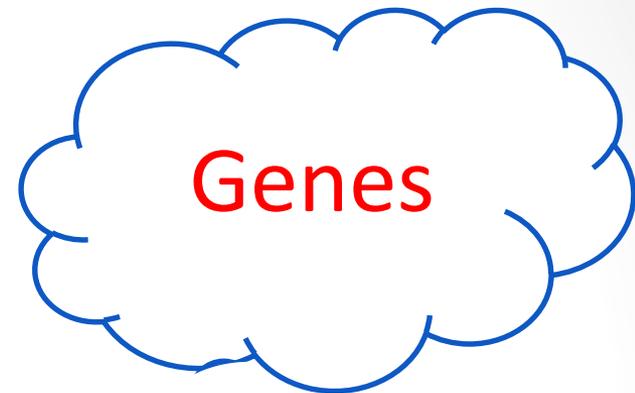
Potential

- **Constant**-does not change over life course
 - We inherit two pieces of genetic information, one from each parent
 - Same for all cells in the body



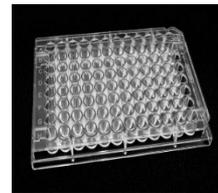
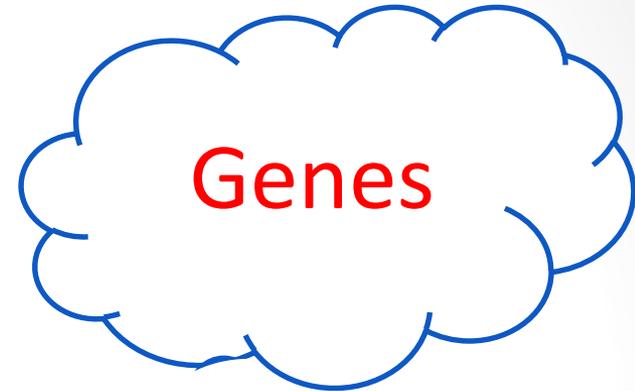
Potential

- **Constant**-does not change over life course
- **Predictive**-heritability estimates
 - Thousands of behavioral measures from 10's of millions of individuals (family-based studies) show genes are involved in behavior (0.49)
 - Fertility-.24-0.33, Education-0.25-.35, depression 0.4-0.6, alcohol use 0.35-0.7, etc
 - Population estimate and assumptions



Potential

- **Constant**-does not change over life course
- **Predictive**—heritability estimates
- **(Relatively) inexpensive**
 - Collection and genotyping can be (much) less than \$100 for millions of variables



Why would social scientists care about genes?

- **Controls**-better measurement of the social contextual effects
 - Improve power to detect social effects
 - Increase strength of argument by removing potential confounders
 - Use as an instrument due to random nature of allele selection (Mendelian Randomization)

Why would social scientists care about genes?

- **Controls**-better measurement of the social contextual effects
- **Moderation**- differential response to the environment



Genetic Differential Sensitivity to Social Environments: Implications for Research

| Colter Mitchell, PhD, Sara McLanahan, PhD, Jeanne Brooks-Gunn, PhD, Irwin Garfinkel, PhD, John Hoxby, BSc, and Daniel Natterman, MD

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

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

Gene–Environment Interaction

Stephen B. Manuck¹ and Jeanne M. McCaffery²

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

²Department 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

Annual Review of Psychology
2014. 65:41–70



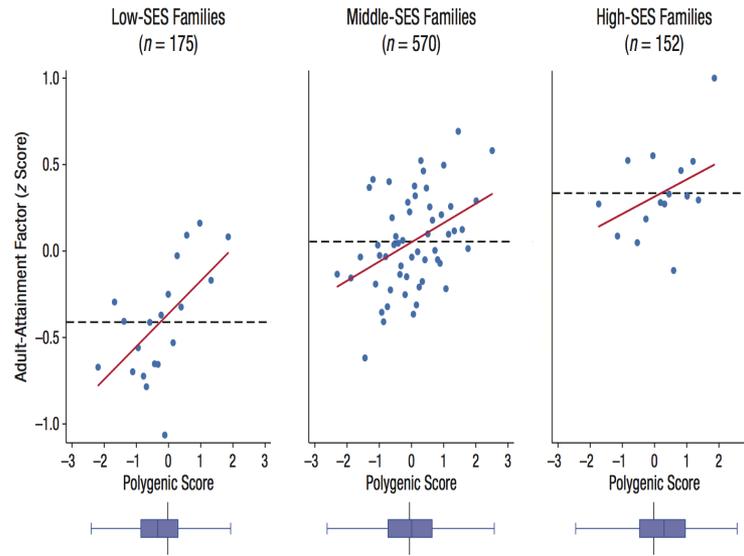
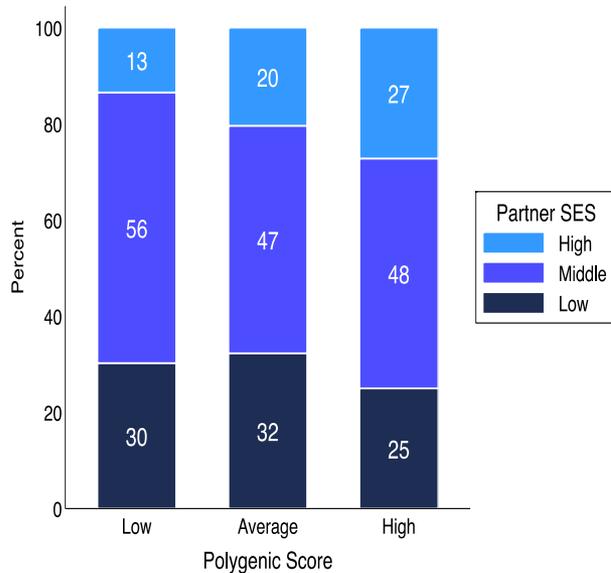
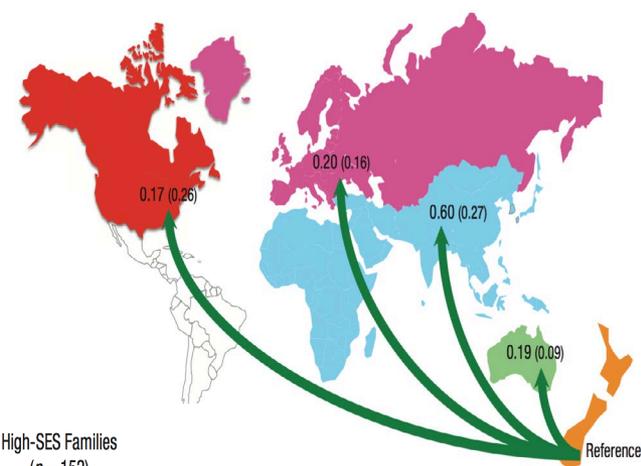
Why would social scientists care about genes?

- **Controls**-better measurement of the social contextual effects
- **Moderation**- differential response to the environment
- **Gene-environment correlation**-why are there consistent patterns of demographic and social sorting over geography and time (and intergenerational)

Gene-environment correlation

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

Do genes shape the environment?



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

aps
ASSOCIATION FOR
PSYCHOLOGICAL SCIENCE

Psychological Science
1-16
© The Author(s) 2016
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0956797616643070
ps.sagepub.com
SAGE

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. Corcoran⁵, Benjamin Domingue⁷, HonaLee Harrington³, Sean Hogan⁸, Renate Houts³, Sandhya Ramrakha⁸, Karen Sugden³, Benjamin S. Williams³, Richie Poulton³, and Avshalom Caspi^{3,4,5,6}

¹Department of Medicine, Duke University School of Medicine; ²Social Science Research Institute, Duke University; ³Department of Psychology & Neuroscience, Duke University; ⁴Department of Psychiatry and Behavioral Sciences, Duke University School of Medicine; ⁵Center for Genomic and Computational Biology, Duke University; ⁶MRC Social, Genetic & Developmental Psychiatry Research Centre,



Passive rGE and Genetic Nurture

Evidence for genetic nurture

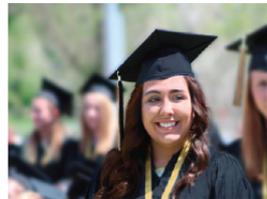


$\beta = .12$ ($p < .01$)

$\beta = .50$ ($p < .01$)



$\beta = .20$ ($p < .01$)

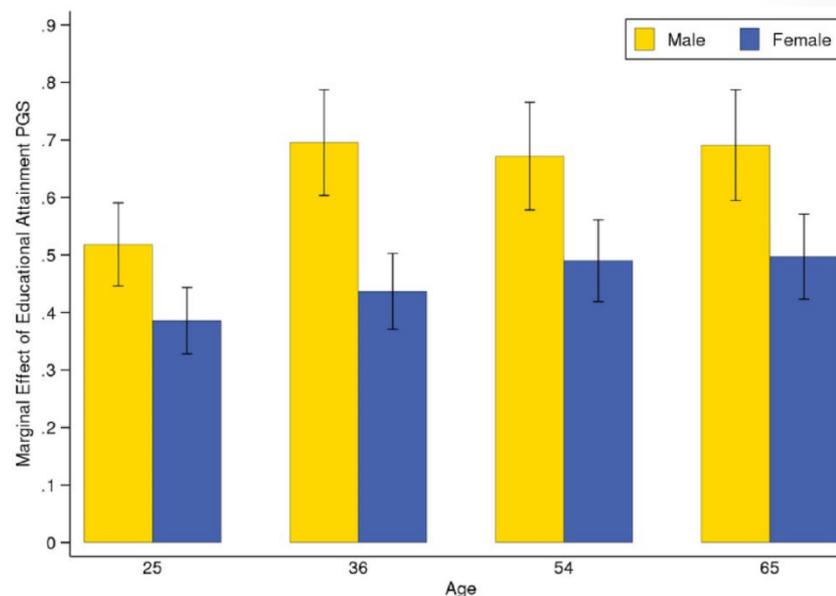
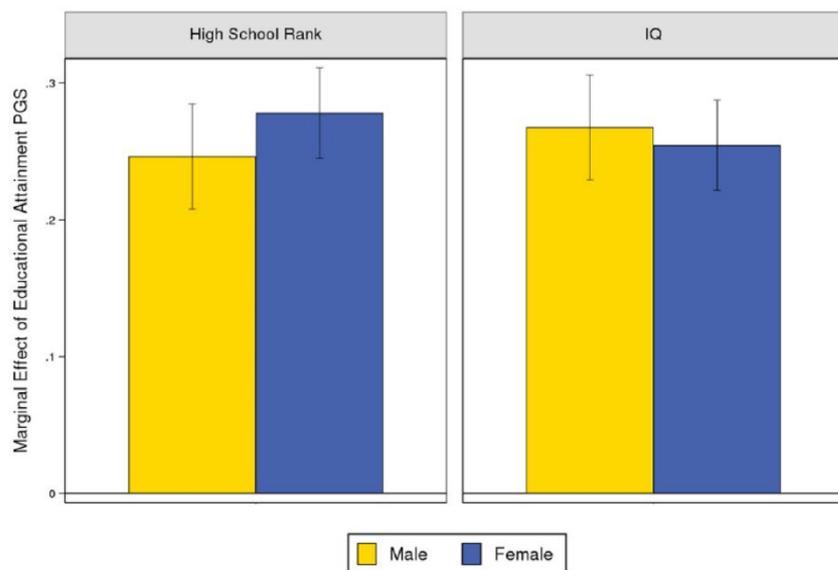


MEDIATION BY MOMS PARENTING	% mediated	P-val.
Cognitive stimulation	75%	<.01
Warm, sensitive parenting	17%	ns
Low household chaos	42%	<.01
Safe, tidy home	25%	<.01

Why would social scientists care about genes?

- **Controls**-better measurement of the social contextual effects
- **Moderation**- differential response to the environment
- **Gene-environment correlation**-why are there consistent patterns of demographic and social sorting over geography and time (and intergenerational)
- **Highlighting the social**-reveal the effect of the environment by comparing the strength of genetic effects

Highlight social influences



Genes, Gender Inequality, and Educational Attainment

Pamela Herd,^a Jeremy Freese,^b Kamil Sicinski,^c
Benjamin W. Domingue,^b Kathleen Mullan Harris,^d
Caiping Wei,^d and Robert M. Hauser^c

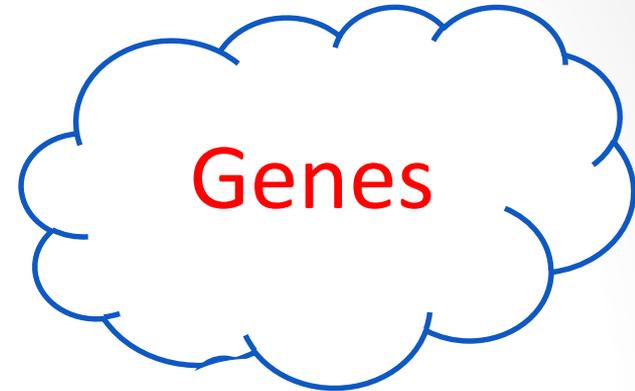
American Sociological Review
2019, Vol. 84(6) 1069–1098
© American Sociological
Association 2019
DOI: 10.1177/0003122419886550
journals.sagepub.com/home/asr



Potential

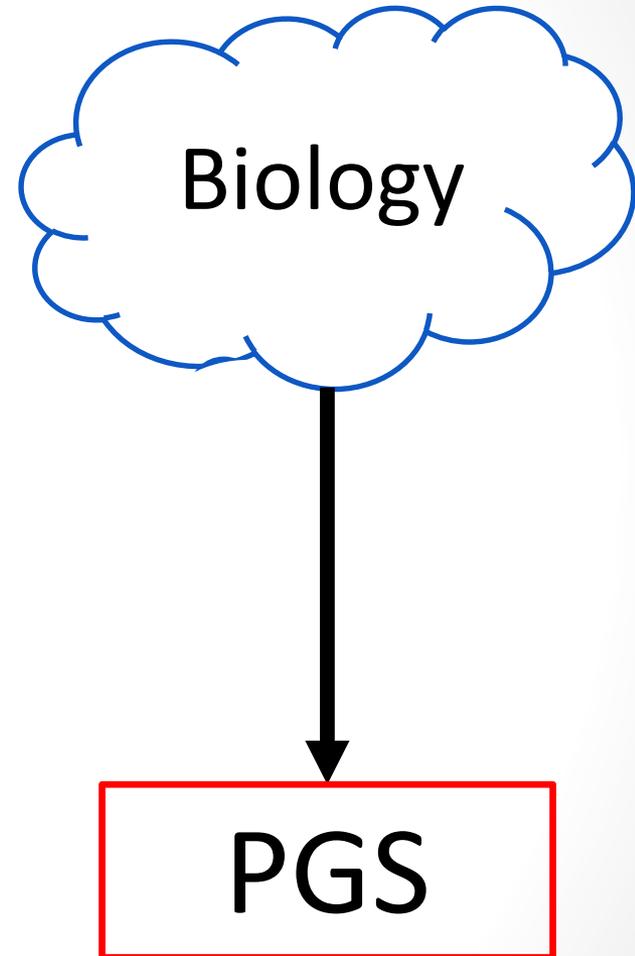
- Constant
- Predictive
- Inexpensive
- Useful for several applications

- So why isn't everyone onboard?

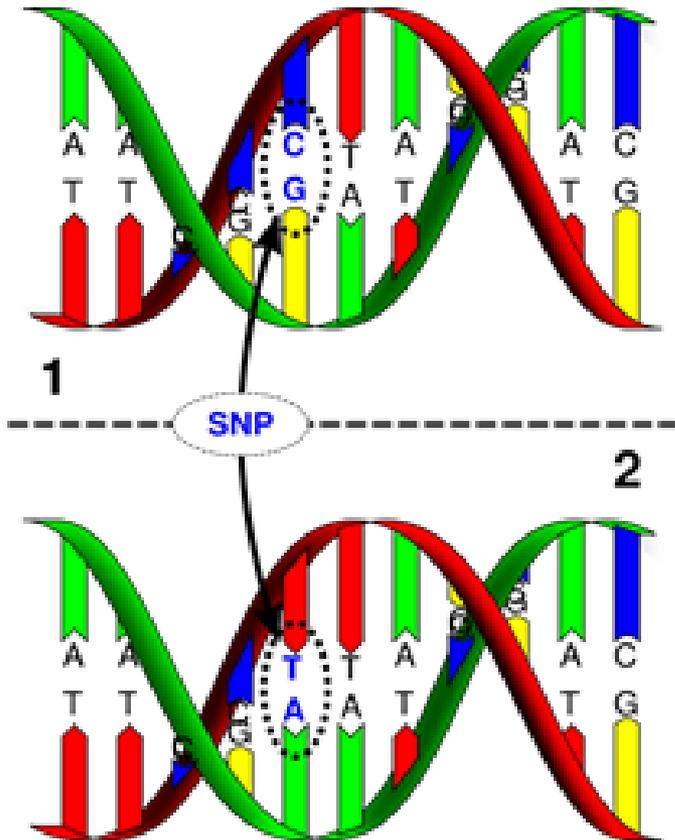


Pitfalls

- Genetic Research
- GWAS
- Construction of PGS
- Ancestry
- Sample Size and Selection

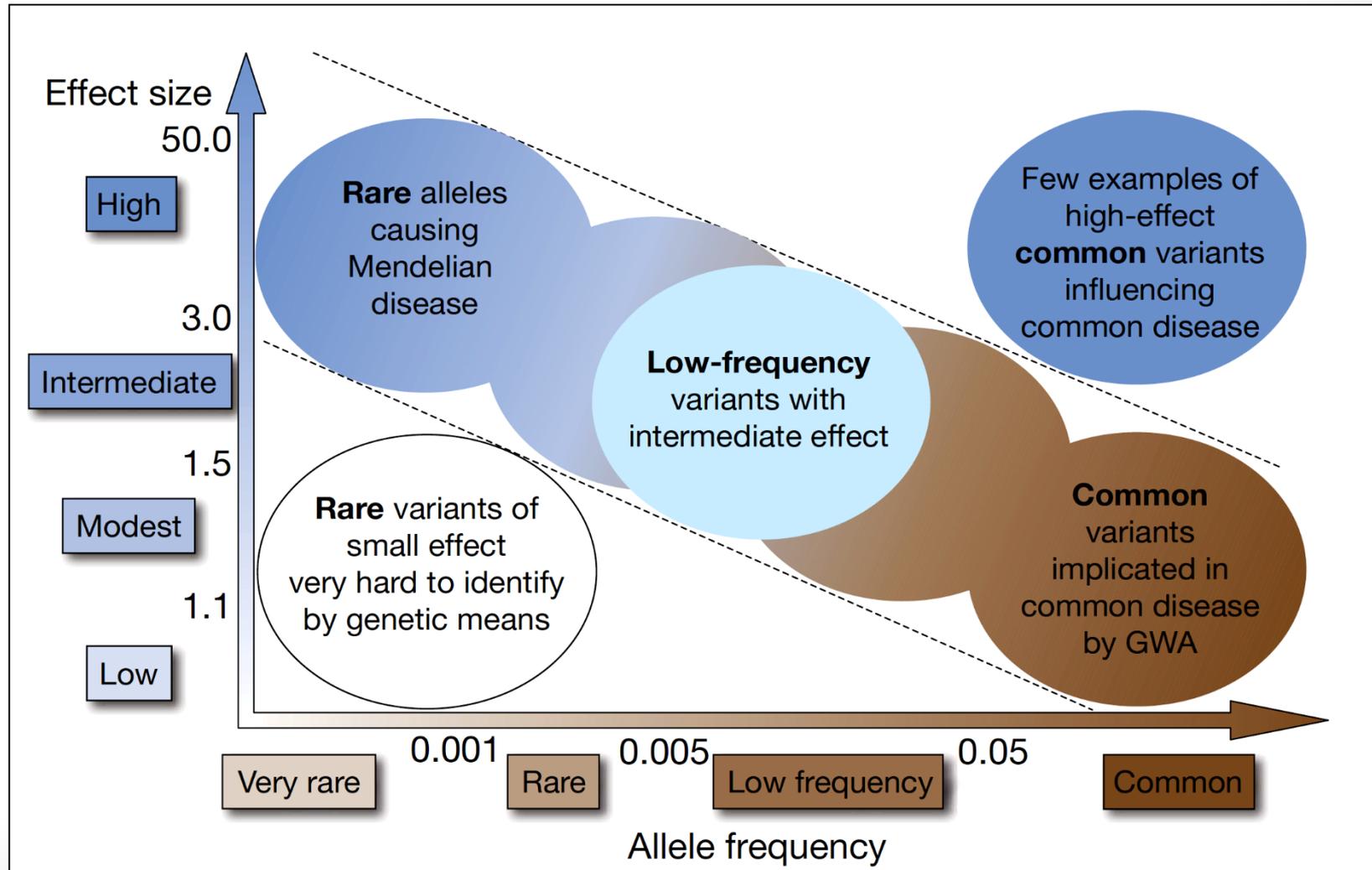


Measuring the genome



- Single Nucleotide Polymorphisms (SNPs)
- Simplest form of genetic variation.
- SNPs occur $\sim 1/300$ base pairs
- In $>1\%$ of the population (allele)
- Simple coding 0,1,2

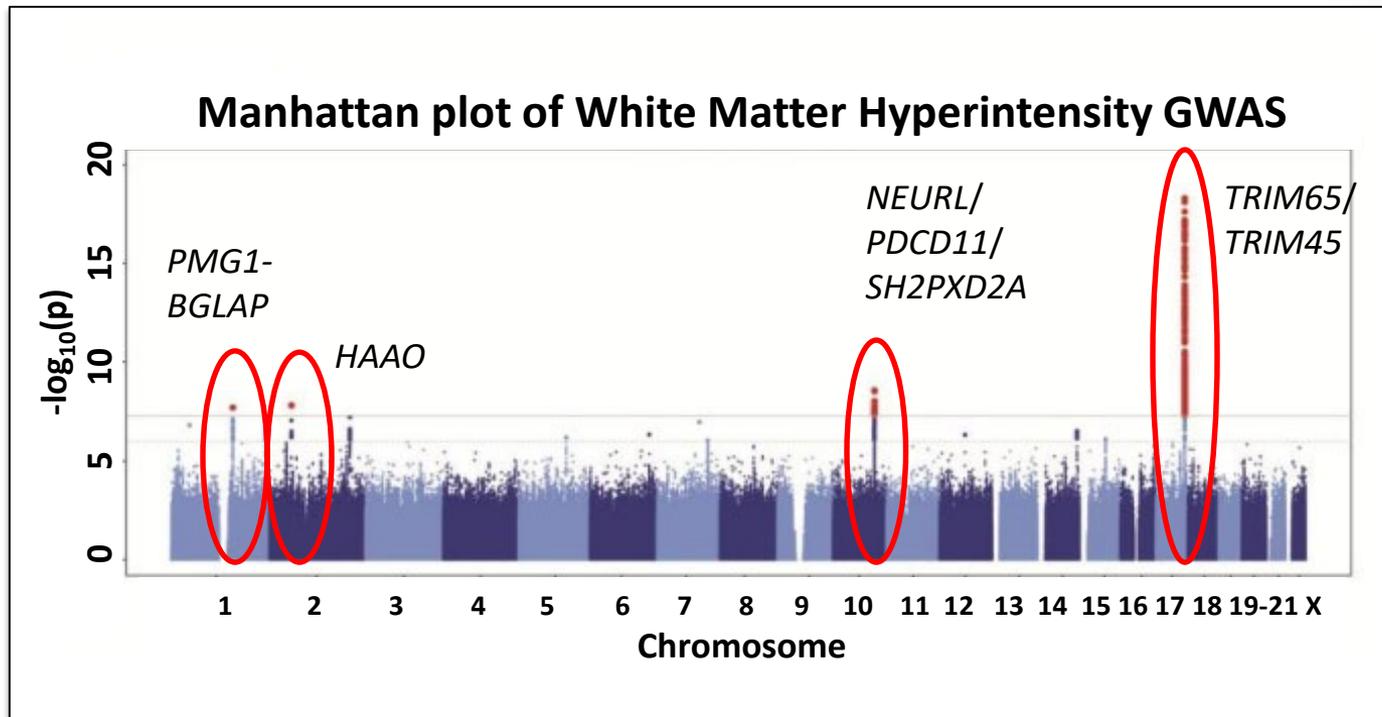
Most traits have many genes with small effects



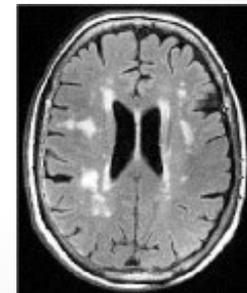
How a Genome-wide Analysis Typically Works

- Discovery sample
 - Simple OLS or Logistic Regression done millions of times
 - Find Genome-wide ($p\text{-value} < 5 \times 10^{-8}$) significant hit
 - 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



GWAS meta-analysis: strengths and limitations

- **Strengths**

- Significantly boosts power and reduces false positives (Dudbridge 2013: Power and Predictive Accuracy of Polygenic...)
- All cohorts measure the same SNPs

- **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

Typical approach for score construction

ARTICLES

Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared

The Wellcome Trust Case Control... There is increasing evidence that genetic identification can predict common diseases and a shared set of ~3,000... 14,000 cases of seven common diseases and 3,000 shared

LETTERS

Genome-wide association study identifies five new schizophrenia loci

The Schizophrenia Psychiatric Genome-Wide Association Study (Schizophrenia PGC) has identified five new schizophrenia loci... We examined the role of common genetic variants in schizophrenia in a genome-wide association study... 5 new schizophrenia loci

ARTICLES

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

Educational attainment is strongly influenced by social and other environmental factors, but genetic factors are thought to account for about 10% of the variance across individuals. Here we report a genome-wide association study (GWAS) of educational attainment that includes over 1 million individuals... 74 loci associated with educational attainment

LETTER

Genome-wide association study identifies 249,796 individuals reveal 18 new loci associated with body mass index

Body mass index (BMI) is a measure of body fat based on height and weight that is widely used to assess whether an individual is underweight, overweight or obese... 18 new loci associated with body mass index

ARTICLES

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

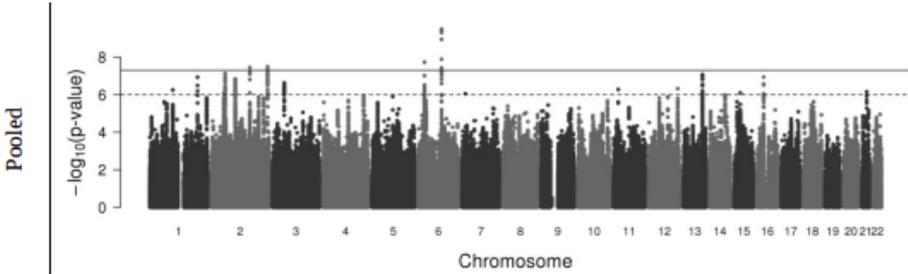
Educational attainment is strongly influenced by social and other environmental factors, but genetic factors are thought to account for about 10% of the variance across individuals. Here we report a genome-wide association study (GWAS) of educational attainment that includes over 1 million individuals... 74 loci associated with educational attainment

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

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



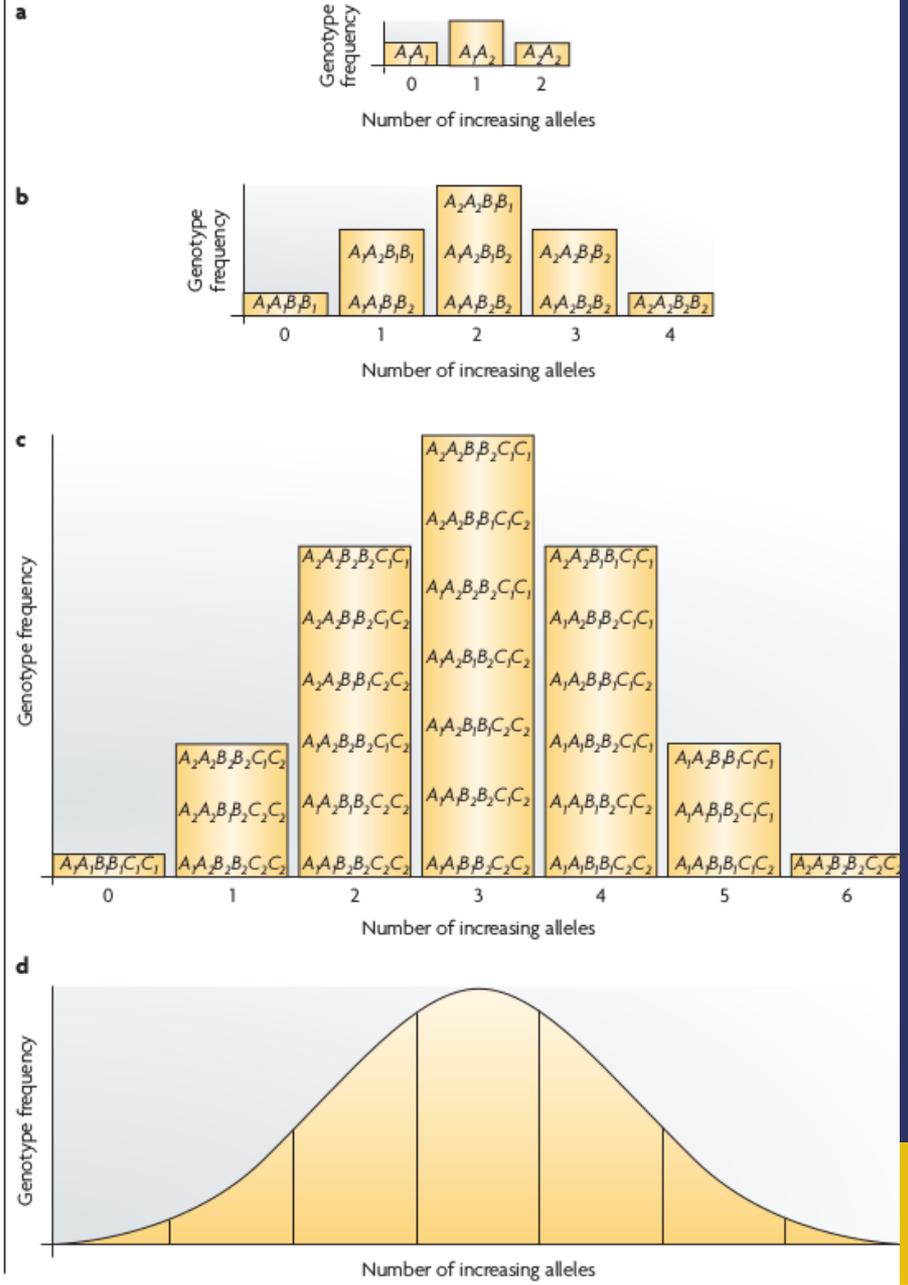
Score Construction



+

	SNP 1	SNP 2	...	SNP 1,000,000
P1	0	1	...	2
P2	1	0	...	0
P3	1	2	...	1
⋮	⋮	⋮	⋮	⋮
P1000	2	1	...	2

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



Polygenic Score Construction

Heterogeneity in polygenic scores for common human traits

Erin B Ware MPH PhD^{*1}, Lauren L Schmitz PhD¹, Jessica Faul MPH PhD¹, Arianna Gard MA², Colter Mitchell PhD¹, Jennifer A Smith MPH PhD^{1,3}, Wei Zhao PhD³, David Weir PhD¹, Sharon LR Kardia PhD³

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

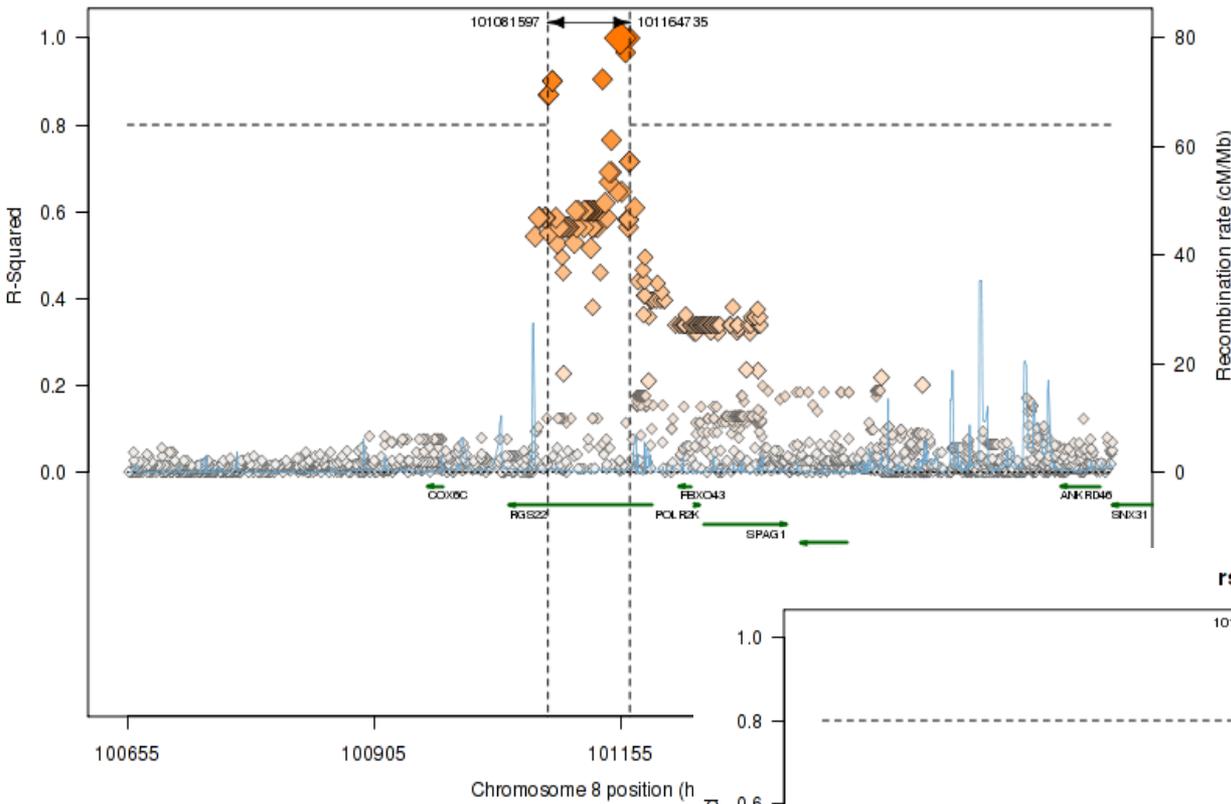
*Corresponding author ebakshis@umich.edu

Race/Ethnicity/Caste and Ancestry

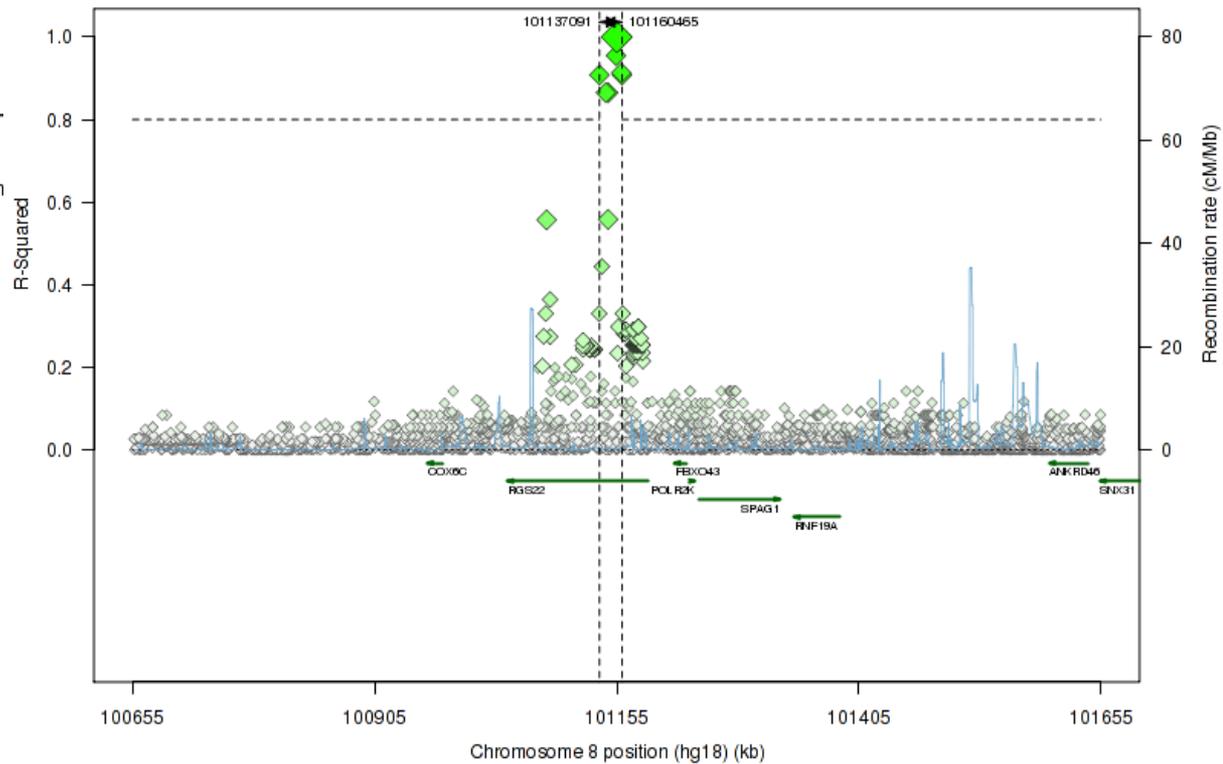
- **Race, Ethnicity, and Caste**-socially constructed and a key measure in examining social and health inequities
- **Ancestry**—variation in genetic architecture between populations
 - Used in genomic analyses
 - Results from many generations of demographic processes: migration, fertility, and mortality (which have been/are under social control)
 - Genetic variation \neq causal effect on biology/health
- The number and homogeneity of ancestries is very different by race and ethnicity

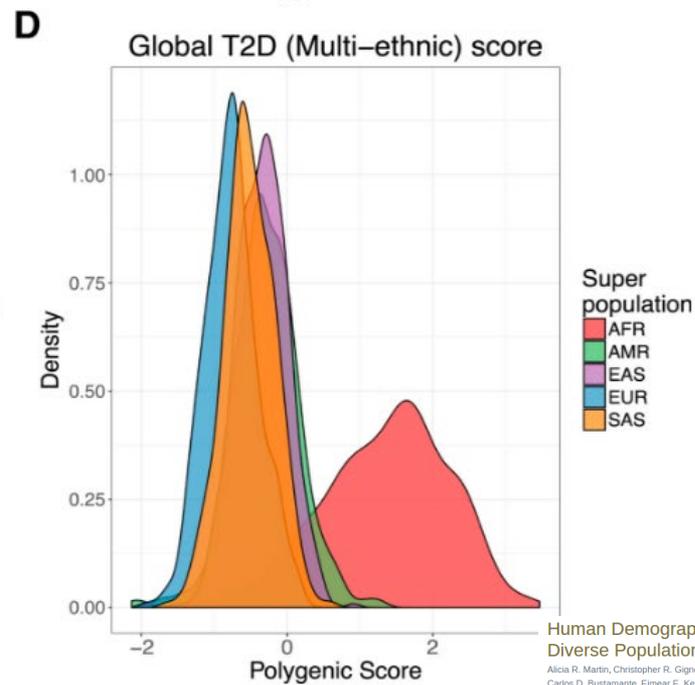
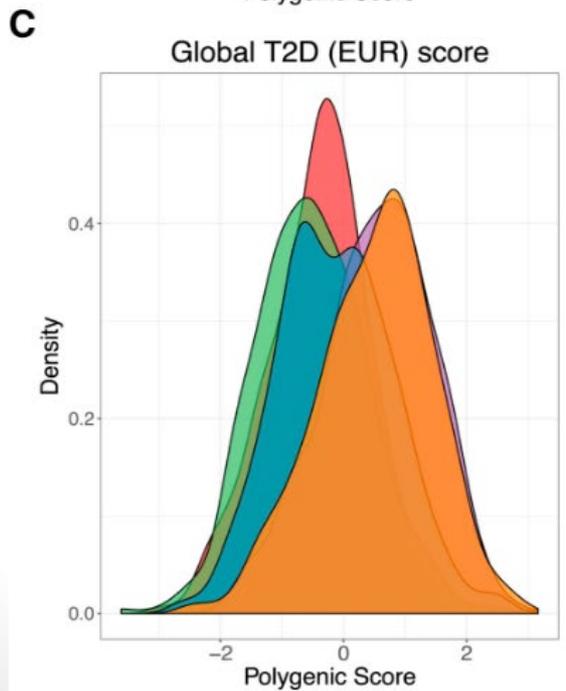
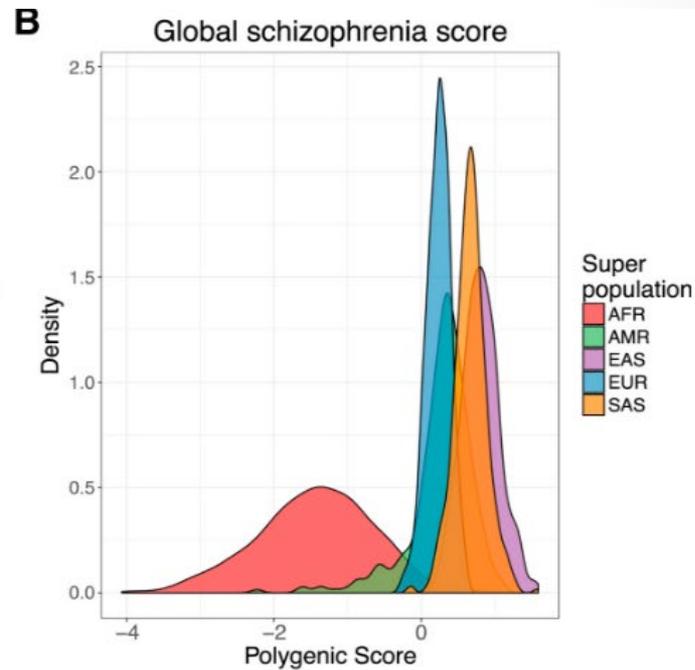
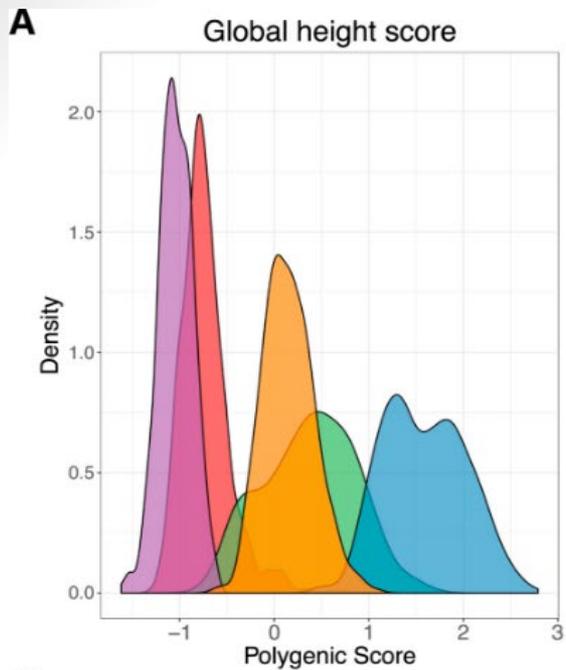
Sensitivity to ancestry

rs1380994 (CEU)



rs1380994 (YRI)





Selection into GWAS

Vast majority of GWAS results are based on those of euro ancestry

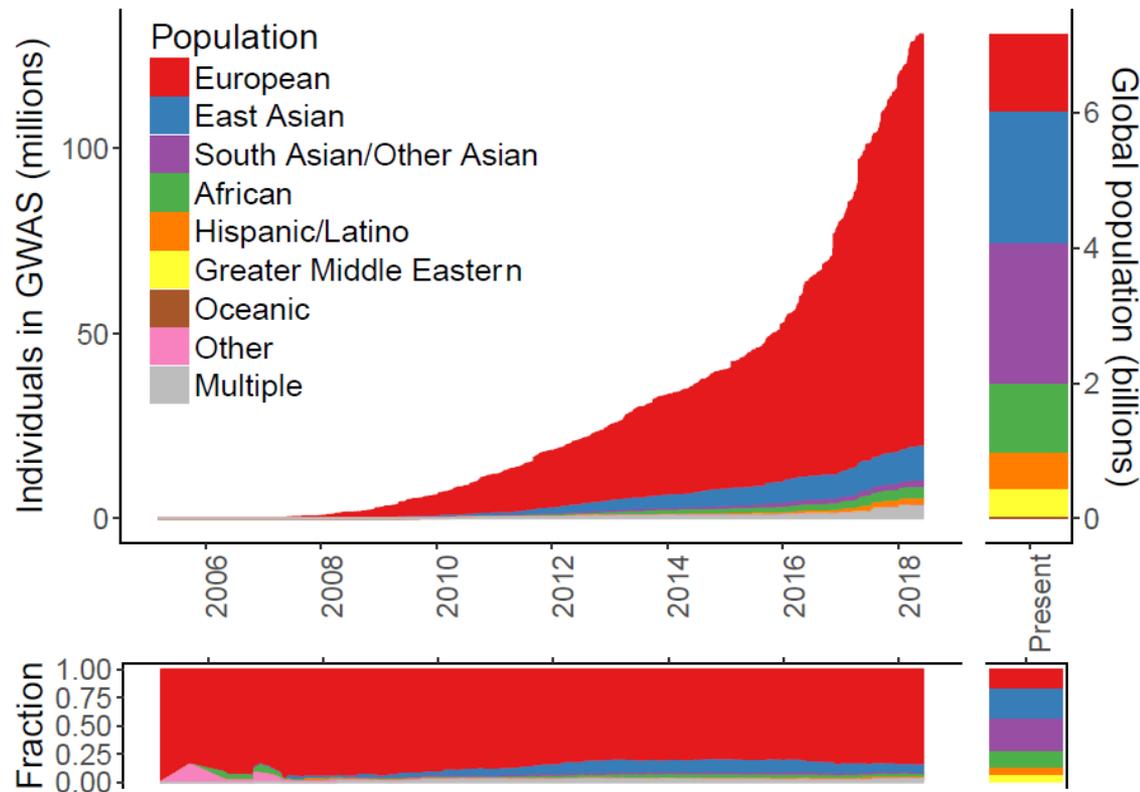
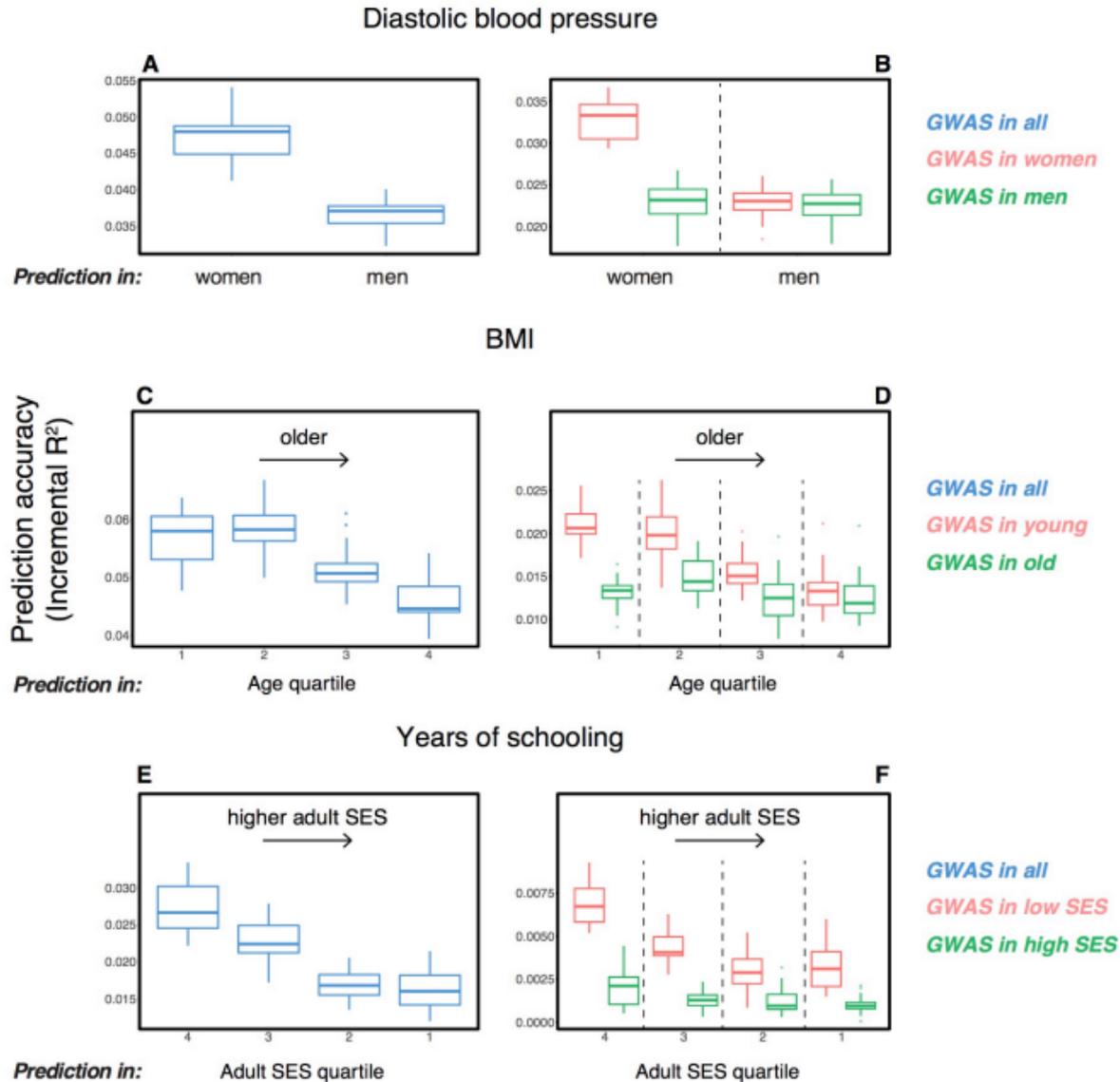


Figure 1 – Ancestry of GWAS participants over time compared to the global population. Cumulative data as reported by the GWAS catalog²³. A notable caveat is

Selection into GWAS

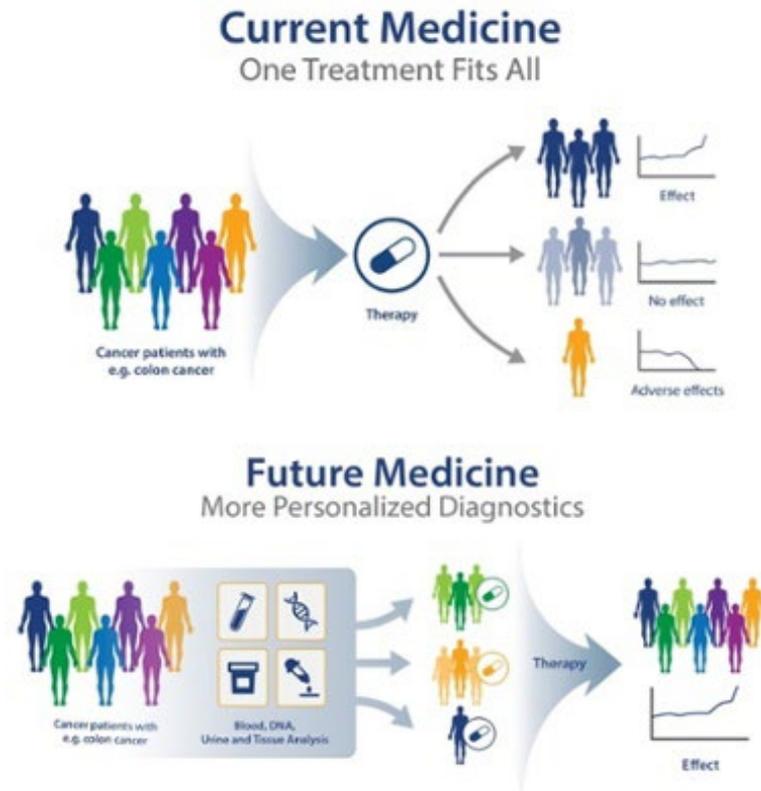
Variable prediction accuracy of polygenic scores within an ancestry group

Hakhamanesh Mostafavi^{1,2}, Arbel Harpak^{1,2}, Dalton Conley^{2,3}, Jonathan K Pritchard^{4,5} and Molly Przeworski^{1,6}



Selection into GWAS

- **Bias**-Ascertainment and assay bias leads to algorithm bias leads to anchor and confirmation bias



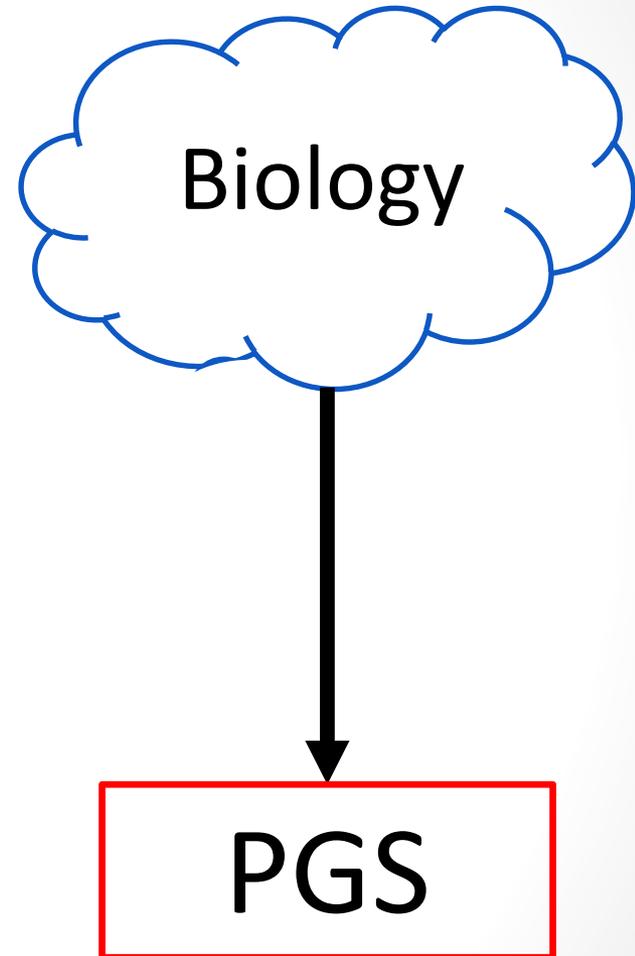
Selected Recent Reviews and Commentaries

- 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)



Practicalities

- GWAS of many traits
- Larger and larger effects
- Correlation within ancestry and improving representation
- Working to improve interpretation of GWAS and PGS

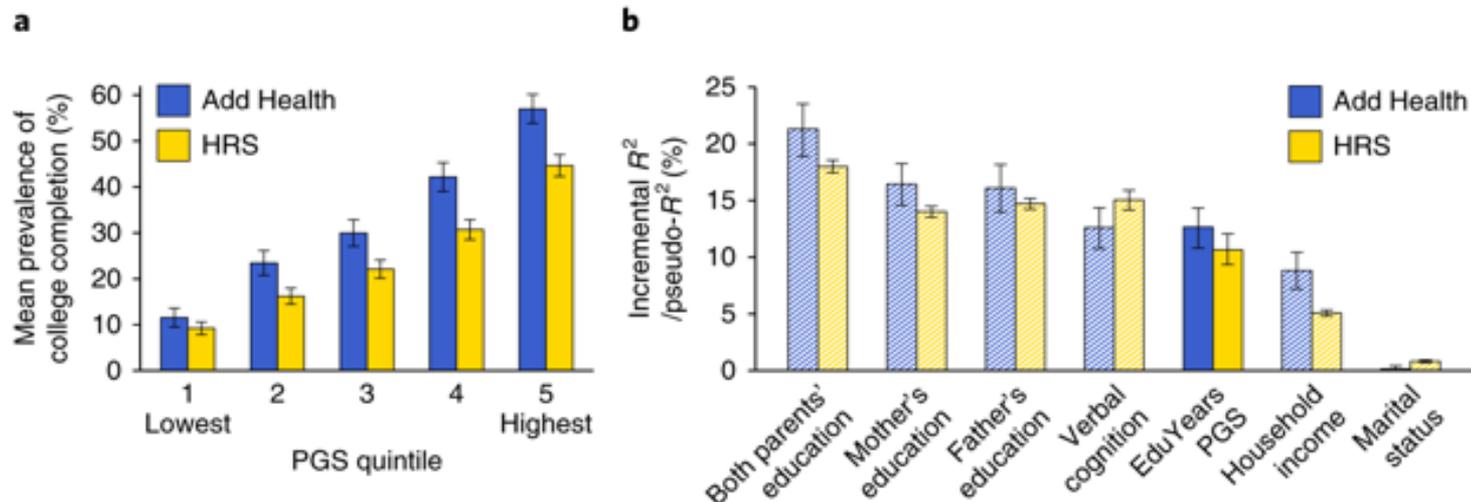


GWAS of many traits

- **Bias**-Ascertainment and assay bias leads to algorithm bias leads to anchor and confirmation bias



Large predictive effects



- **Ancestry**-within ancestry comparisons are possible but explanatory power is cut by at least 50%.
 - More multi-ethnic and larger non-European studies are in progress (see CVFS and LASI)

Genomics of Behavior needs social scientists

- Expertise on measurement and modeling
- Generalizability
- Rigorous interpretation of PGS
- Biological determinism and Eugenics

Conclusions

- **Potential**—why the idea of a genetic measure is important for social & demographic research
- **Pitfalls**—How PGS operationalize genetic measures and concerns about them
- **Practicalities**—What is possible now and likely in the future

