Genetics and Demographic Behavior

Colter Mitchell (cmsm@umich.edu)

January 15, 2025

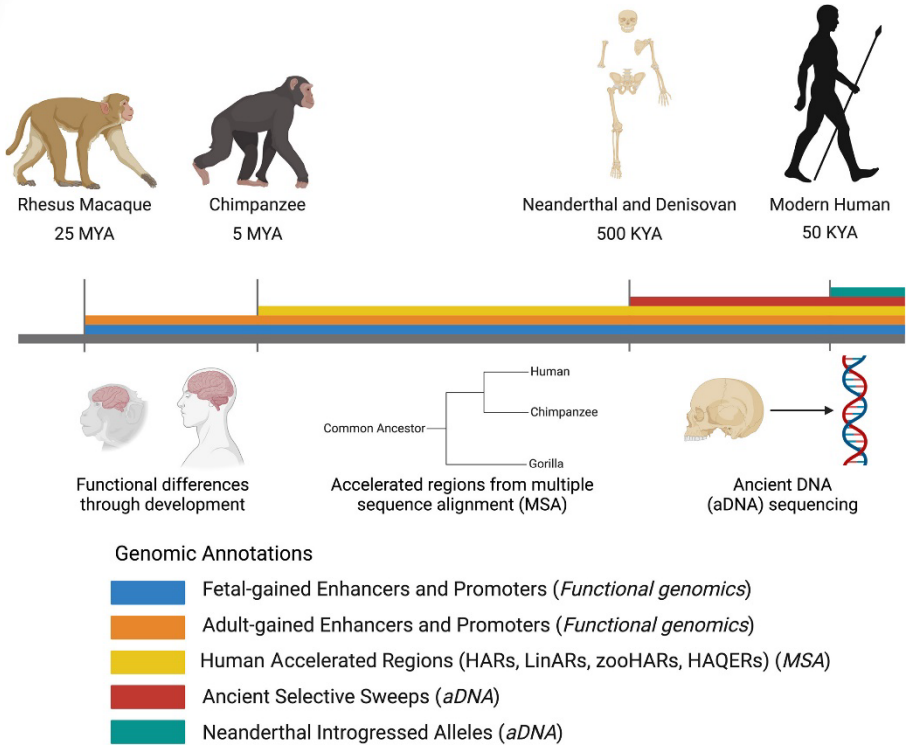
Chitwan Valley Family Study Webinar Series

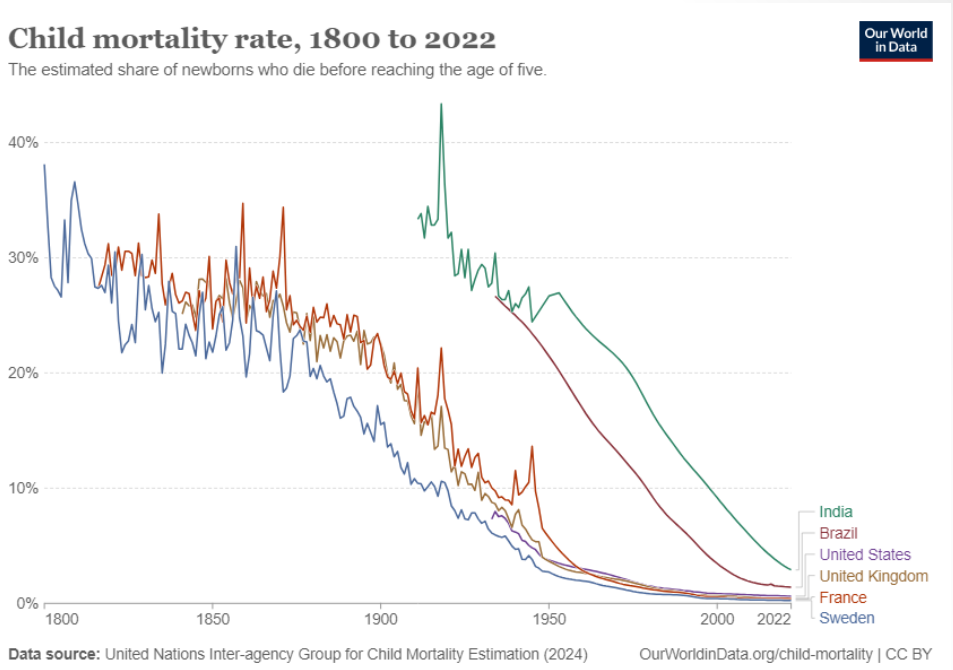
Outline

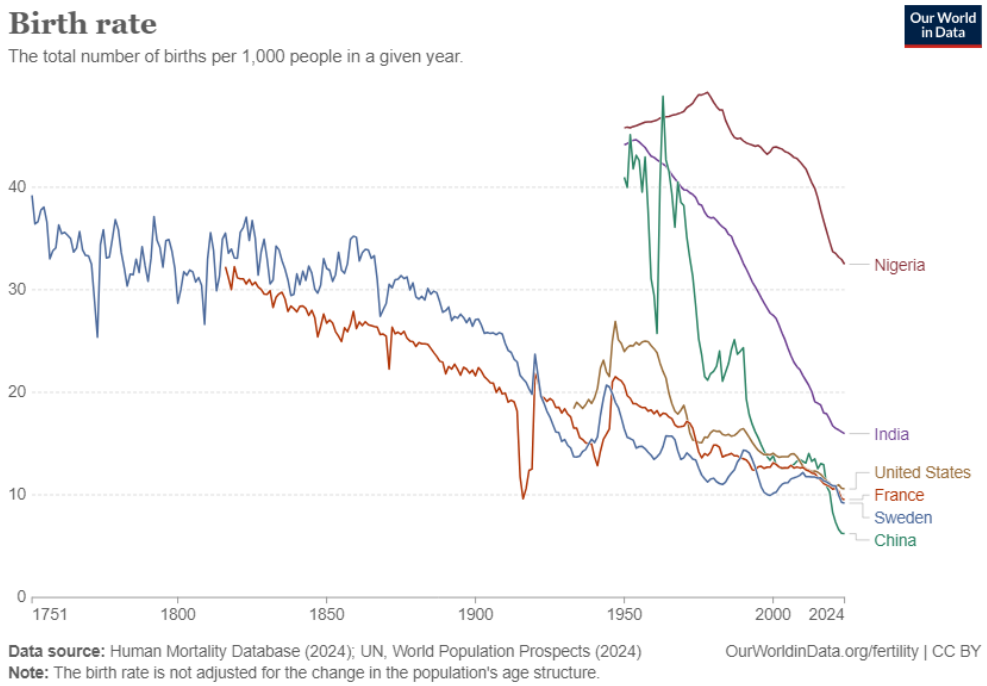
- Key issues for understanding linking genetics and demographic behavior
- The Genetics of
  - Fertility
  - Mortality
  - Migration
  - Brief reminders of some genetic analytic techniques
  - Focus is mostly on recent polygenic associations
- CVFS genetic data description
- Potential in CVFS and beyond

KEY ISSUES

Demographic Change







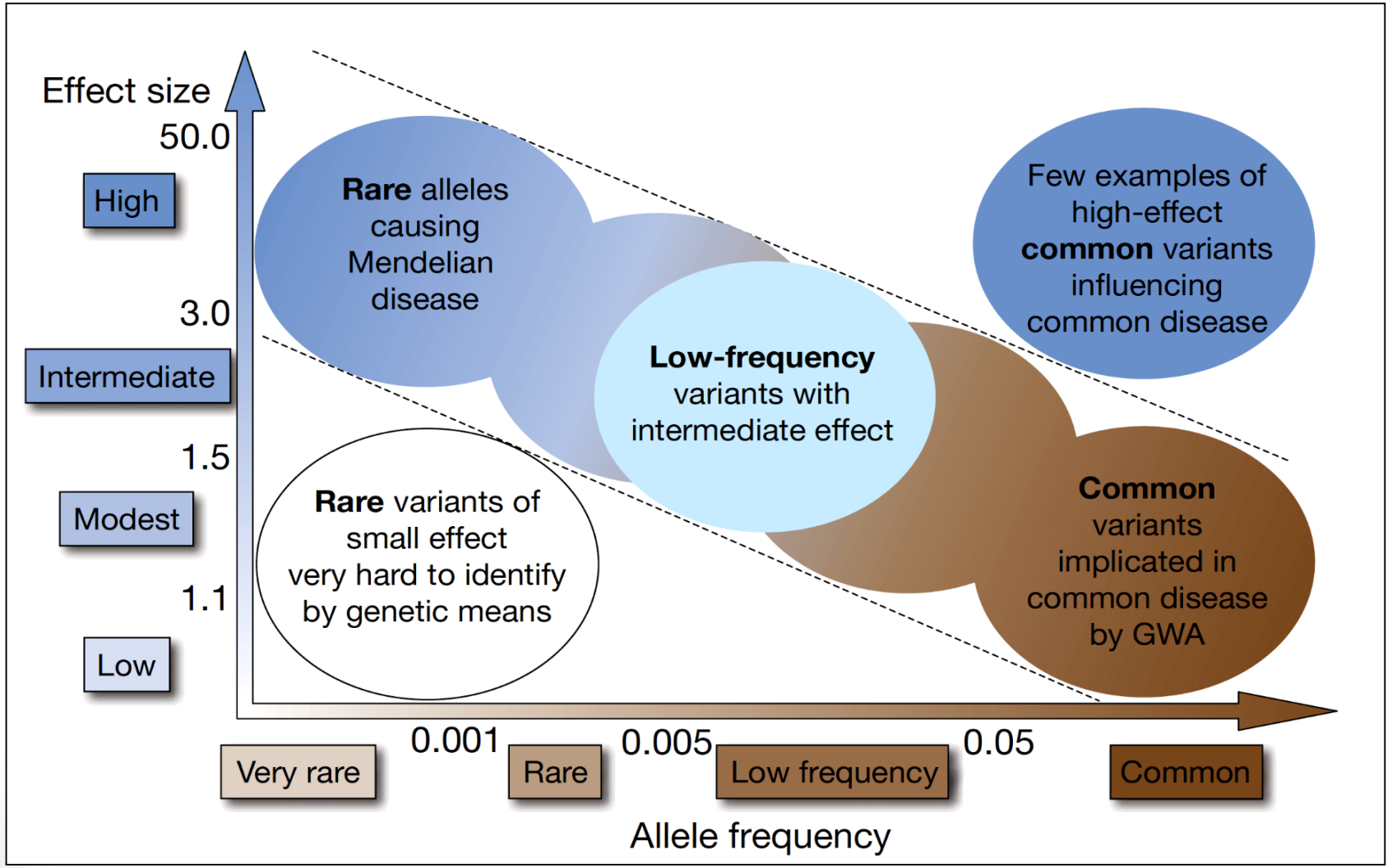
- Most within species changes requires thousands of years
- Powerful rare mutations still require many generations (hundreds) years

Kun, E., Sohail, M. and Narasimhan, V.M., 2025. *Cell Genomics*, *5*(1).; Subbaraman, N. *Nature* (2012).

So then why genetics?

- Most recent demography change **cannot** be through genetics
- Yet genetics still influences demographic outcomes
- Constant genetic diversity (i.e. genetics related to lower fertility, mortality)
- New historical environment—new genetic effects?
- Genes are correlated with environments (often through social factors)
- Highlight the social with genetics
  - Those dubious about social factors

Most traits have many genes with small effects



- Large literatures on rare genetic effects
- More recent focus on polygenic effects (primary focus)

Gene-environment correlation

- **Active**- people choose or create environment based on genetic tendencies

- **Evocative**-environment responds to a phenotype caused by genes (e.g. facial structure, height, skin pigmentation, temperament, etc.).

- **Passive**-genes and environment are passed together

Mitchell, et al. 2013. *AJPH*

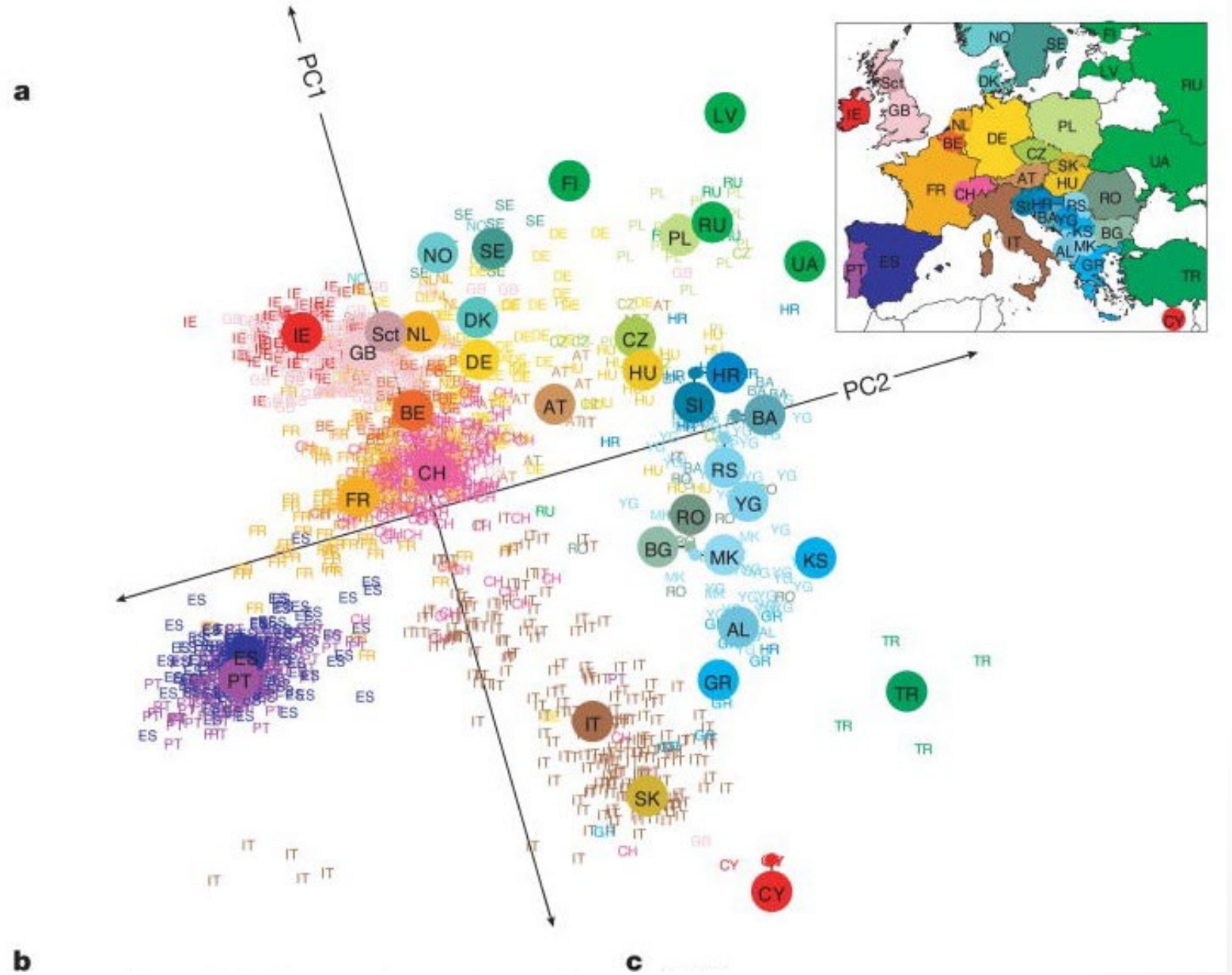
Race/Ethnicity/Caste and Ancestry

- **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

Ancestry Population Stratification



Novembre, 2008, *Nature*

FERTILITY

Genetics of Infertility

- Around 7% of couples
- Large literature with many strong genetic effects (i.e. rare variants)
- Chromosomal abnormalities: XXY and XXX syndromes, Y and X microdeletions
- Over 20 single gene (typically different by sex) effects
- Novel or stronger genetic effects due to delaying fertility is being studied

Qiao, J., et al. (2018). *Nature Reviews Genetics;* Laissue, P. (2019). *Reproduction;* Krausz, C., & Riera-Escamilla, A. (2018). *Nature Reviews Urology*

Polygenic Explorations

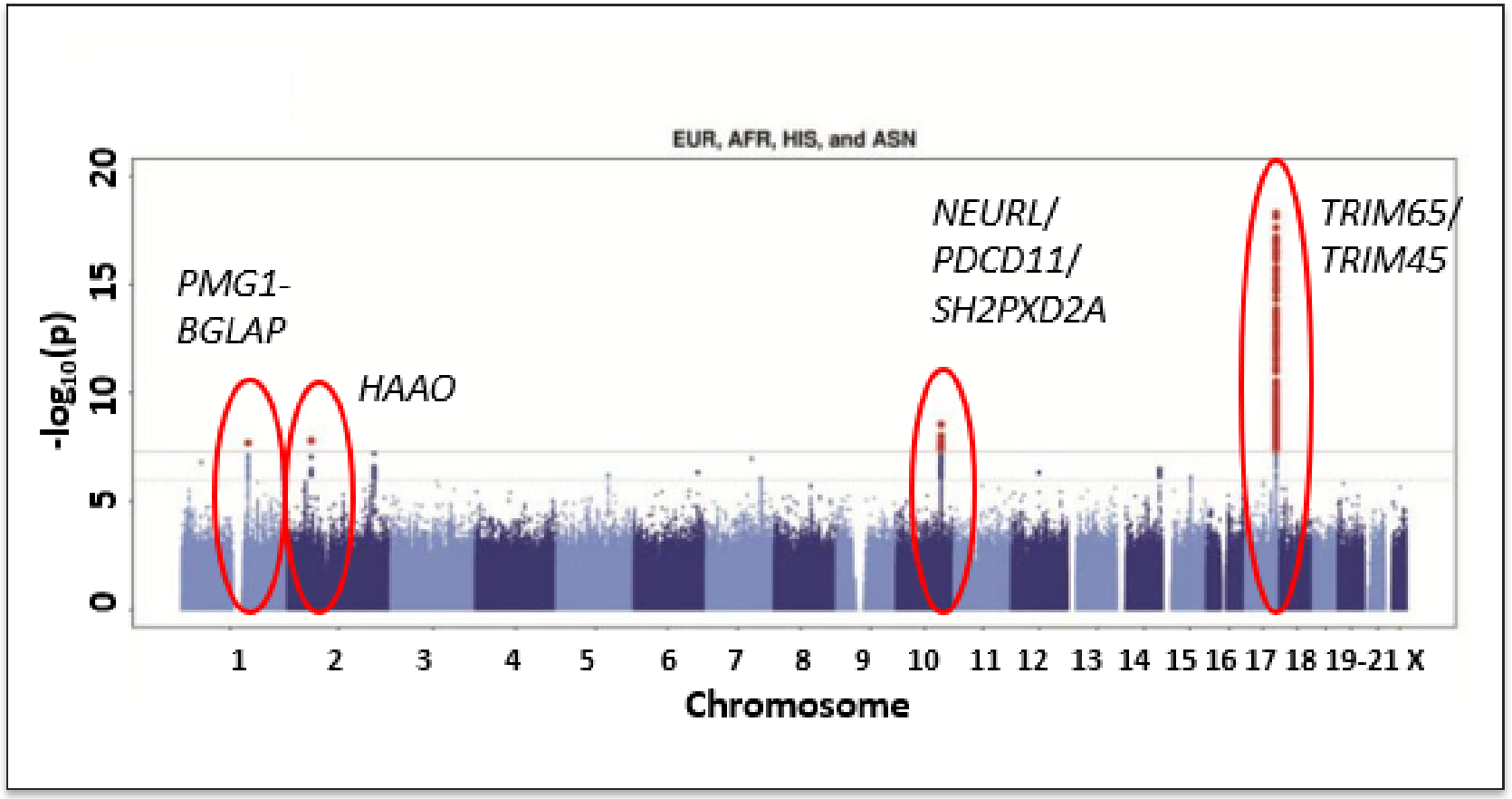
- High throughput techniques have allowed for more genome-wide analyses
- Genome-wide Association Studies (see next slide) identify genetic variants associated with complex traits (such as fertility) by scanning a strategic sub-sample of genetic variants
- Fertility outcomes, such as the number of children ever born, age at first birth, and age at menarche/menopause (not covered here), are influenced by genetic and environmental factors

Barban, N., et al. (2016). *Nature Genetics;* Tropf, F. C., et al. (2017). *PLOS ONE;* Mills, M. C., et al. (2018). *Human Nature;* Hwang, J. Y., et al. (2017). *Genetics;* Mills, M. C., (2021). *Nature human behaviour*.

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 (100Ks of people)
- Replication Sample
  - Smaller n, but often still large
  - Sometimes a meta-analysis of studies not in discovery

Visualizing GWAS results: Manhattan plots



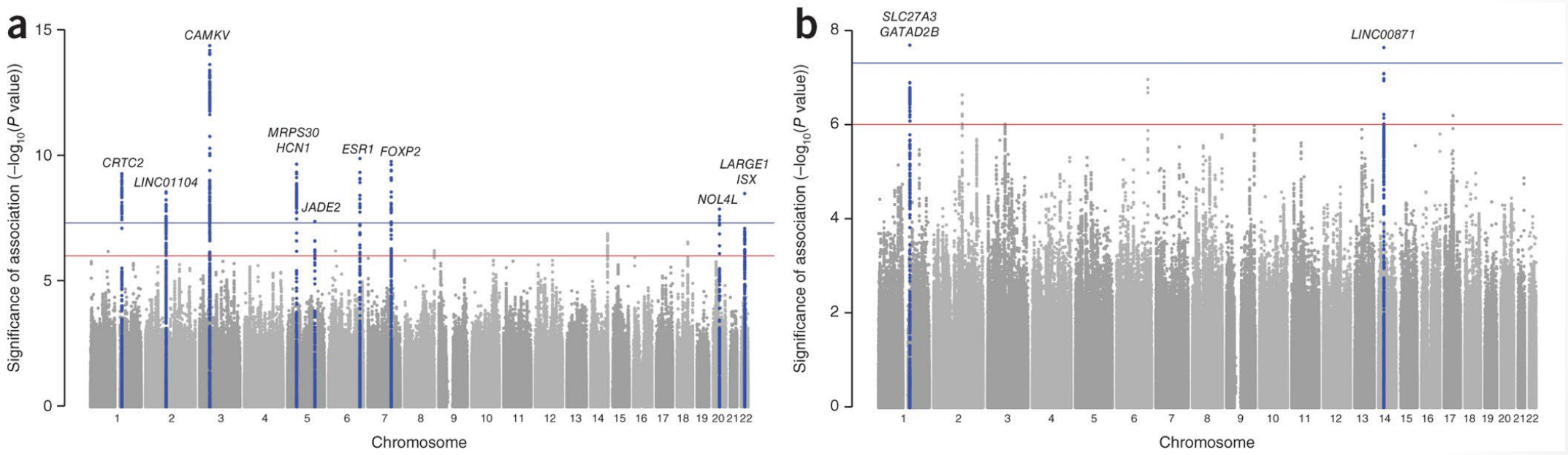
- X-axis = chromosomal location
- Y-axis = the **observed** p-values (higher=smaller p-value)

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

Fertility GWAS

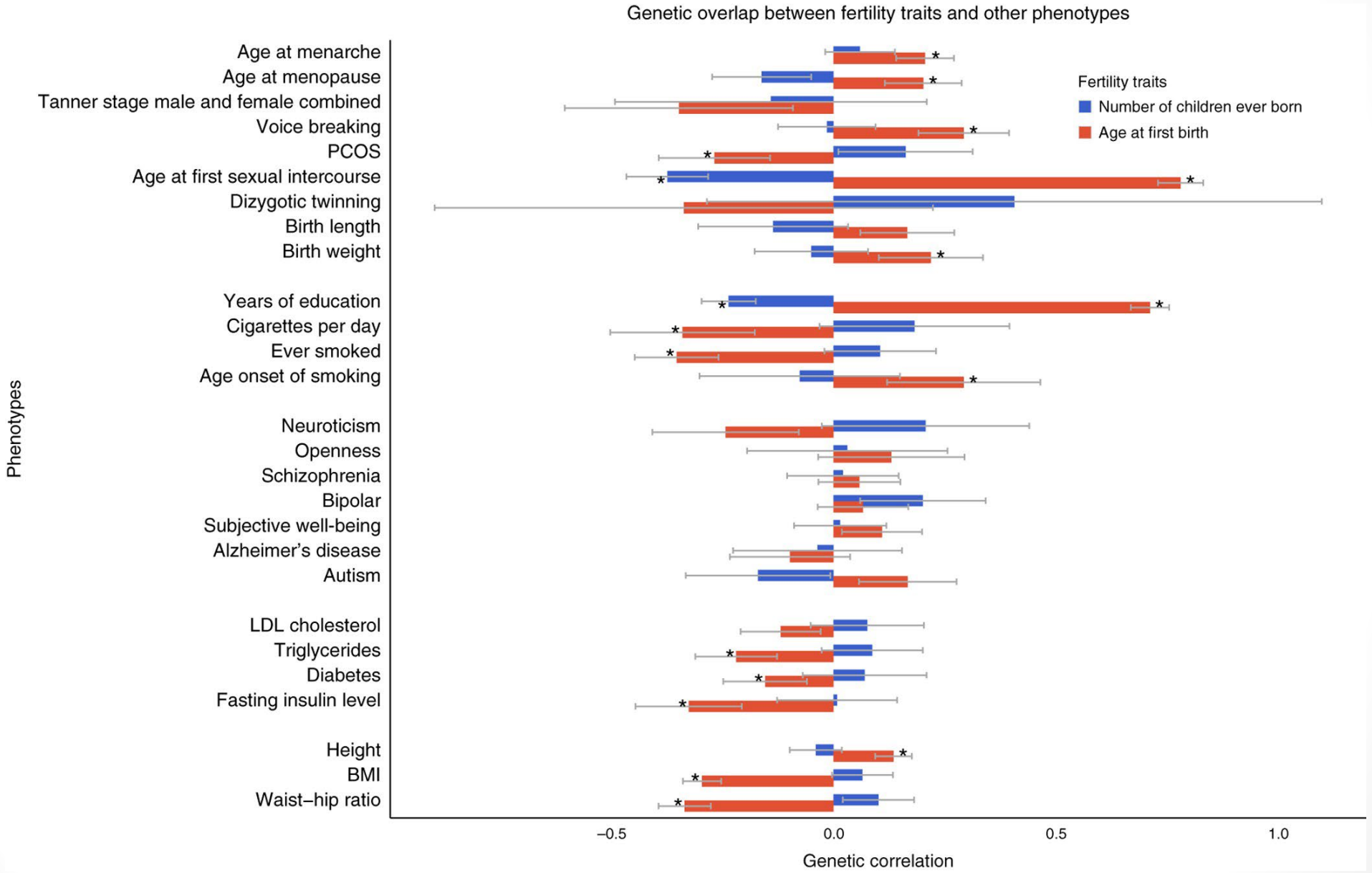
- Barban, N., et al. (2016). Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nature Genetics*
- Age at First Birth (AFB) and Number of Children Ever Born
- 62 European Ancestry cohorts
- N=251,151 individuals for AFB
- N=343,072 individuals for NEB
- 10 new gene-regions discovered

GWAS of Age 1st Birth (L) and Number of Children Ever Born(R)



Solid blue line indicates the threshold for genome-wide significance ($P < 5 \times 10^{-8}$), and the red line represents the threshold for suggestive hits ($P < 5 \times 10^{-6}$).

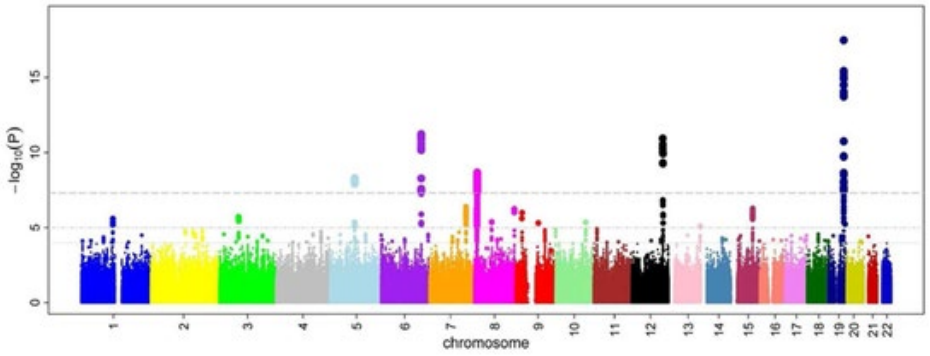
Genetic Correlations with other traits



Typical approach for Polygenic 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

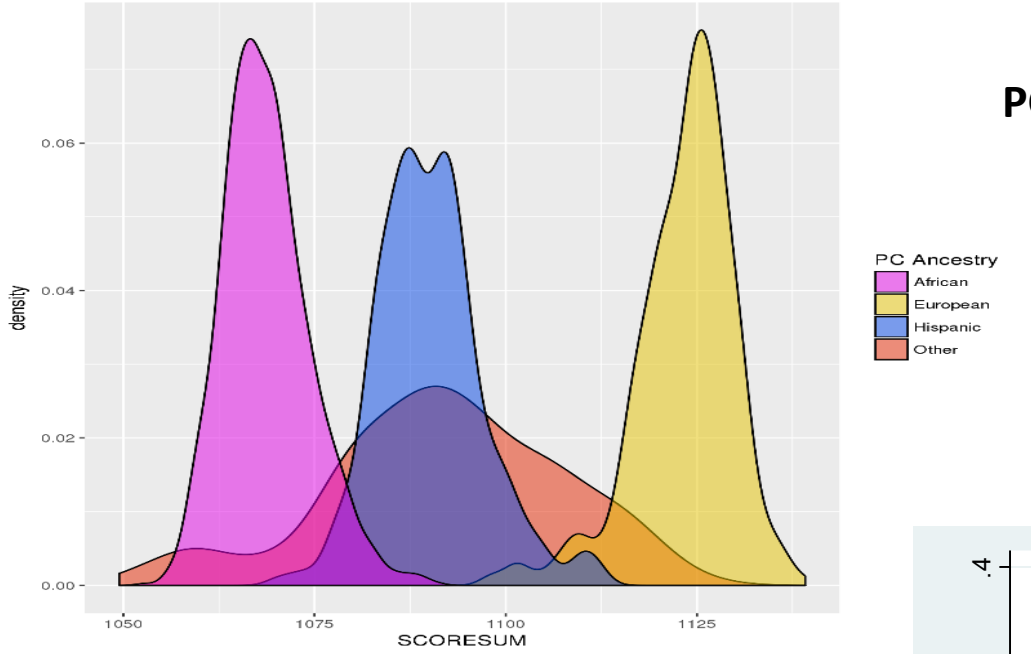


Heterogeneity in polygenic scores for common human traits

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 Portability by Ancestry

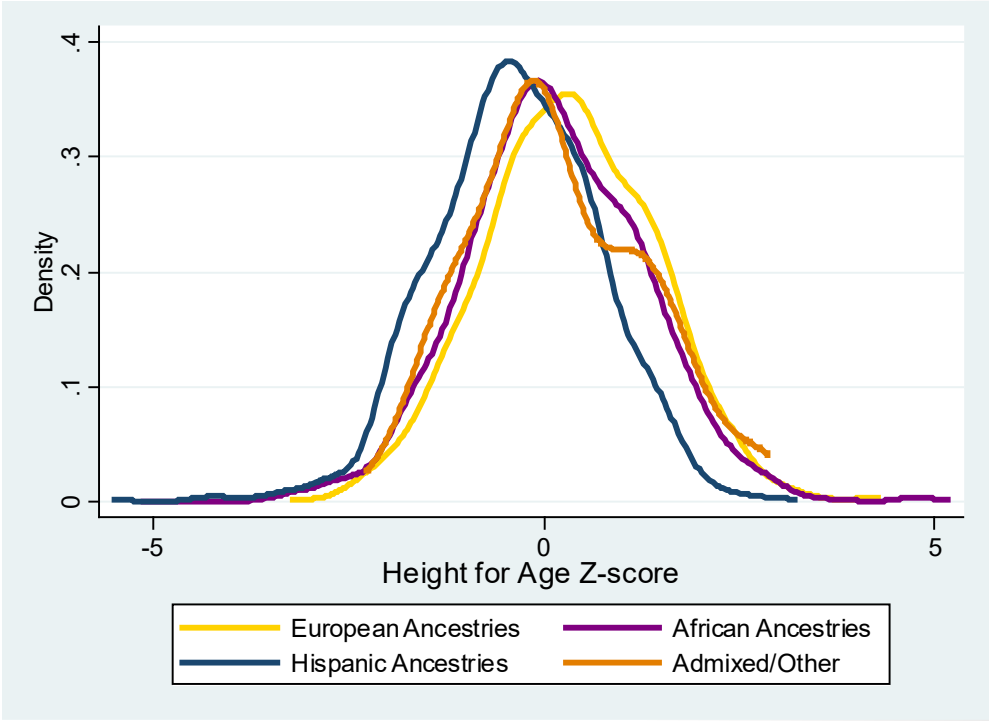


PGS- Expected difference in height:

- EA 2.9 inches taller than HA and Other
- EA 4.3 inches taller than AA
- HA 1.4 inches taller than AA

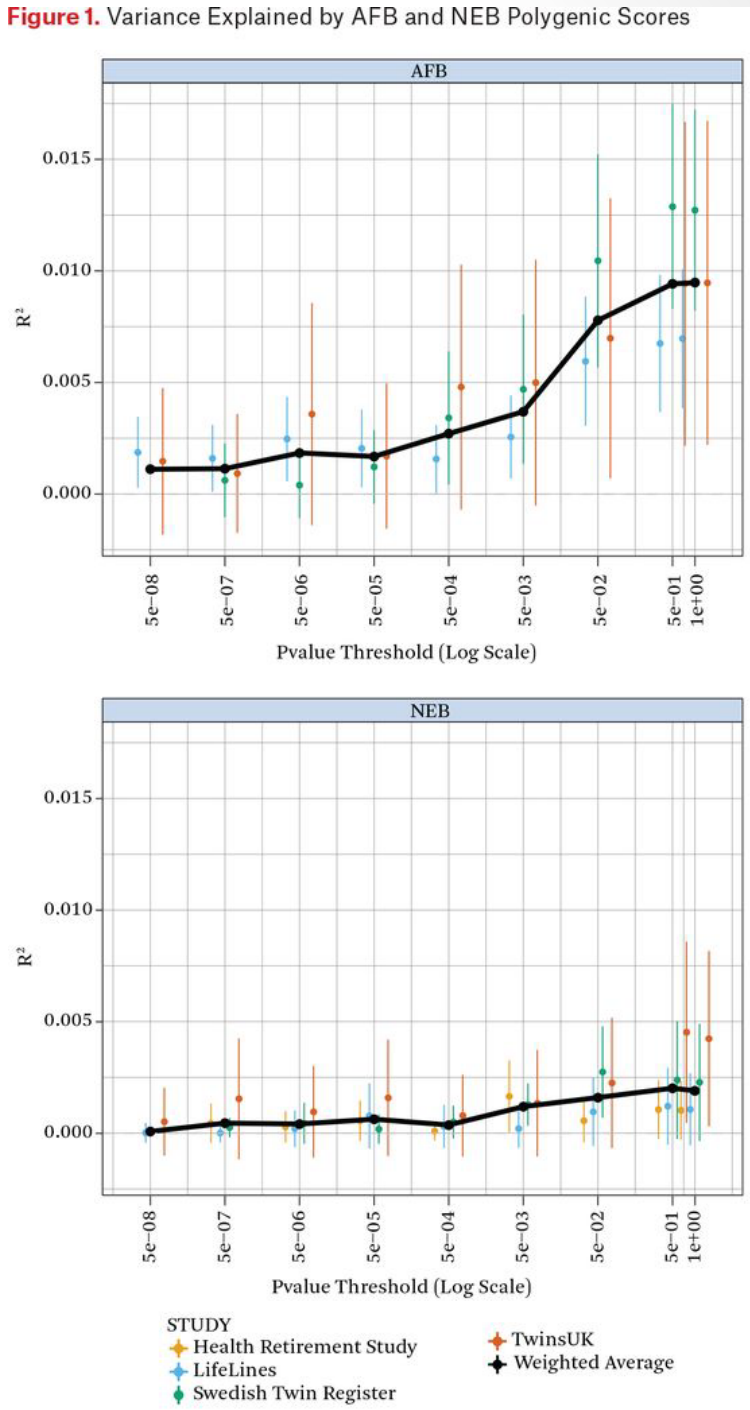
Real difference in height (z-score):

- HA not significantly different from AA and Other
- EA not significantly taller than AA or Other
- EA 0.8 inch taller than HA



Polygenic Scores of Fertility

- Predict other Fertility traits
  - Small variance explained
- Associated with many other traits—obesity, education, extraversion, puberty and menopause timing
- Within family designs (siblings) show it is likely causal but very small
- **Need more non-EA studies and rGE**



Mills, et al. 2018. Russell Sage Foundation Journal

MORTALITY

The Genetics of Mortality and Lifespan

- Lifespan and Mortality is influenced by a complex interplay of genetic, environmental, and lifestyle factors.
- Large-scale genome-wide association studies (GWAS) have identified key loci linked to mortality.
  - Notable genes include APOE, which impacts lipid metabolism, and FOXO3, linked to cellular stress resistance.
- Genetic predispositions can partly explain variations in lifespan across populations.

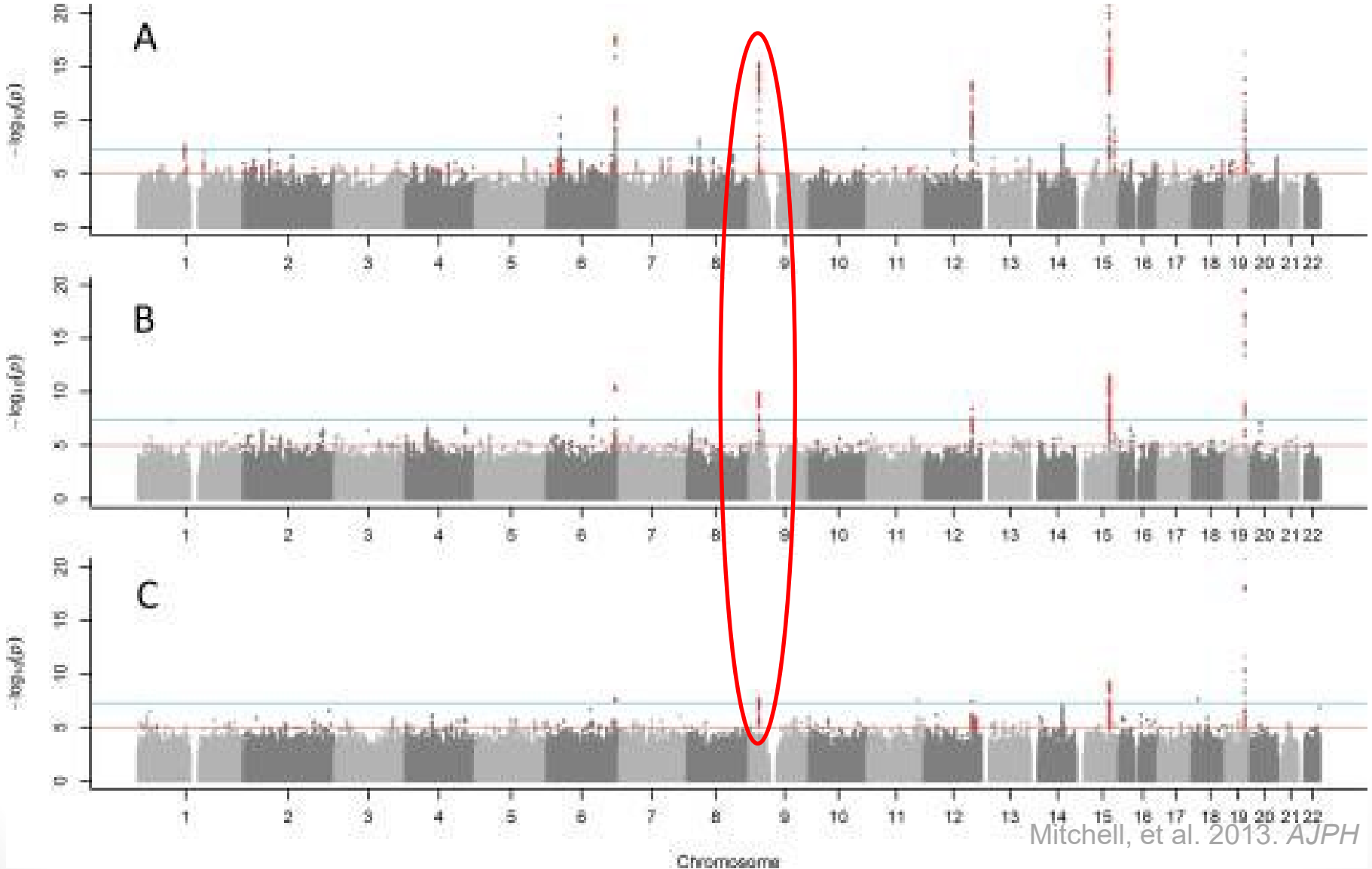
Timmers, P. R., et al. (2019). *eLife*; Codd, V., et al. (2018). *Nature Communications*; Deelen, Joris, et al. (2019) *Nature communications* 10.1 (2019): 3669.

GWAS of Lifespan (25 new genes)

- Pilling, Luke C., et al. "Human longevity: 25 genetic loci associated in 389,166 UK biobank participants." *Aging* 9.12 (2017): 2504.
- European descent
  - UK biobank response rate ~5%
- Data recorded on parents' current ages or parents' ages of death (n=181,048 had one or more living parents)
- A) Combined Age of Parents, B) Combined age of death of parents, C) Parents lived to 90% percentile (90 for Female 87 for Male)
- Discovered several new loci for biological exploration
- Polygenic scores predicted mortality in multiple longitudinal cohorts

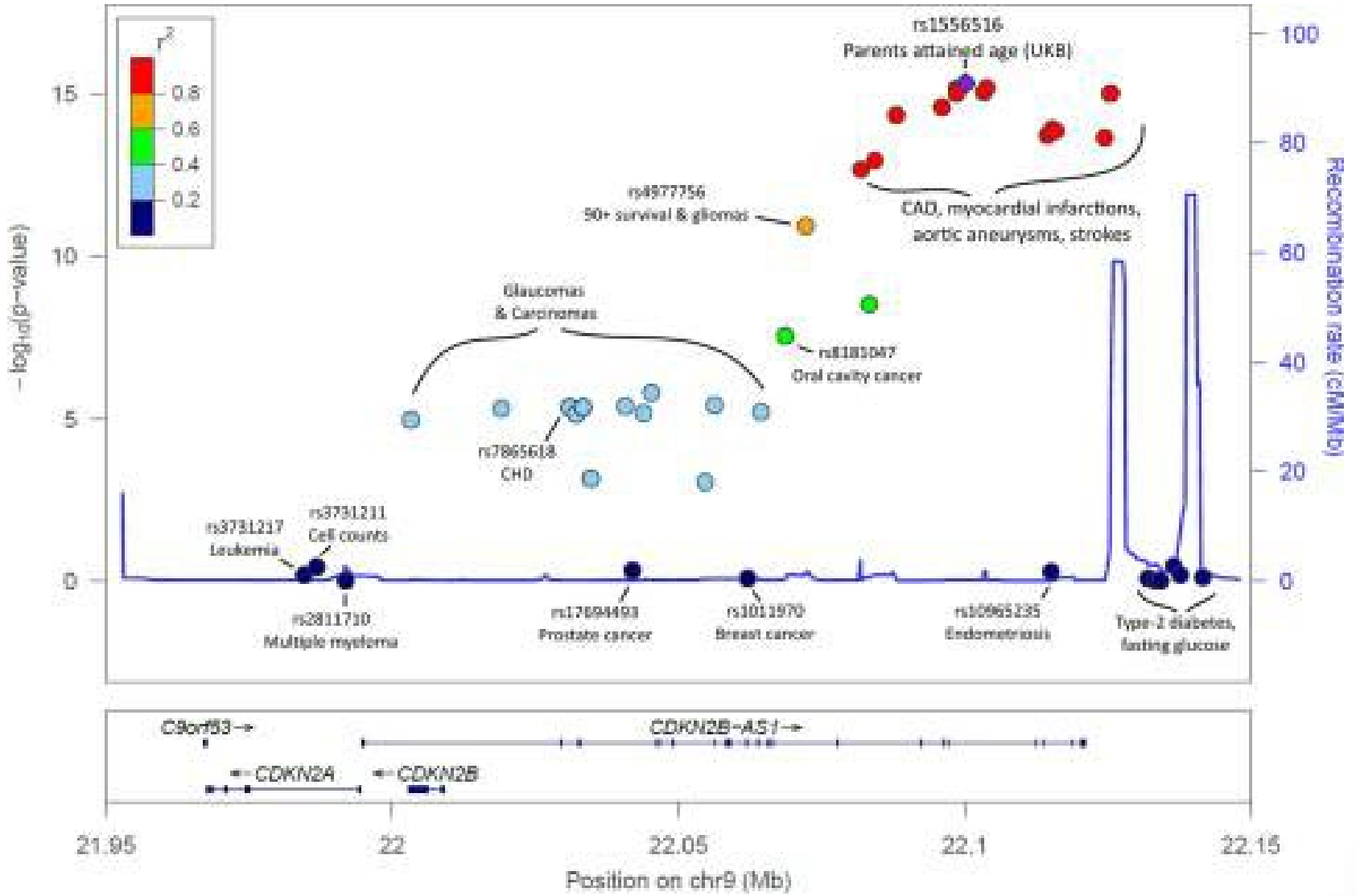
GWAS of Lifespan

- A) Combined Age, B) Combined Age of Death, D) Parents Lived to 90%



Mitchell, et al. 2013. *AJPH*

Locus Zoom Plot



Pilling et al 2017 *Aging*

Disease specific analyses

- Much larger literature on the genetic association with disease-specific mortality, such as cancer, neurological diseases, and cardiovascular disease.
- Advancements in genetics hold promise for targeted therapies and personalized medicine approaches to enhance longevity.
- But also, are likely to lead to health disparities without more representative data.
- Ethical considerations include potential discrimination based on genetic predispositions to mortality.

Kuchenbaecker, K. B., et al. (2017). Risks of breast, ovarian, and contralateral breast cancer…. *JAMA*; Inouye, M., et al. (2018). Genomic risk prediction of coronary artery disease…. *Journal of the American College of Cardiology*.; Martin, Alicia R., et al. (2019) Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature genetics*.

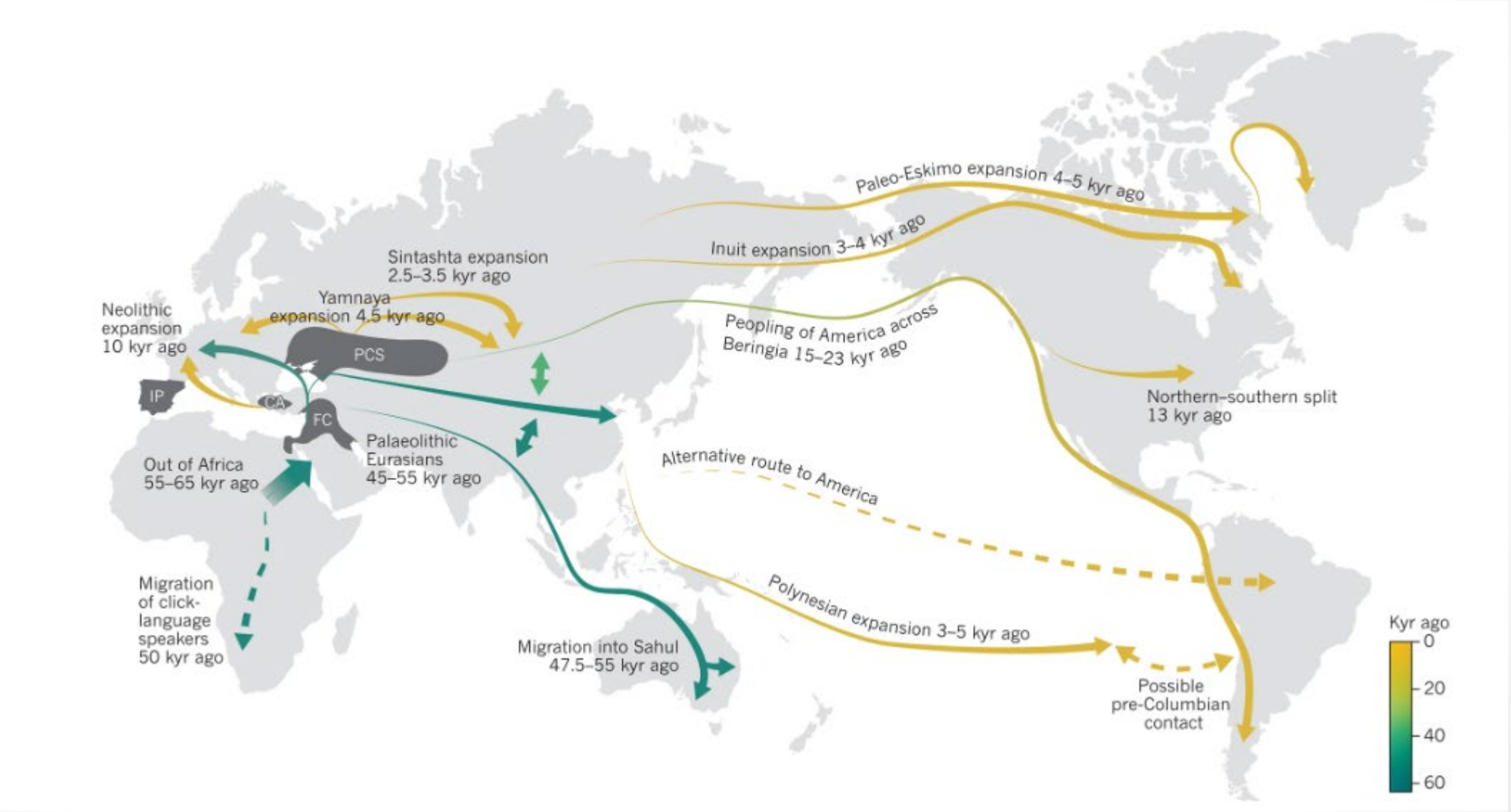
MIGRATION

Genetics reveals historical migration patterns

- Migration is a key factor in human evolution, shaping genetic diversity and population structure.
- Majority of genetic studies of migration investigate how historical movements of populations have influenced modern genetic variation.
- Recent advancements in genomic technologies have enabled detailed studies of human migration patterns and their genetic impacts.
- “Recent” migrations (last 1000 years) continue to impact genetic diversity and population structure.
- The focus is descriptive not causal

Nielsen, R., et al. (2018). Tracing the peopling of the world through genomics. *Nature*.

Major migrations inferred through genetics



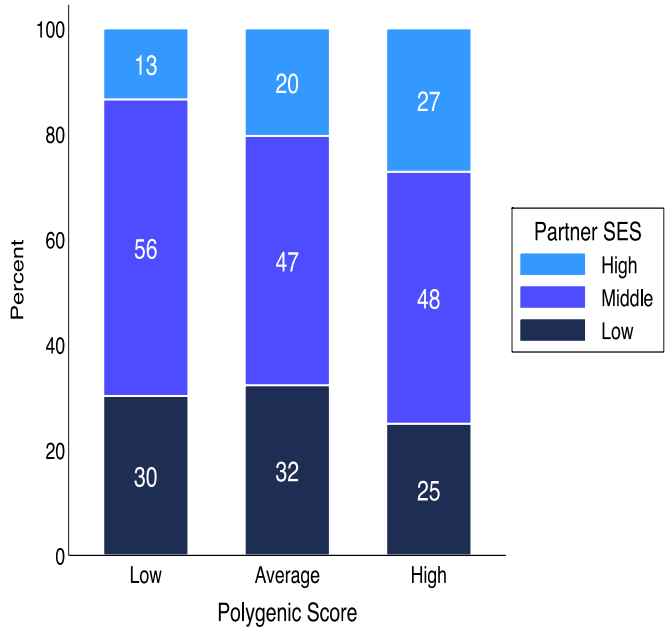
Environmental Selection

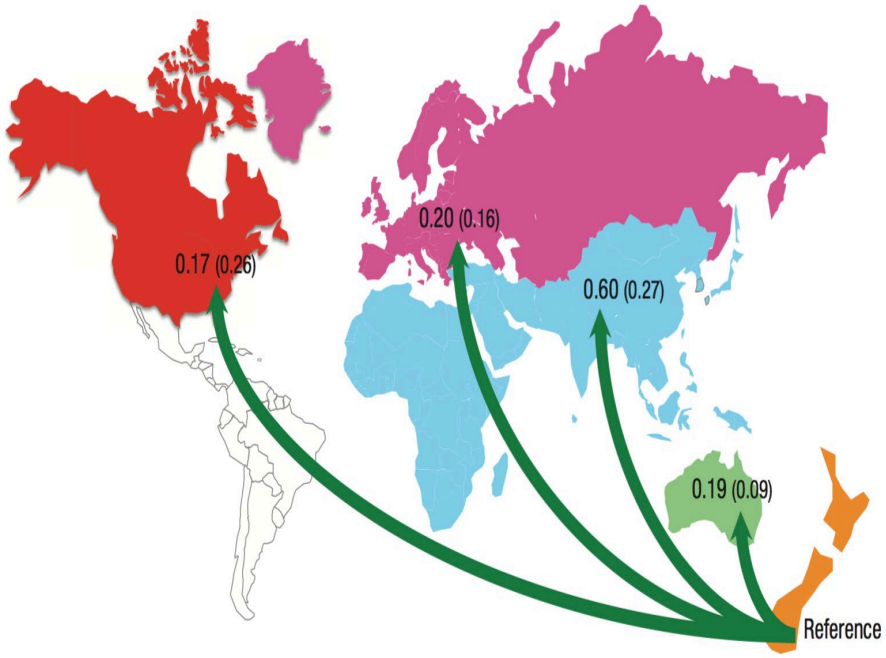
- Migration has facilitated evolutionary adaptations in diverse environments
  - Genetic adaptations to high-altitude, diet, and infectious diseases are influenced by historical migration patterns.
  - Typically takes several generations to manifest

- Less focus on what genetic traits lead to migration (next slide)
- No focus on evocative vs active gene-environment correlation
- Potential for exploration of genetics of different traits predicting migration
- Difficult to conduct a GWAS of migration without more non-EA data or a limited definition of migration (i.e. moves within country)

Huerta-Sánchez, E., & Casey, F. P. (2019). Current Biology.

Active rGE?





Dunedin Multidisciplinary Health & Development Research Unit

DMHDRU

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

Psychological Science
1–16

DOI: 10.1177/0956797616643070
pss.sagepub.com
SAGE

Daniel W. Belsky[1,2], Terrie E. Moffitt[3,4,5,6], David L. Corcoran[5], Benjamin Domingue[7], HonaLee Harrington[3], Sean Hogan[8], Renate Houts[3], Sandhya Ramrakha[8], Karen Sugden[3], Benjamin S. Williams[3], Richie Poulton[8], 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 3/14/16

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

CVFS 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

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

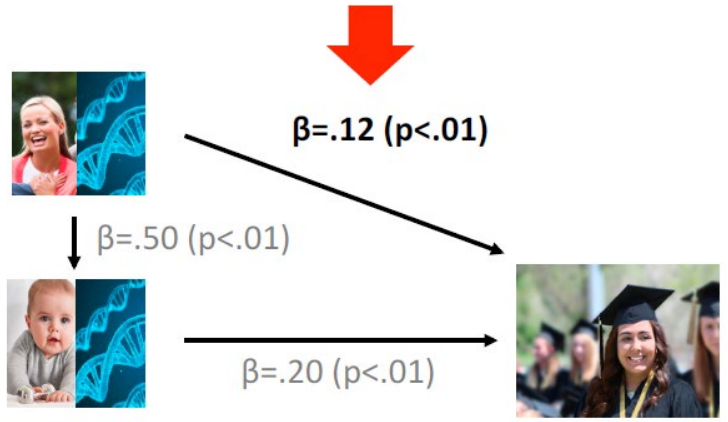
CVFS Genetic Data Family Relatedness

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

CVFS Strengths

- Prospectively measured demographic behavior
- Strong focus on fertility and migration
- Large sample size for testing of PGS
- Use of Trio (Mother-Father-Child; n= 2,000 trios) analyses-break the effect of population stratification
- Genetic nurture (parent genes directly and indirectly influencing child outcomes)
  - ~5,100 mother-child pairs
  - ~4,500 father-child pairs
- Sibling comparisons (n=~2,000) to estimate potential causal differences



Kong, Augustine, et al. "The nature of nurture: Effects of parental genotypes." *Science* 359.6374 (2018): 424-428.

Prior Lectures of Note

- https://cvfs.isr.umich.edu/webinars/past-webinars/
- Introduction to CVFS: September 15, 2021
- Migration: Oct. 27, 2021; April 20, 2022; Dec. 11, 2024
- Health: Sept 29, 2021; Jan 12, 2022; May 18 2022
- Genetic designs and methods: Dec. 1 2021; March 9, 2022; Jan. 24, 2024

Resources for Social Scientists Studying Genetics

- NIA Biomarker Network

https://biomarker-network.isr.umich.edu/

- Genomics for Social Scientists

https://genomicsworkshop.isr.umich.edu/genomics/

- Social Science Genetic Association Consortium

https://www.thessgac.org/

- Psychiatric Genomics Consortium

https://www.med.unc.edu/pgc/