Genome-wide polygenic scores and common diseases

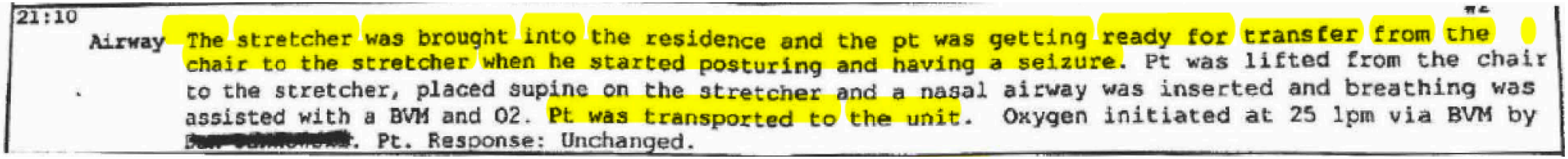
Sekar Kathiresan, M.D.
Director, Center for Genomic Medicine, MGH
Director, Cardiovascular Disease Initiative, Broad
Professor of Medicine, Harvard Medical School

May 14, 2019

Massachusetts General Hospital

Harvard Medical School

Health care scenario: 42yo male with dizziness, profuse sweating

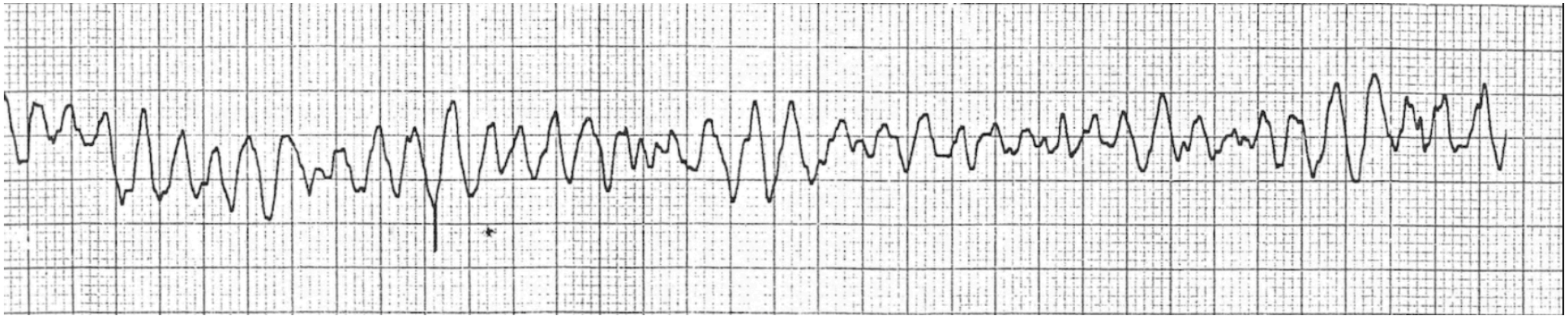


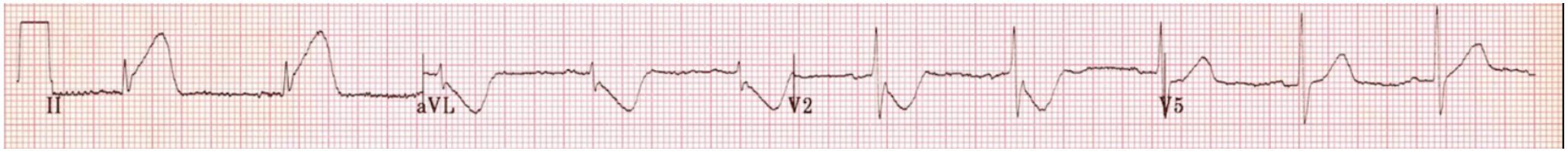
21:10 Airway The stretcher was brought into the residence and the pt was getting ready for transfer from the chair to the stretcher when he started posturing and having a seizure. Pt was lifted from the chair to the stretcher, placed supine on the stretcher and a nasal airway was inserted and breathing was assisted with a BVM and O2. Pt was transported to the unit. Oxygen initiated at 25 lpm via BVM by [redacted]. Pt. Response: Unchanged.

42yo male with cardiac arrest due to acute myocardial infarction (MI)

21:10	Airway	The stretcher was brought into the residence and the pt was getting ready for transfer from the chair to the stretcher when he started posturing and having a seizure. Pt was lifted from the chair to the stretcher, placed supine on the stretcher and a nasal airway was inserted and breathing was assisted with a BVM and O2. Pt was transported to the unit. Oxygen initiated at 25 lpm via BVM by [redacted]. Pt. Response: Unchanged.

21:15	000	V-Fib	3041A	#2
Initiate IV	Once inside the unit the pt went into cardiac arrest, ALS back-up was called, CPR was started, V-Fib was noted on the monitor, precardial thump, CPR continued. Pupils noted to be dialated and fixed. Peripheral IV initiated by [redacted] with 18ga. at LF. Attempts: 1, successful. Authorization: Via Protocol. Pt. Response: Unchanged. 1000cc's of NSS wide open.			

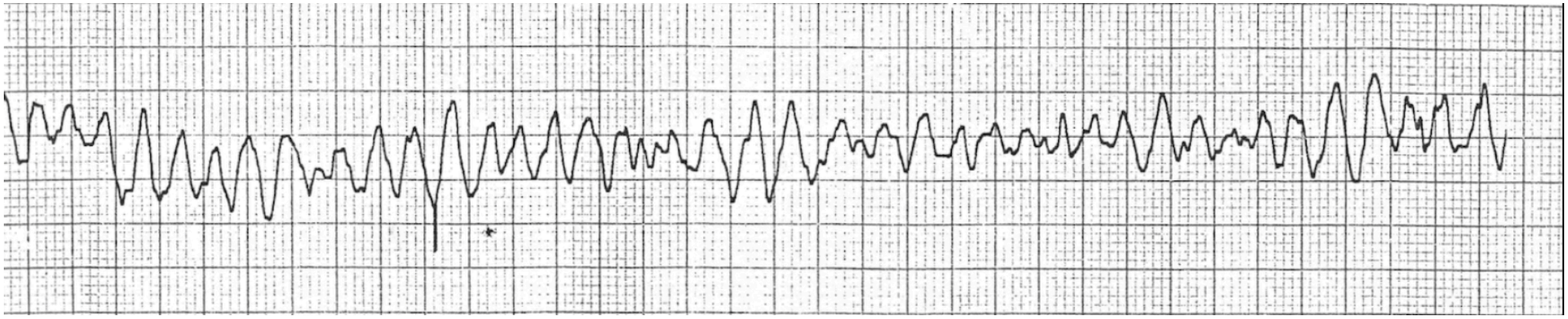


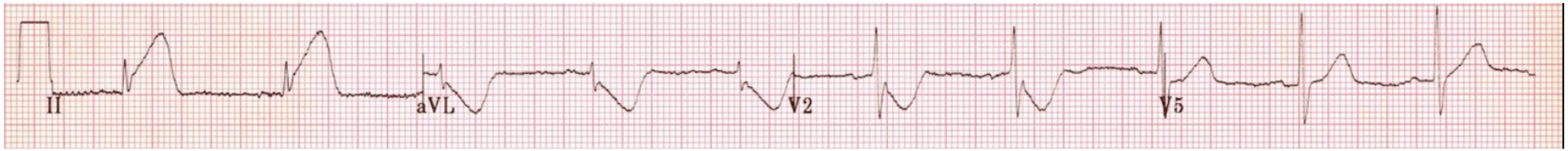


Anoxic brain injury
Expired after 10 days in hospital

21:10 Airway The stretcher was brought into the residence and the pt was getting ready for transfer from the chair to the stretcher when he started posturing and having a seizure. Pt was lifted from the chair to the stretcher, placed supine on the stretcher and a nasal airway was inserted and breathing was assisted with a BVM and O2. Pt was transported to the unit. Oxygen initiated at 25 lpm via BVM by [redacted]. Pt. Response: Unchanged.

21:15 000 V-Fib 3041A #2
Initiate IV Once inside the unit the pt went into cardiac arrest, ALS back-up was called, CPR was started, V-Fib was noted on the monitor, precardial thump, CPR continued. Pupils noted to be dialated and fixed. Peripheral IV initiated by [redacted] with 18ga. at LF. Attempts: 1, successful. Authorization: Via Protocol. Pt. Response: Unchanged. 1000cc's of NSS wide open.





42yo male with fatal, early-onset MI

MI risk factors prior to event

Total cholesterol 198 mg/dl

LDL cholesterol 124 mg/dl

HDL cholesterol 40 mg/dl

Triglycerides 170 mg/dl

Blood pressure 122/78

Body mass index 26

Non-smoker

No type 2 diabetes

Family history: father with MI at 54

ACC/AHA10y ASCVD risk calculator typically used for statin allocation decision: 1.7% ('low-risk')

Pooled Cohort Risk Assessment Equations

Predicts 10-year risk for a first atherosclerotic cardiovascular disease (ASCVD) event

ClinCalc.com » Cardiology » Pooled Cohort 10-Year ASCVD Risk Assessment Equations

Risk Factors for ASCVD

Gender	Male / Female	Systolic BP	122 mmHg
Age	42 years	Receiving treatment for high blood pressure (if SBP > 120 mmHg)	No / Yes
Race	White or other	Diabetes	No / Yes
Total Cholesterol	198 mg/dL	Smoker	No / Yes
HDL Cholesterol	40 mg/dL		

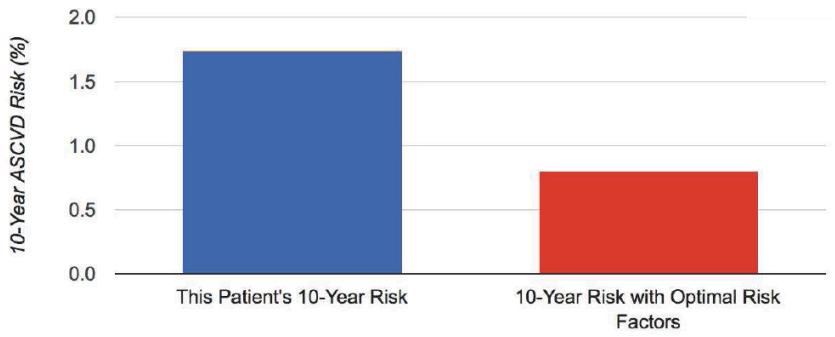
Reset | Calculate

US units

ASCVD Risk Evaluation

10-year risk of atherosclerotic cardiovascular disease: 1.7%

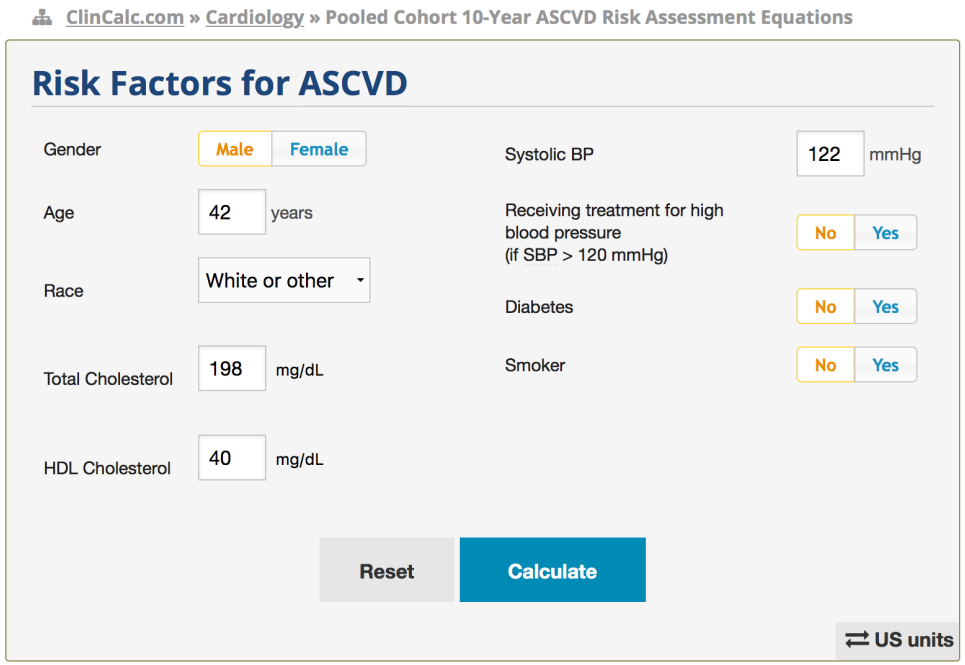
10-year risk in a similar patient with optimal risk factors: 0.8%



Why is the ACC/AHA pooled cohort equation <u>not useful</u> in young people?

Pooled Cohort Risk Assessment Equations

Predicts 10-year risk for a first atherosclerotic cardiovascular disease (ASCVD) event



<u>Model almost entirely driven by 'age'</u>

In population, older you are, more likely you are to have a heart attack!

Health care scenario

What is predicted?	Risk for heart attack
Intended target population	Men/women < 55yo
How	Gene variant(s)
For what purpose	Statin initiation at early age

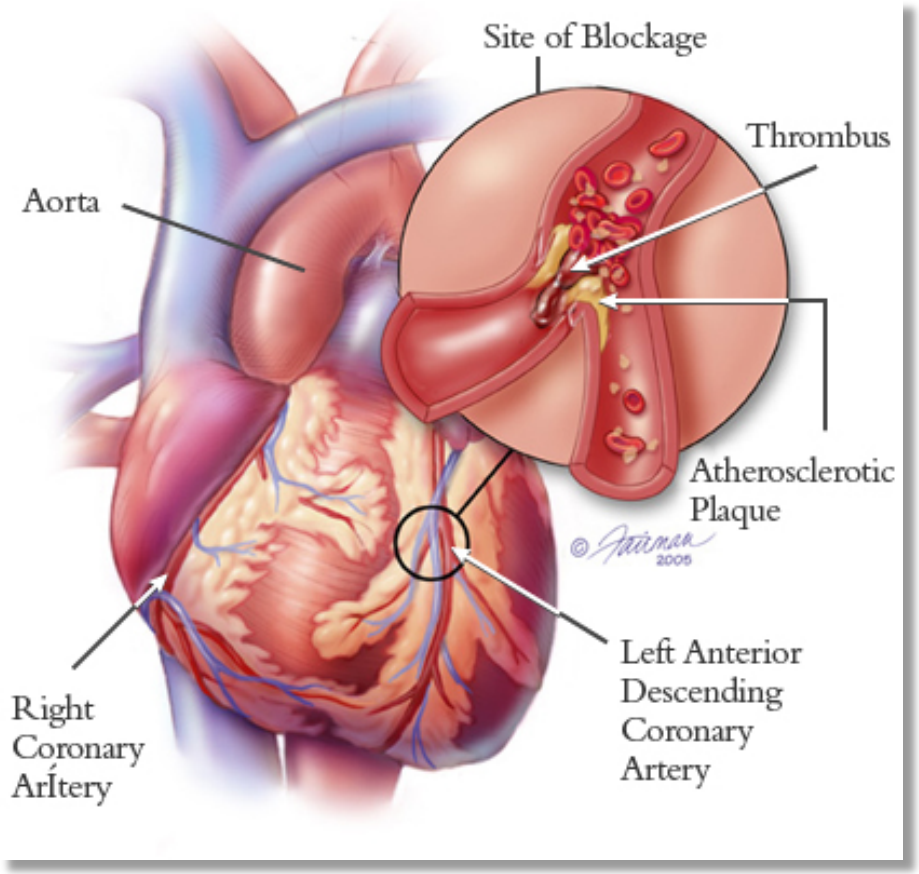
Janssens, Martens, Prediction Research Manual
http://www.cecilejanssens.org/wp-content/uploads/2018/01/PredictionManual2.0.pdf

For early-onset disease, stratifying individuals based on inborn DNA variation an option

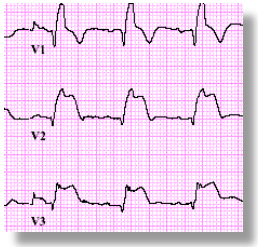
Most diseases inherited component

Stratify individuals based on inherited DNA variation

Myocardial infarction (MI) or heart attack: a classic common, complex disease



Symptoms

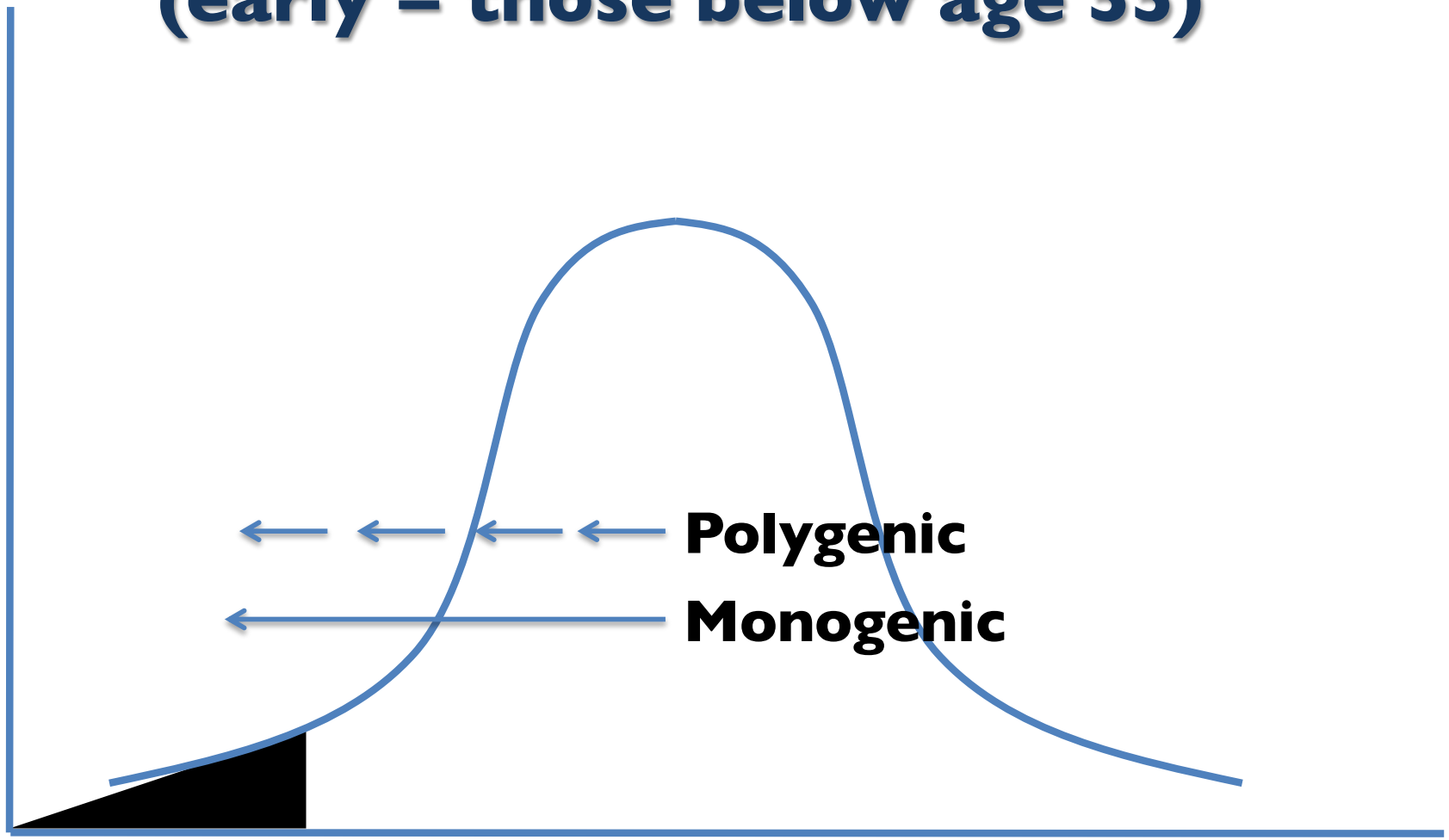


EKG change

Elevation in cardiac biomarkers

Heritable & lifestyle components

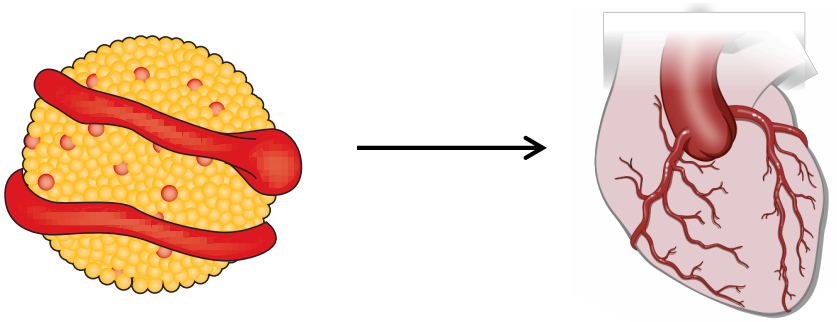
Inherited component to early heart attack (early = those below age 55)



Traditional approach:
Genetic prediction focuses on rare, monogenic mutations

Traditional approach:
Genetic prediction focuses on rare, monogenic mutations

Familial hypercholesterolemia

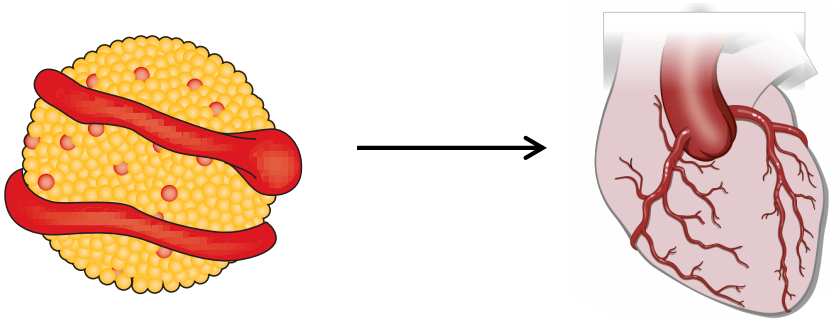


↑ **Cholesterol**

Heart attack ~3x increased risk

Traditional approach:
Genetic prediction focuses on rare, monogenic mutations

Familial hypercholesterolemia



↑ Cholesterol

Heart attack ~3x increased risk

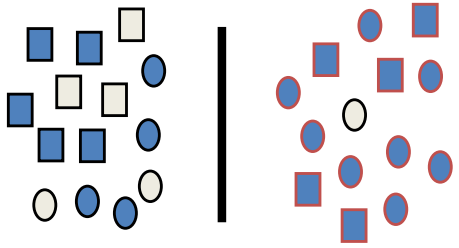
0.4% of the general population
2% of early MI patients

Question: Can we identify additional at-risk individuals with a polygenic risk model?

Hypothesis: a polygenic score including a genome-wide set of SNPs can identify individuals with risk equivalent to a familial hypercholesterolemia mutation

Step 1

Training data set: effect sizes for 6.6 million variants from genome-wide association study

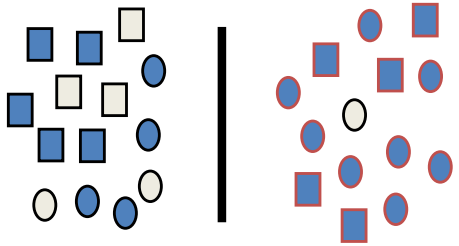


Genotypes: from arrays + imputation

Khera*, Chaffin*, *Nat Genet* (2018)

Hypothesis: a polygenic score including a genome-wide set of SNPs can identify individuals with risk equivalent to a familial hypercholesterolemia mutation

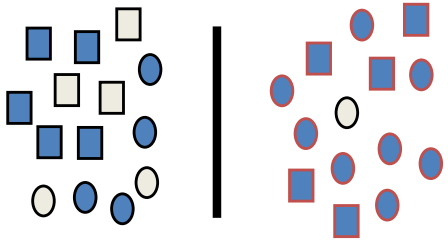
Step 1
Training data set: effect sizes for 6.6 million variants from genome-wide association study



Cases
N = 60K

Controls
N = 120K

Step 2
Validation Dataset: ~125K



Cases
N = 4K

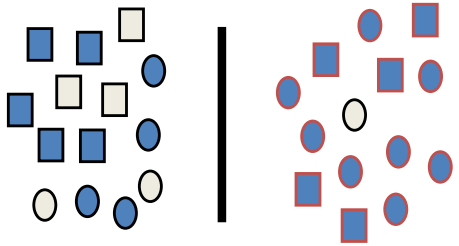
Controls
N = 120K

Genotypes: from arrays + imputation

Khera*, Chaffin*, *Nat Genet* (2018)

Hypothesis: a polygenic score including a genome-wide set of SNPs can identify individuals with risk equivalent to a familial hypercholesterolemia mutation

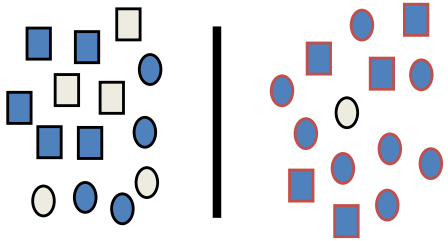
Step 1
Training data set: effect sizes for 6.6 million variants from genome-wide association study



Cases
N = 60K

Controls
N = 120K

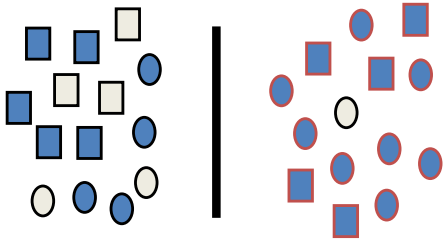
Step 2
Validation
Dataset: ~125K



Cases
N = 4K

Controls
N = 120K

Step 3
Testing
Dataset: ~300K



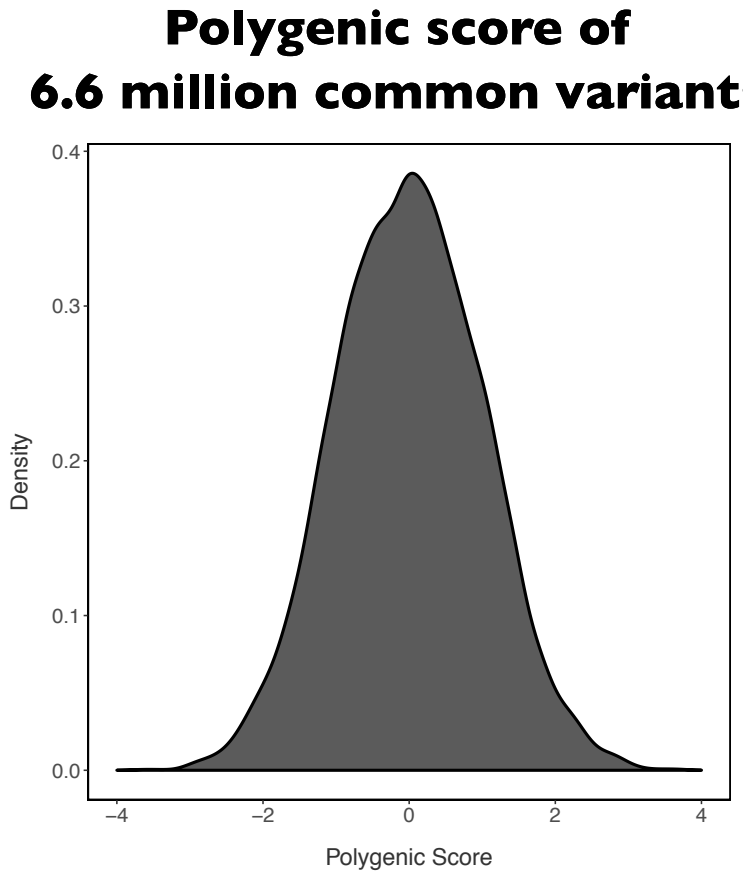
Cases
N = 8.7K

Controls
N = 288K

Genotypes: from arrays + imputation

Khera*, Chaffin*, *Nat Genet* (2018)

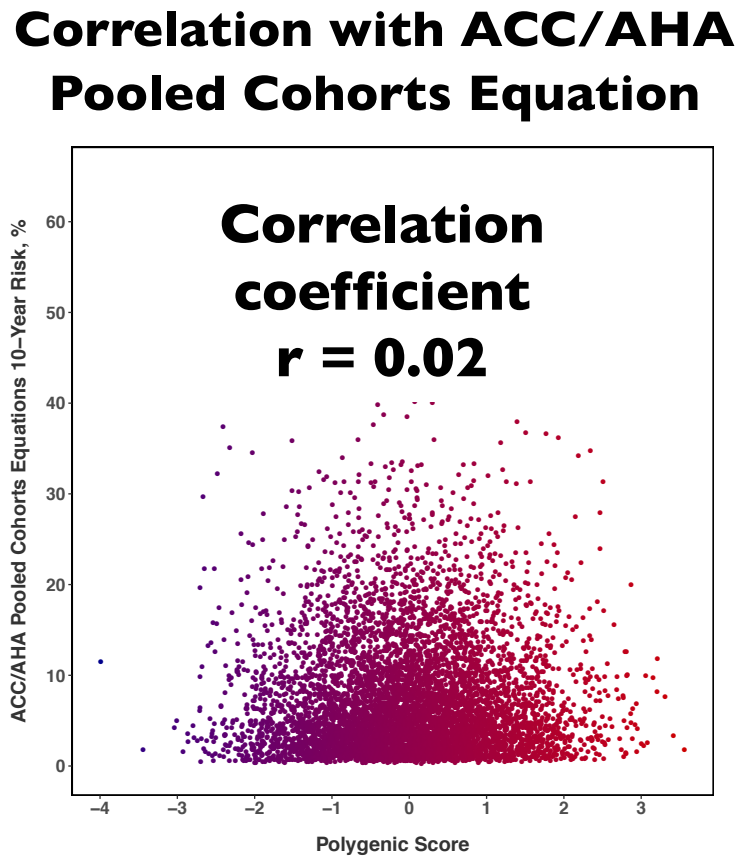
A new quantitative metric of genetic liability to heart attack



Amit V. Khera

Khera*, Chaffin*, *Nat Genet* (2018)

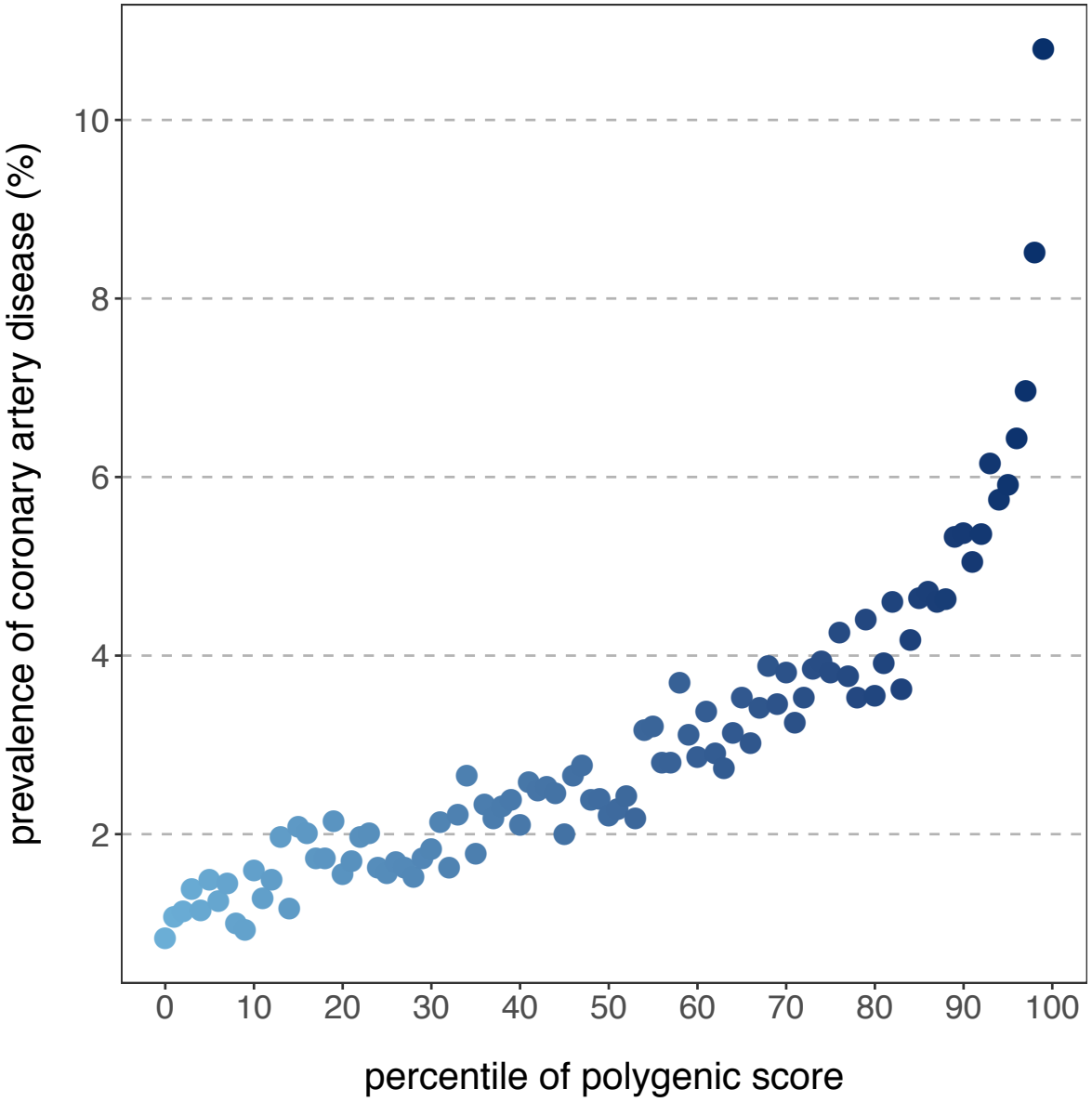
Genome-wide polygenic score: little correlation with currently measured MI risk factors



Khera*, Chaffin*, *Nat Genet* (2018)

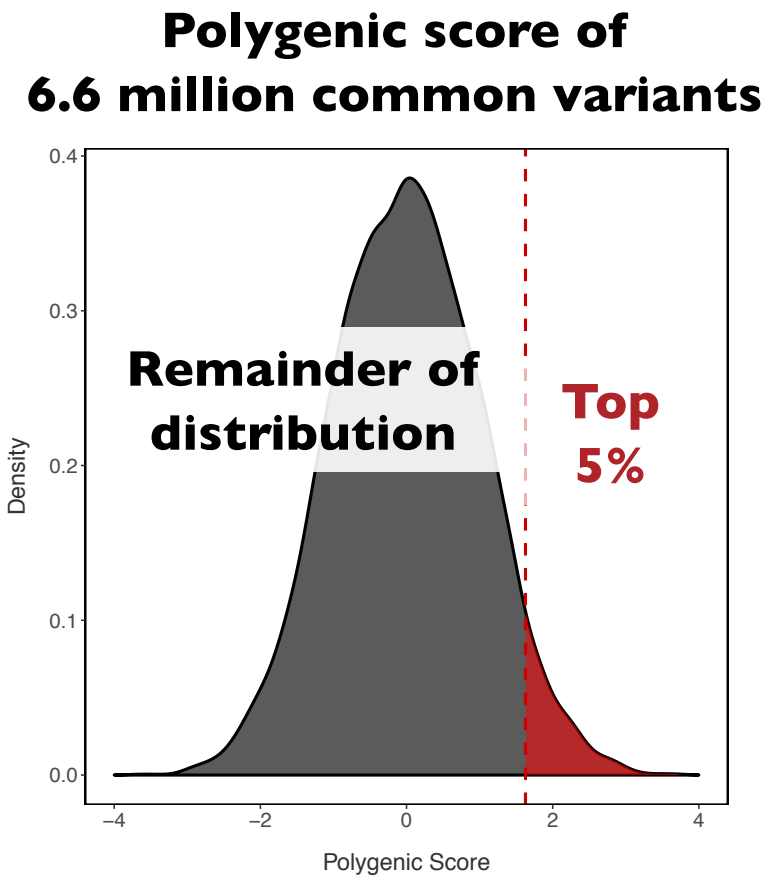
Using polygenic model, can we identify group with risk for MI equivalent to a familial hypercholesterolemia mutation?

>20-fold gradient in risk across bins of score



Khera*, Chaffin*, *Nat Genet* (2018)

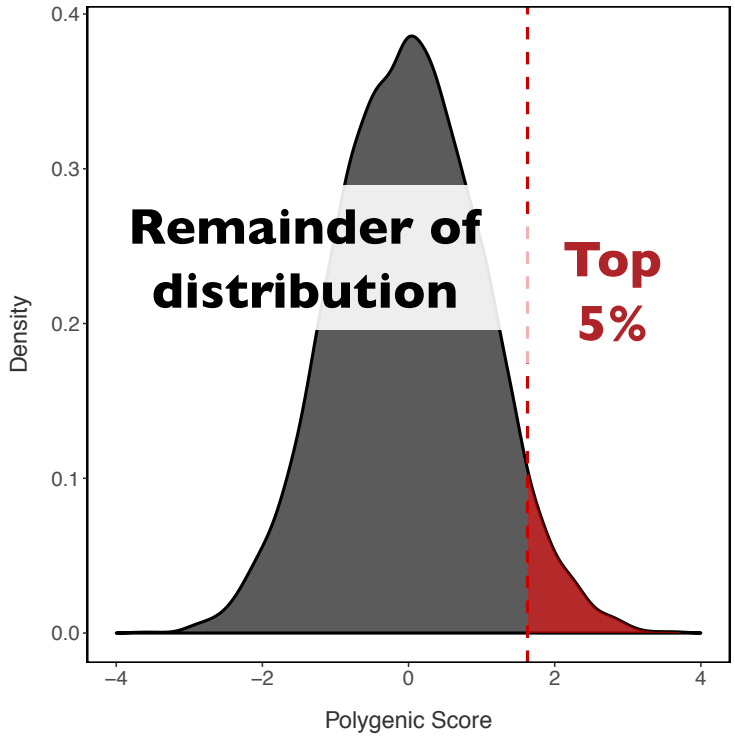
What if we label top 5% tail of distribution as ‘carriers’ and remainder as ‘non-carriers’?



Khera*, Chaffin*, *Nat Genet* (2018)

Top 5% of polygenic MI score: risk equivalent to monogenic mutations

Polygenic score of 6.6 million common variants



High polygenic score definition	Odds ratio
Top 5%	**3.3**
Top 1%	4.7

Khera*, Chaffin*, *Nat Genet* (2018)

Putting it all together:

External generalizability outside UK?

Extension to those of non-European ancestry?

Simultaneously evaluate monogenic & polygenic models

2,081 Early-onset MI patients | 3,761 Controls
United States

30X whole genome sequences

MI Cases:

- **VIRGO:** Patients hospitalized across US with first MI at age **≤ 55** years

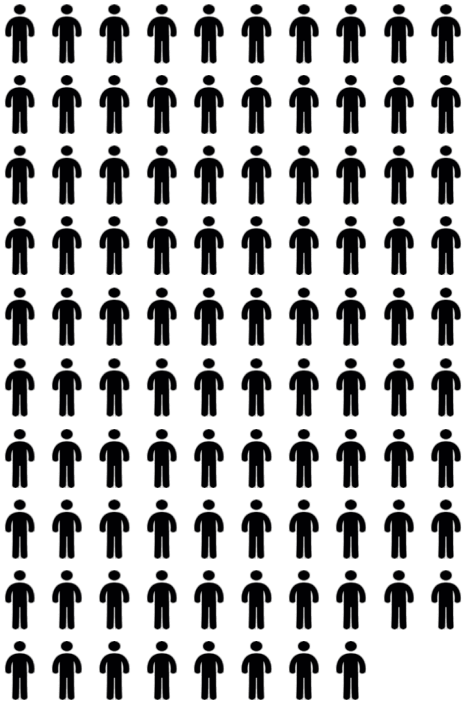
Controls:

- **MESA:** Multiethnic population free of cardiovascular disease

NHGRI Centers for Common Disease Genomics

Monogenic familial hypercholesterolemia mutation identified in 1.7% patients -> 3.8-fold increased risk

100 patients with myocardial infarction

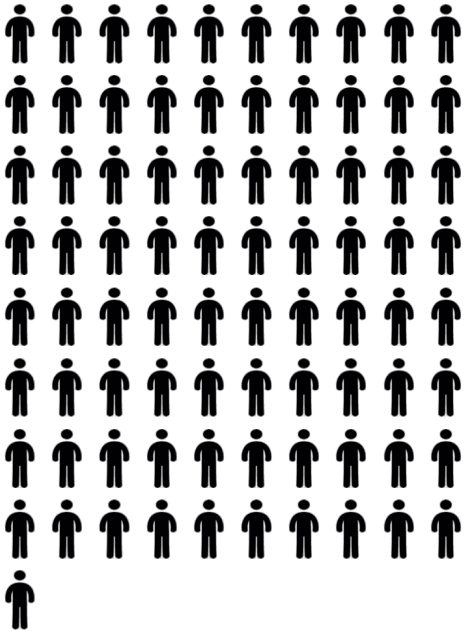


	↑ Risk
Monogenic	3.8-fold

Khera*, Chaffin*, *Circulation* 2019

High polygenic score identified in 17% of patients and confers a 3.7-fold increase in risk

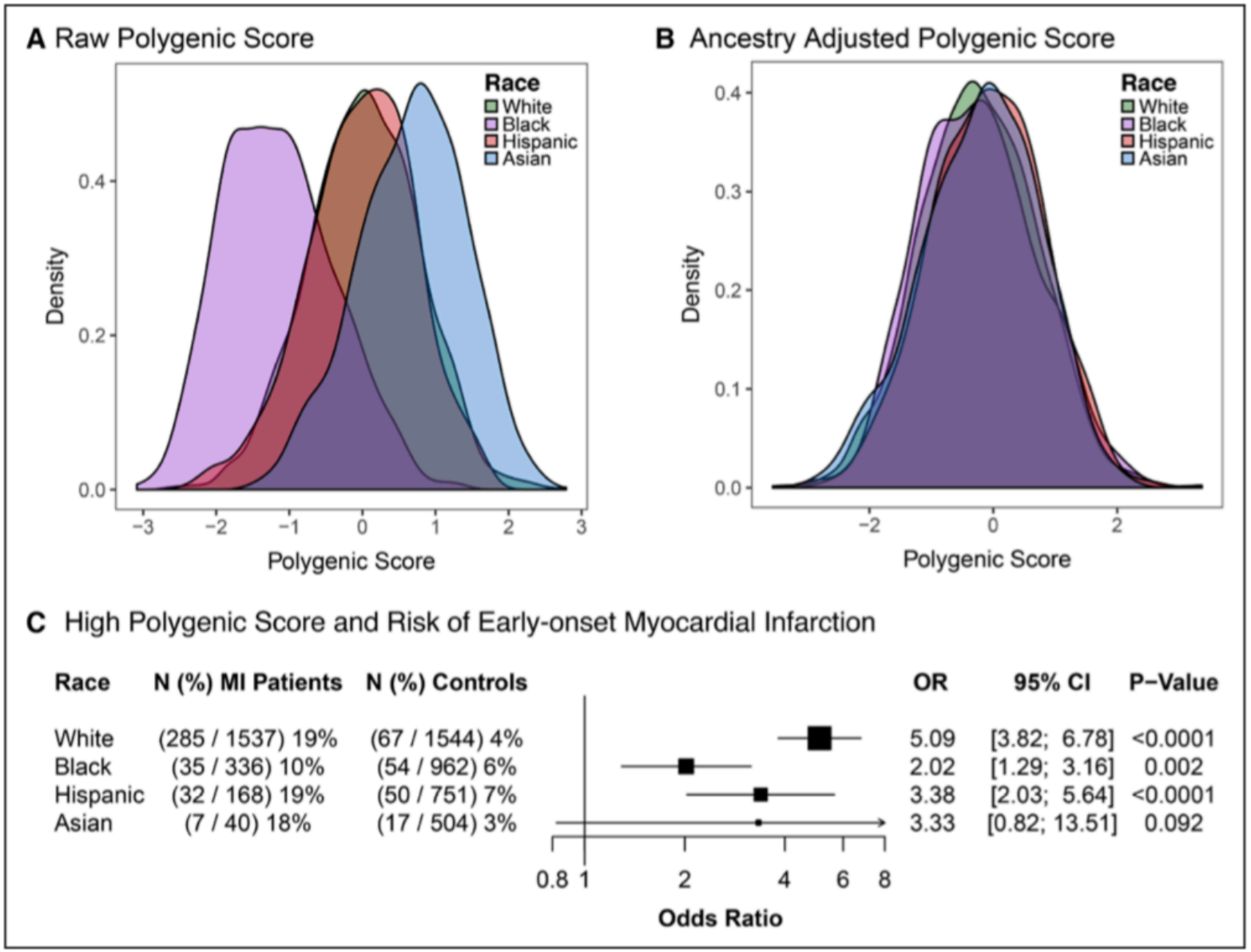
100 patients with myocardial infarction



	↑ Risk
Monogenic	**3.8-fold**
High polygenic	**3.7-fold**

Khera*, Chaffin*, *Circulation* 2019

What about extension to those of non-European ancestry?



Khera*, Chaffin*, *Circulation* 2019

Comparison

	Monogenic	Polygenic
Prevalence among early MI cases	1.7%	17%
Odd ratio for MI	3.8	3.7
Mode of detection	↑ LDL cholesterol	**Currently UNAWARE**
Mechanism of risk	apoB lipoproteins	'Gemish'
Intervention	Lifestyle Medications	**?**

Is polygenic risk for MI modifiable? Yes

Lifestyle

↓48%

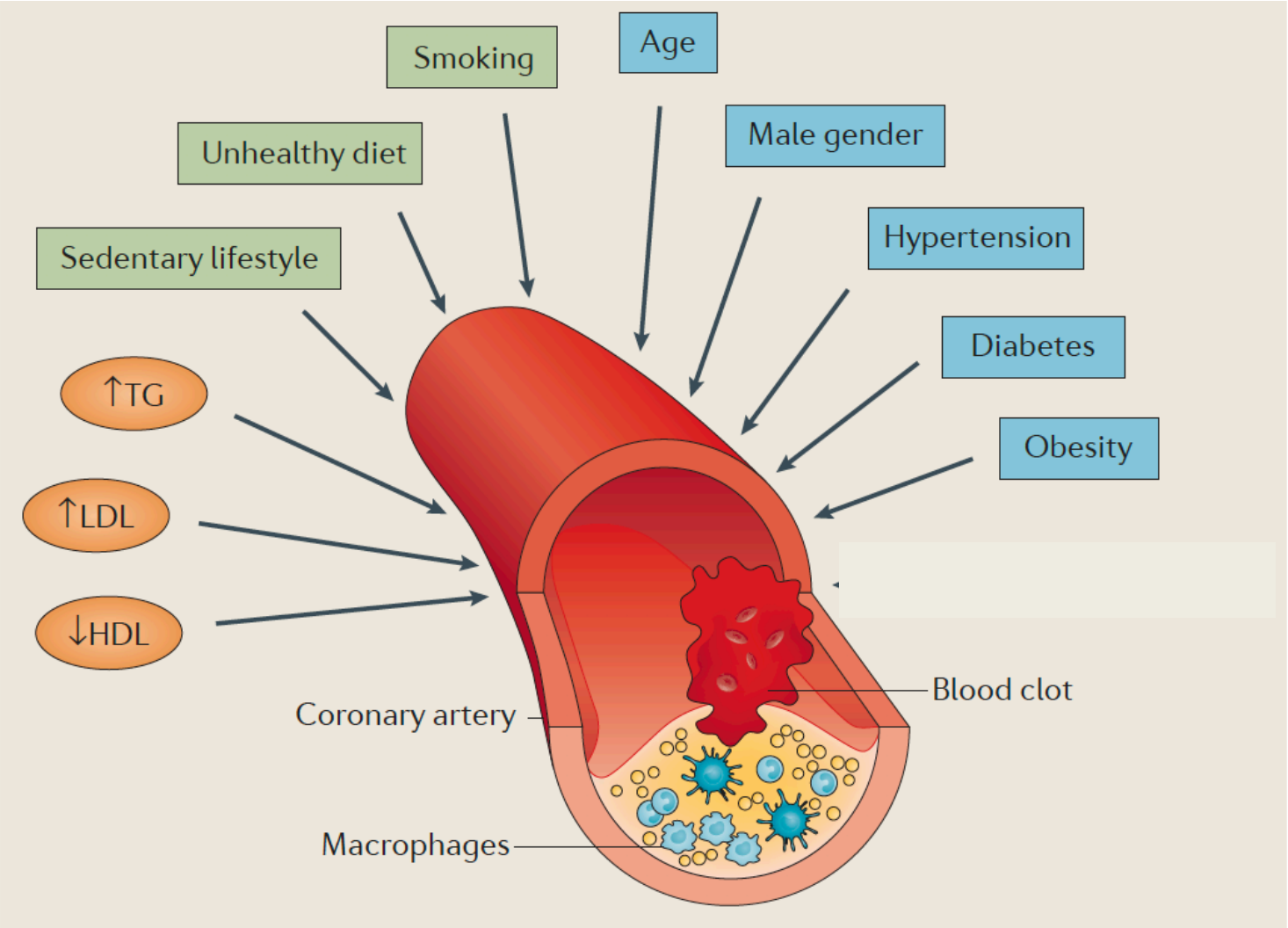
Khera, *N Engl J Med* (2016)

Medicines

↓44%

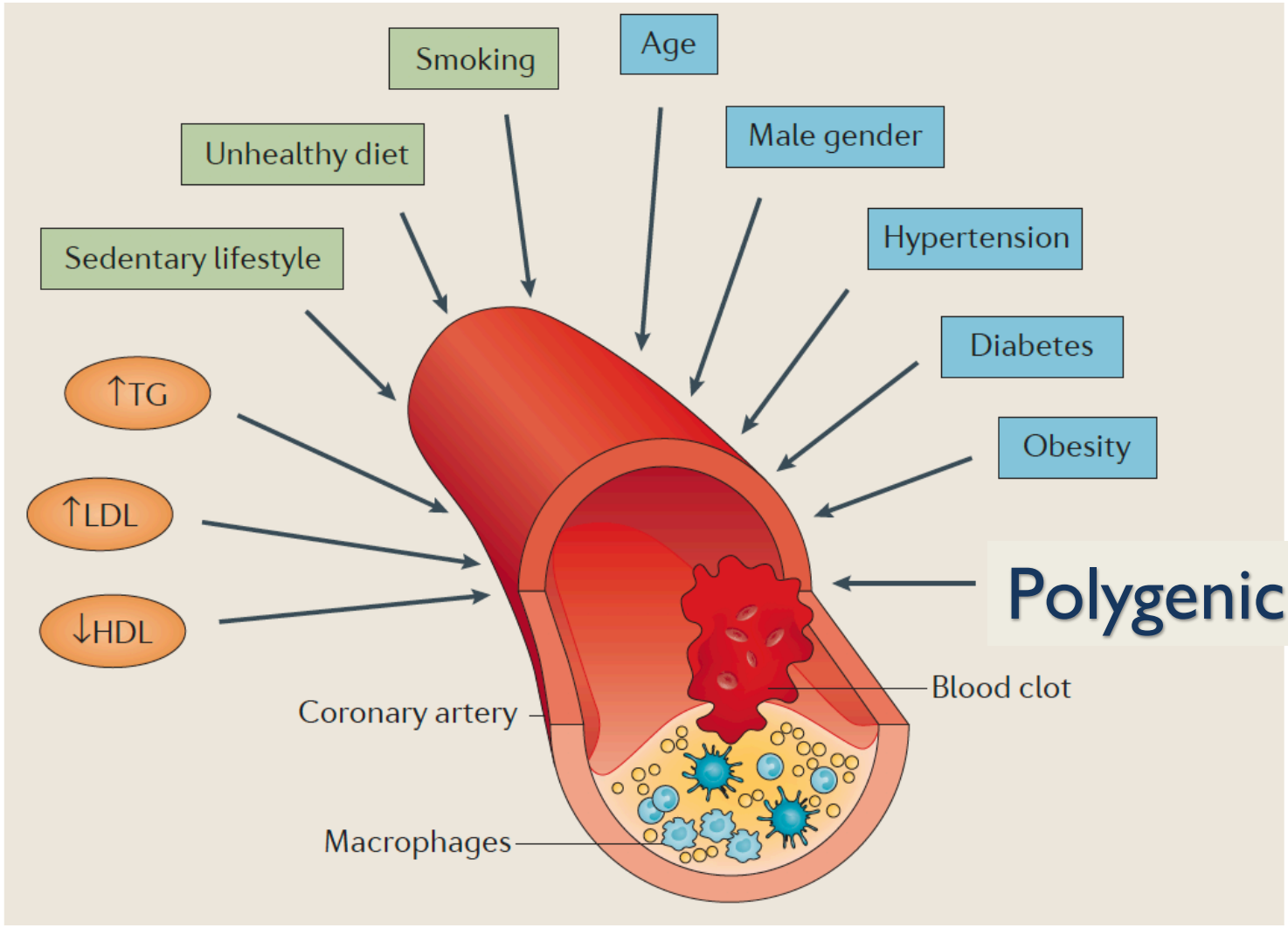
Mega*, Stitziel*, *Lancet* (2015)
Natarajan, *Circulation* (2017)

Heart attack risk assessment today



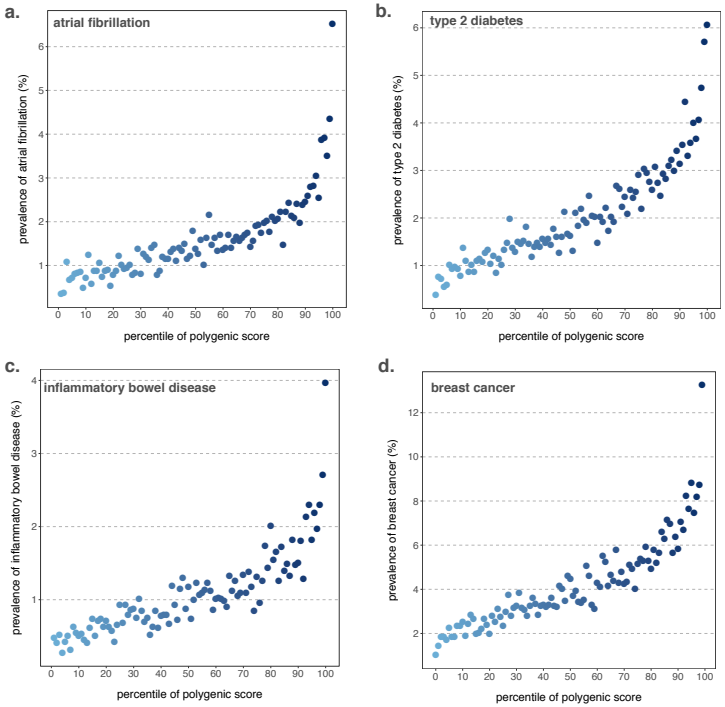
Khera, Kathiresan, *Nat Rev Genet* (2017)

Heart attack risk assessment next 5y: incorporation of polygenic score



Khera, Kathiresan, *Nat Rev Genet* (2017)

Polygenic risk for other common diseases . . . including those without monogenic risk factors



	% of population at >3-fold risk
Atrial fibrillation	**6.2%**
Diabetes	**3.6%**
Inflammatory Bowel Disease	**3.0%**
Breast cancer	**2.1%**

Potential for impact on clinical practice

Conclusions

- Now possible to score polygenic component to any complex trait (from genotyping array data, simultaneous for many diseases, at birth)

- Those in extremes of score: at risk for disease approaching monogenic mutations

- Use scores to
  - guide prevention, treatment, or screening strategies
  - gain new biologic understanding of common diseases

Center for Genomic Health

Mount Sinai

Predicting Genetic Risk for Diverse Populations and Complex Traits

Eimear Kenny, PhD

Founding Director, The Center for Genomic Health
Associate Professor of Medicine
Associate Professor of Genetics and Genomic Medicine
Icahn School of Medicine at Mount Sinai

From Fisher to Visscher

"The Correlation between Relatives on the Supposition of Mendelian Inheritance" by R.A. Fisher was published in the *Transactions of the Royal Society of Edinburgh* (1918)

"Complex Trait Prediction from Genome Data: Contrasting EBV in Livestock to PRS in Humans" by N.R. Wray, K.E. Kemper, B.J. Hayes, M.E. Goddard, P.M. Visscher was published in *Genetics* (2019)

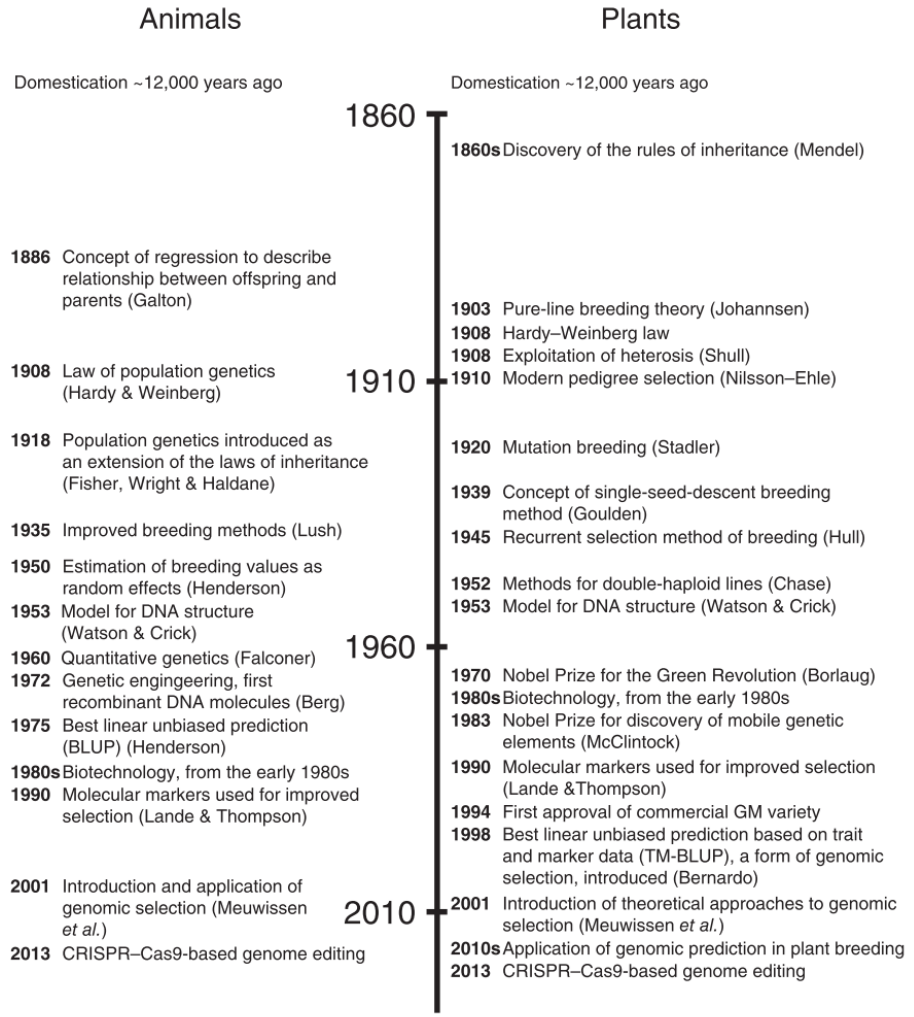
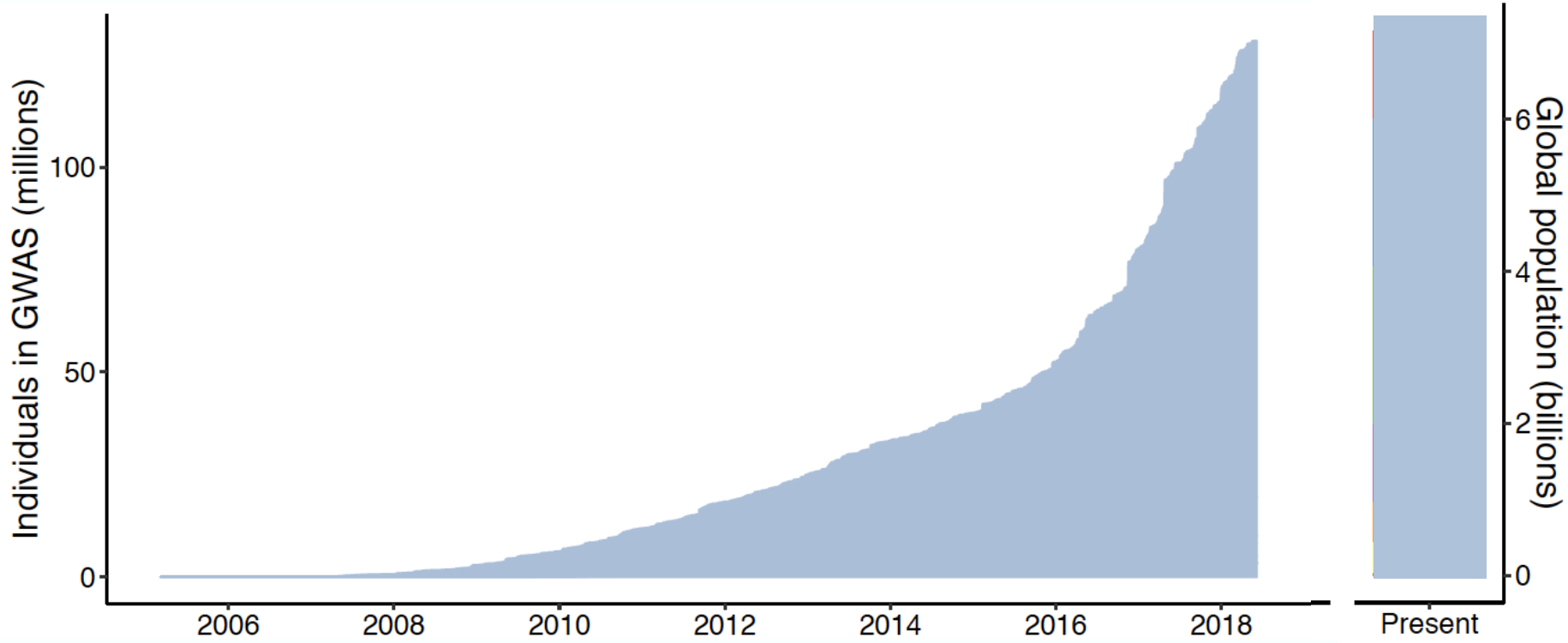


Figure 1 Some key milestones of selective animal and plant breeding.

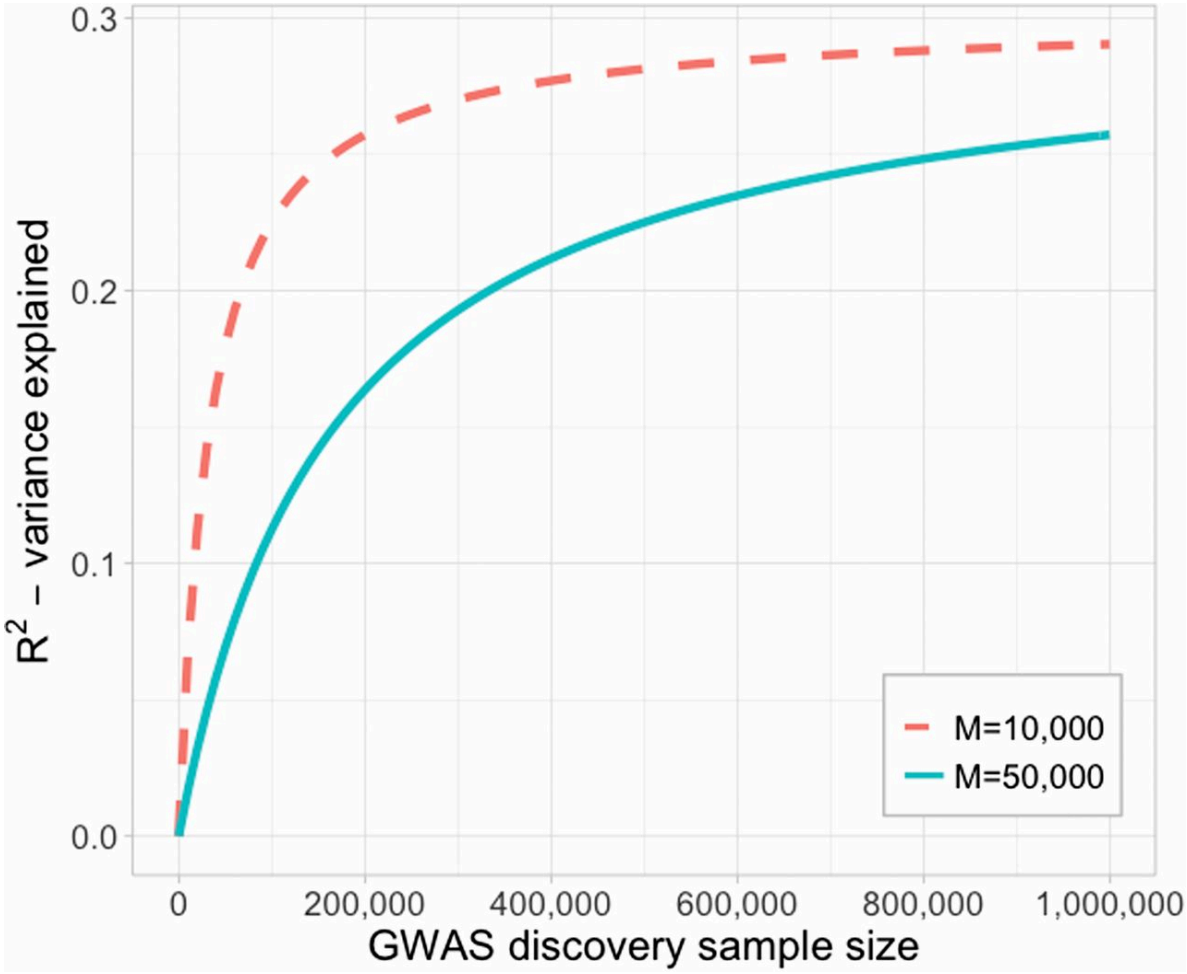
Hickey, J.M., et al (2017). Nature Genetics. 49, 1297-1303

GWAS are increasing in size and scope



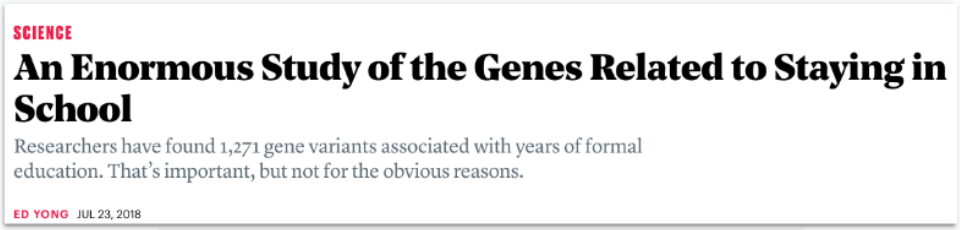
Martin, et al. (2019) Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet

Sample size is the key factor for maximizing accuracy



N.R. Wray, K.E. Kemper, B.J. Hayes, M.E. Goddard, P.M. Visscher (2019) Complex Trait Prediction from Genome Data: Contrasting EBV in Livestock to PRS in Humans. Genetics

Applications of polygenic risk prediction


SCIENCE

An Enormous Study of the Genes Related to Staying in School

Researchers have found 1,271 gene variants associated with years of formal education. That's important, but not for the obvious reasons.

ED YONG JUL 23, 2018


The New York Times

Clues to Your Health Are Hidden at 6.6 Million Spots in Your DNA

By Gina Kolata

Aug. 13, 2018

With a sophisticated new algorithm, scientists have found a way to forecast an individual's risks for five deadly diseases.


The New York Times

Opinion

Why Progressives Should Embrace the Genetics of Education

By Kathryn Paige Harden

Dr. Harden is a psychologist who studies how genetic factors shape adolescent development.

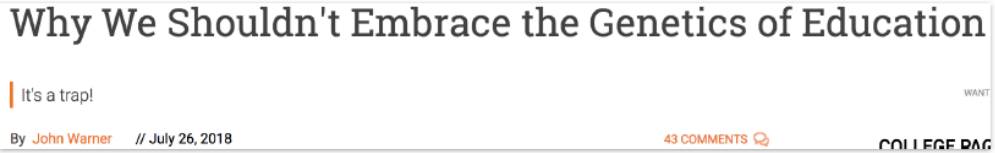
July 24, 2018

365


MIT Technology Review

Forecasts of genetic fate just got a lot more accurate

by Antonio Regalado February 21, 2018


Why We Shouldn't Embrace the Genetics of Education

It's a trap!

By John Warner // July 26, 2018

43 COMMENTS


A Population Genetic Signal of Polygenic Adaptation

Jeremy J. Berg, Graham Coop

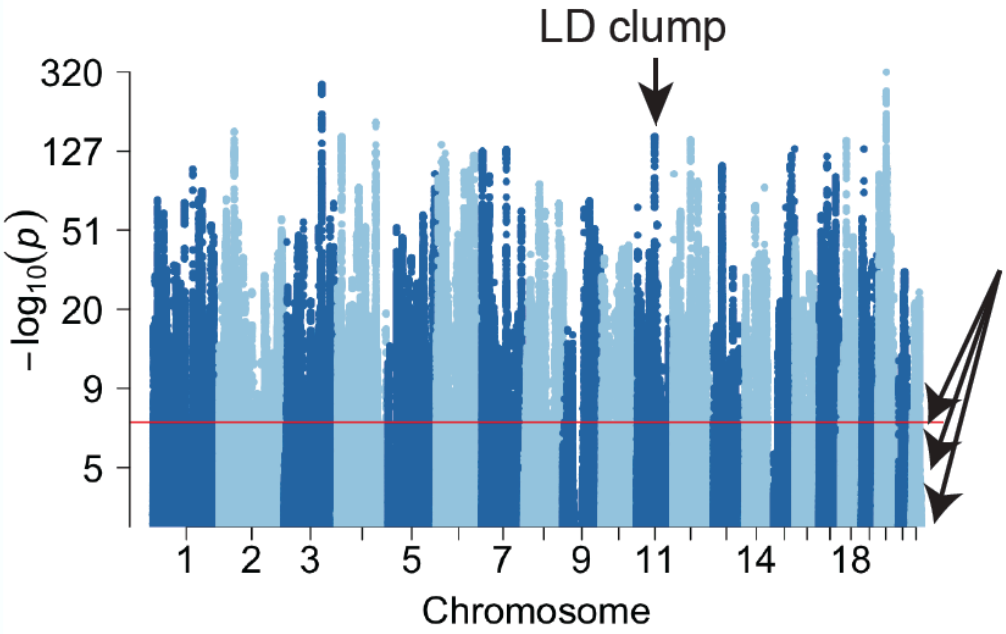
Published: August 7, 2014 • https://doi.org/10.1371/journal.pgen.1004412 • >> See

PLOS GENETICS

- Evolutionary genetics
- Social sciences
- Disease risk
- Preventive Health

Slide adapted from Alicia Martin

What is a Polygenic Risk Score?



Polygenic risk score: genetic prediction of an individual's phenotype. Calculated by summing across the products of GWAS effect sizes and number of trait-increasing alleles.

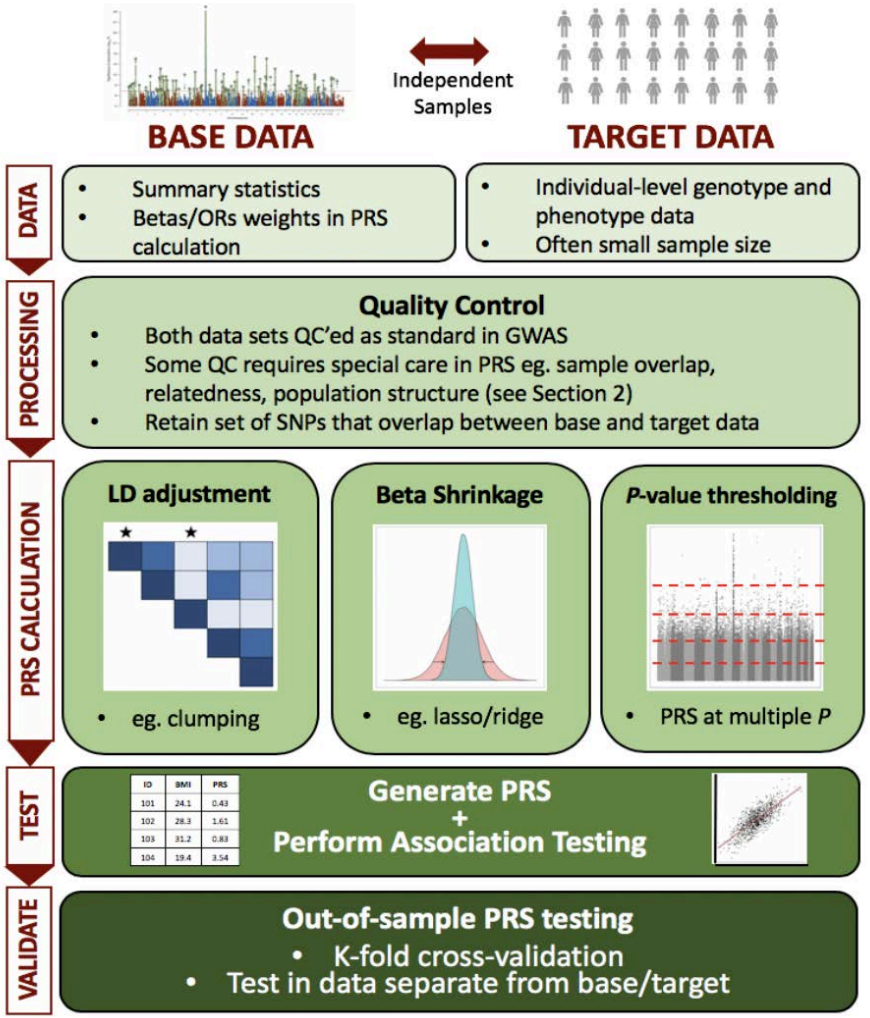
For each threshold (e.g. p-value):

$$Y = \sum_{j=1}^{m} g_j \beta_j$$

Measure accuracy with R^2 (continuous), pseudo-R^2 (binary), or AUC.

Martin, et al. (2018) Predicting polygenic risk of psychiatric disorders. Biological Psychiatry

How to calculate a Polygenic Risk Score



S.W. Choi, T.S. Heng Mak, P.F. O'Reilly (2018) A guide to performing Polygenic Risk Score analysis. BioRxiv

Model development and evaluation

nature REVIEWS GENETICS

Opinion | Published: 18 June 2013

Pitfalls of predicting complex traits from SNPs

Naomi R. Wray, Jian Yang, Ben J. Hayes, Alkes L. Price, Michael E. Goddard & Peter M. Visscher

Nature Reviews Genetics **14**, 507–515 (2013) | Download Citation

nature genetics

Analysis | Published: 03 March 2013

Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies

Nilanjan Chatterjee, Bill Wheeler, Joshua Sampson, Patricia Hartge, Stephen J Chanock & Ju-Hyun Park

Nature Genetics **45**, 400–405 (2013) | Download Citation

CSH Cold Spring Harbor Laboratory

bioRχiv

THE PREPRINT SERVER FOR BIOLOGY

New Results

A guide to performing Polygenic Risk Score analyses

Shing Wan Choi, Timothy Shin Heng Mak, Paul O'Reilly

doi: https://doi.org/10.1101/416545

This article is a preprint and has not been peer-reviewed [what does this mean?].

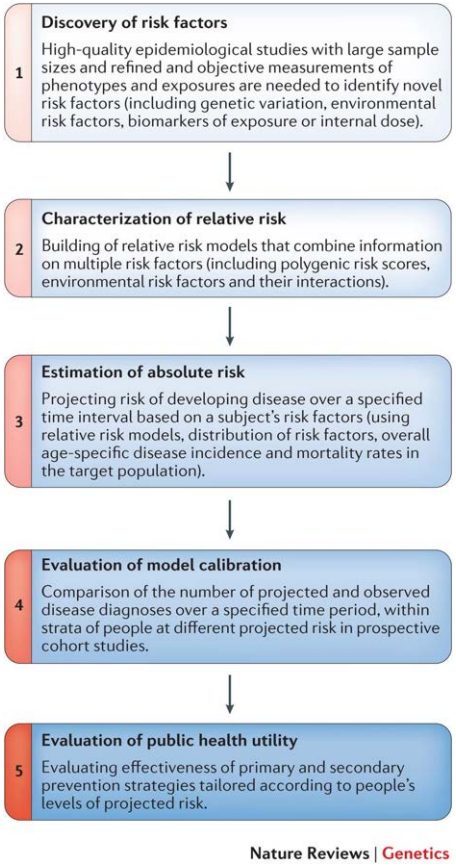
nature genetics

Technical Report | Published: 02 February 2015

LD Score regression distinguishes confounding from polygenicity in genome-wide association studies

Brendan K Bulik-Sullivan, Po-Ru Loh, Hilary K Finucane, Stephan Ripke, Jian Yang, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Nick Patterson, Mark J Daly, Alkes L Price & Benjamin M Neale

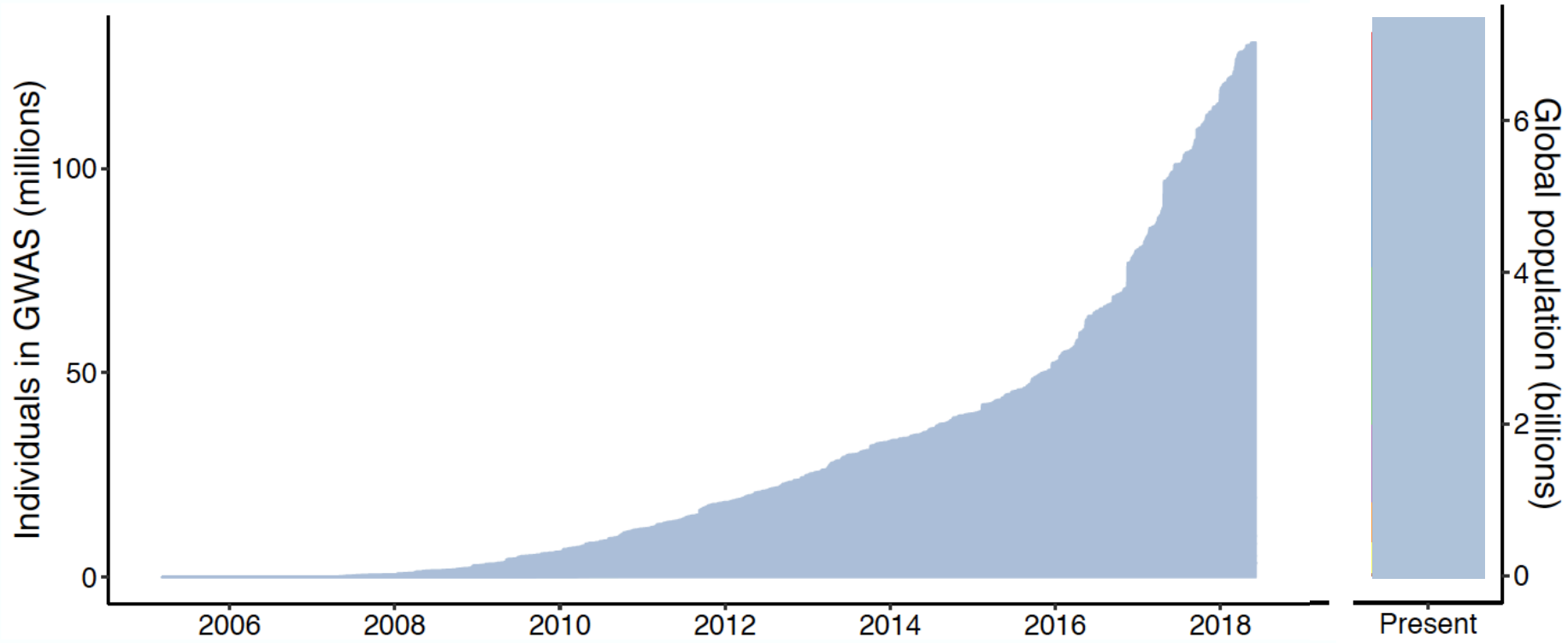
Nature Genetics **47**, 291–295 (2015) | Download Citation



Building and evaluating absolute risk models for general population
(Chatterjee et al, 2016)

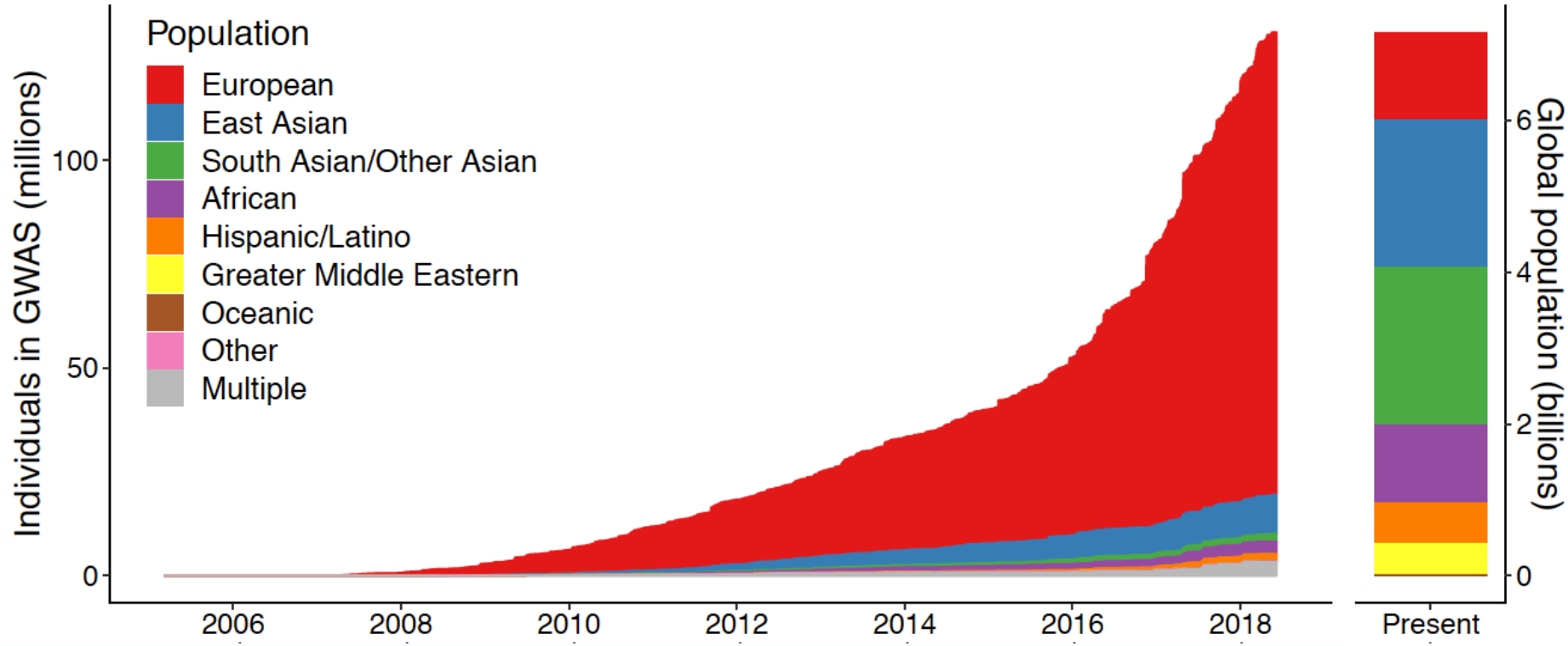
Slide adapted from Gen Wojcik

GWAS are increasing in size and scope …



Martin, et al. (2019) Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet

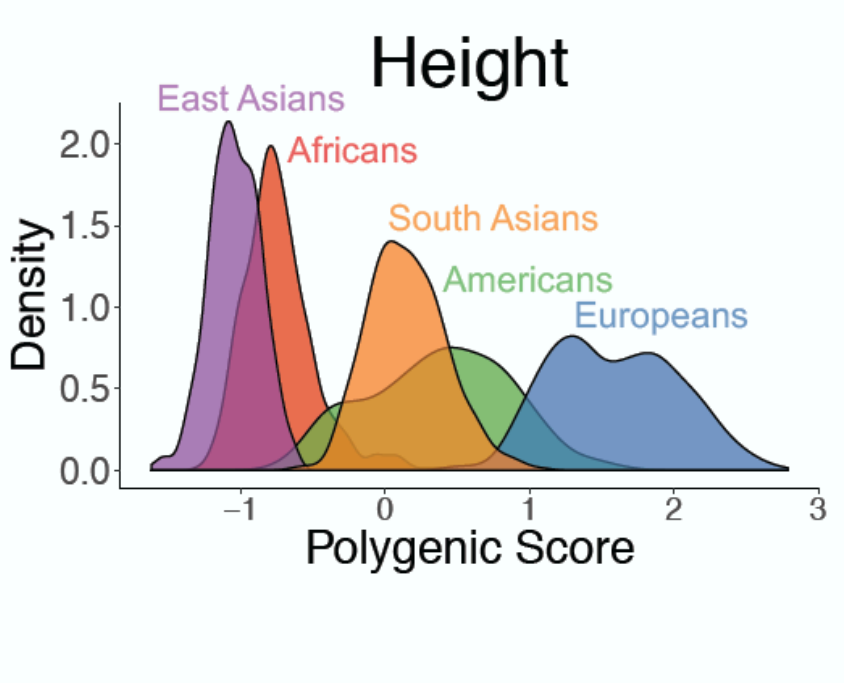
…. but genomics is failing on diversity



Martin, et al. (2019) Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet

What effect does ancestry have on prediction?

European ascertainment of GWAS signals yield unpredictably biased risk scores in other populations

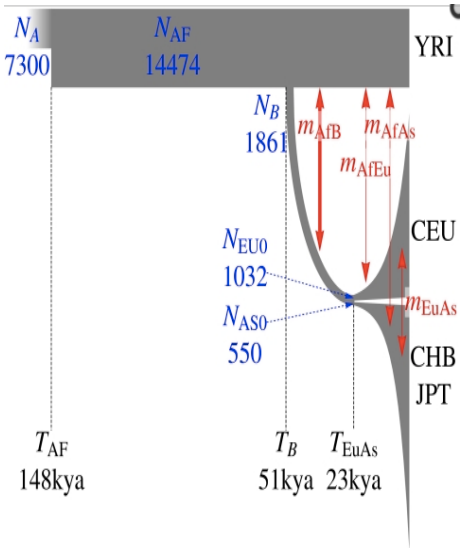


Alicia Martin

Martin et al., Population genetic history and polygenic risk biases in 1000 Genomes populations, AJHG 2017

What effect does ancestry have on prediction?

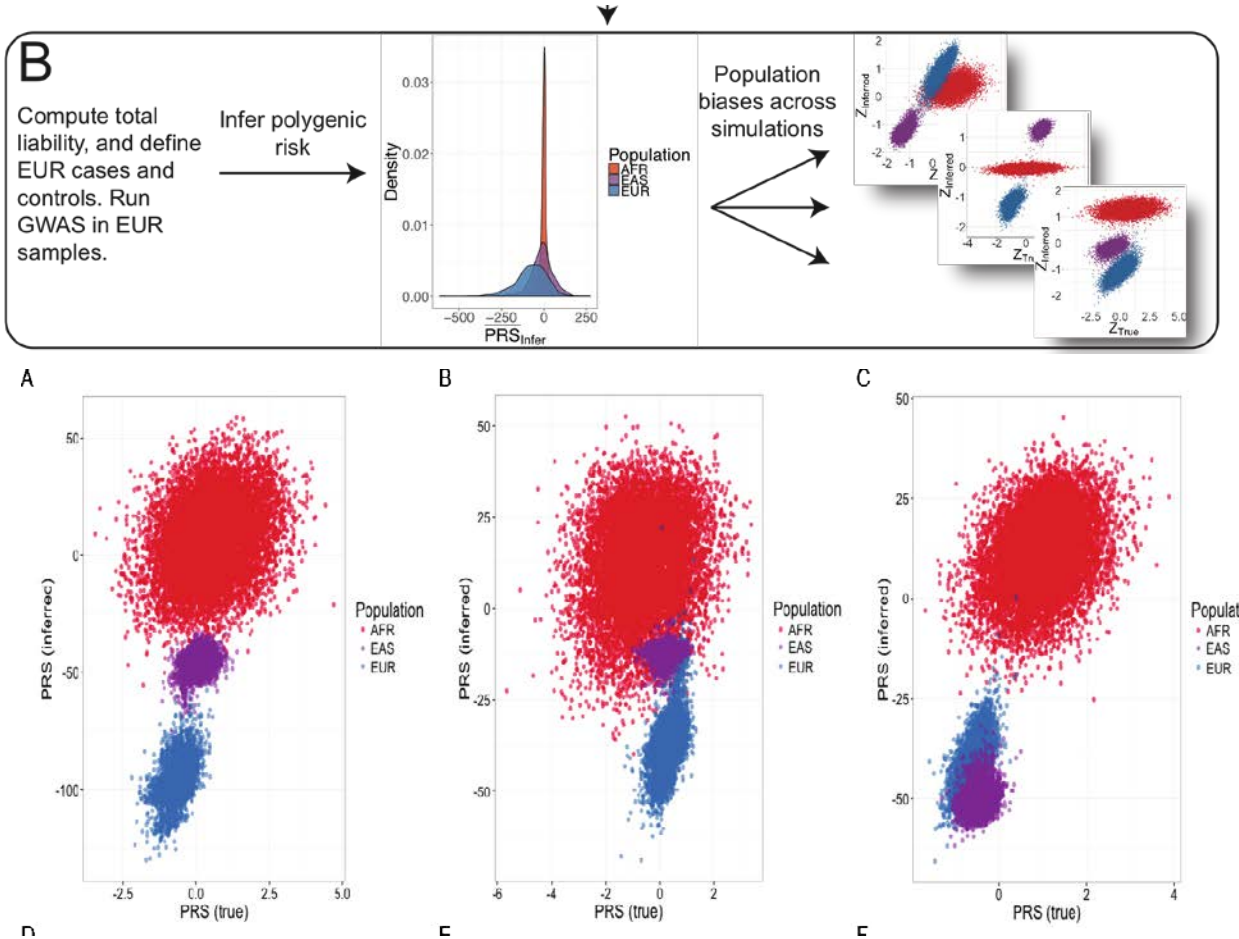
Gravel *et al* 2011



Alicia Martin

Martin et al., Population genetic history and polygenic risk biases in 1000 Genomes populations, AJHG 2017

What effect does ancestry have on prediction?

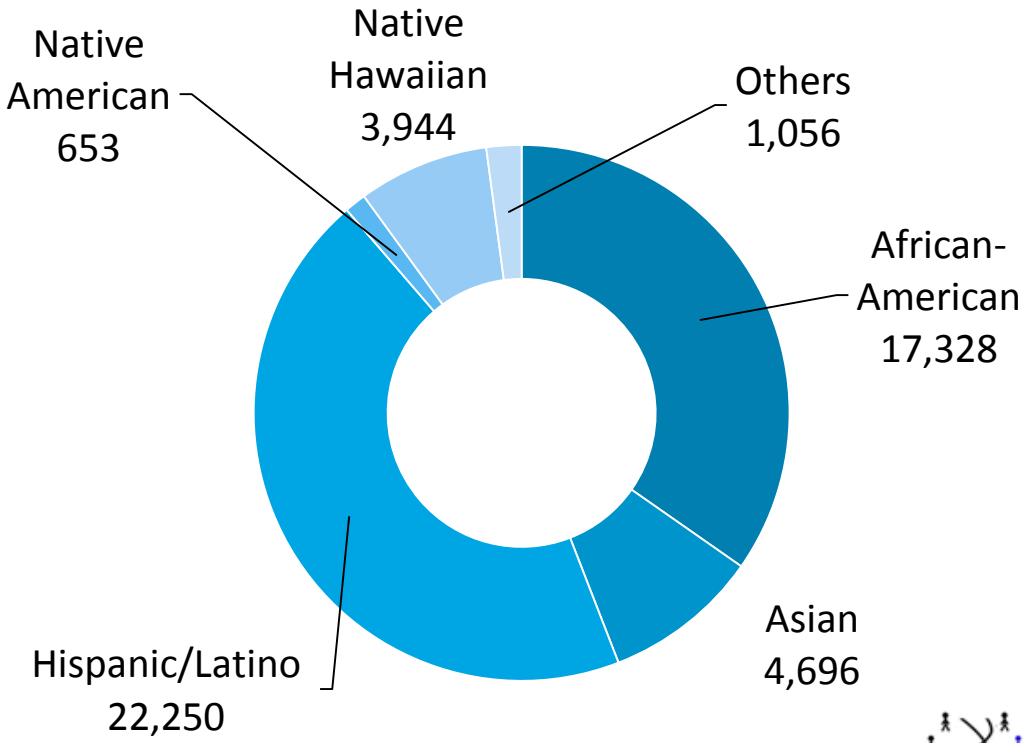


Alicia Martin

Martin et al., Population genetic history and polygenic risk biases in 1000 Genomes populations, AJHG 2017

Genetic diversity improves our understanding of complex traits

- **Goal:** Investigate ancestrally diverse populations to gain a better understanding of how genetic factors influence susceptibility to disease.
- **Focus on US minority populations.**



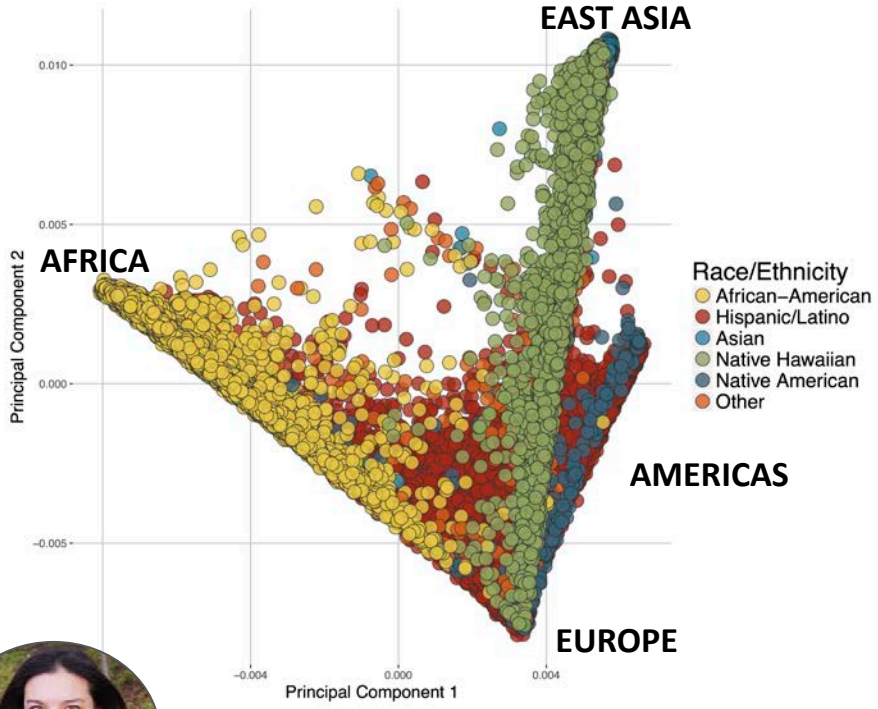
Gen Wojcik

The PAGE Network

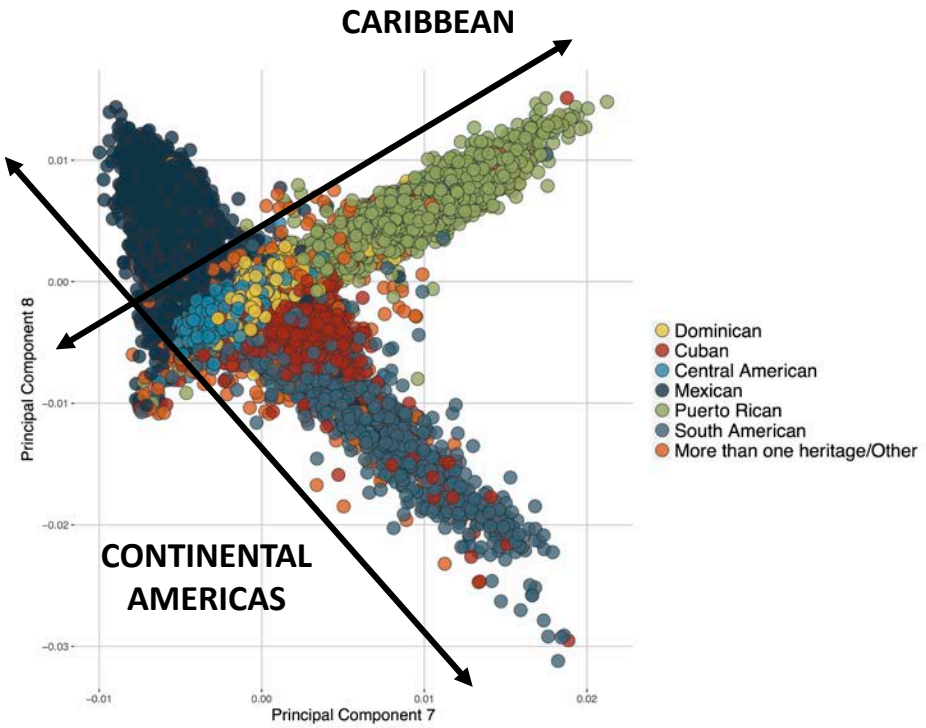
Population Architecture using Genomics and Epidemiology

Wojcik GL[#], Graff M[#], Nishimura KK[#], Tao R[#], Haessler J[#], Gignoux CR[#], Highland HM[#], Patel YM[#], … Kenny EE[$], Carlson CS[$]. Genetic diversity improves our understanding of complex traits. BioRxiv (2019)

Challenge defining population groups



All PAGE Populations

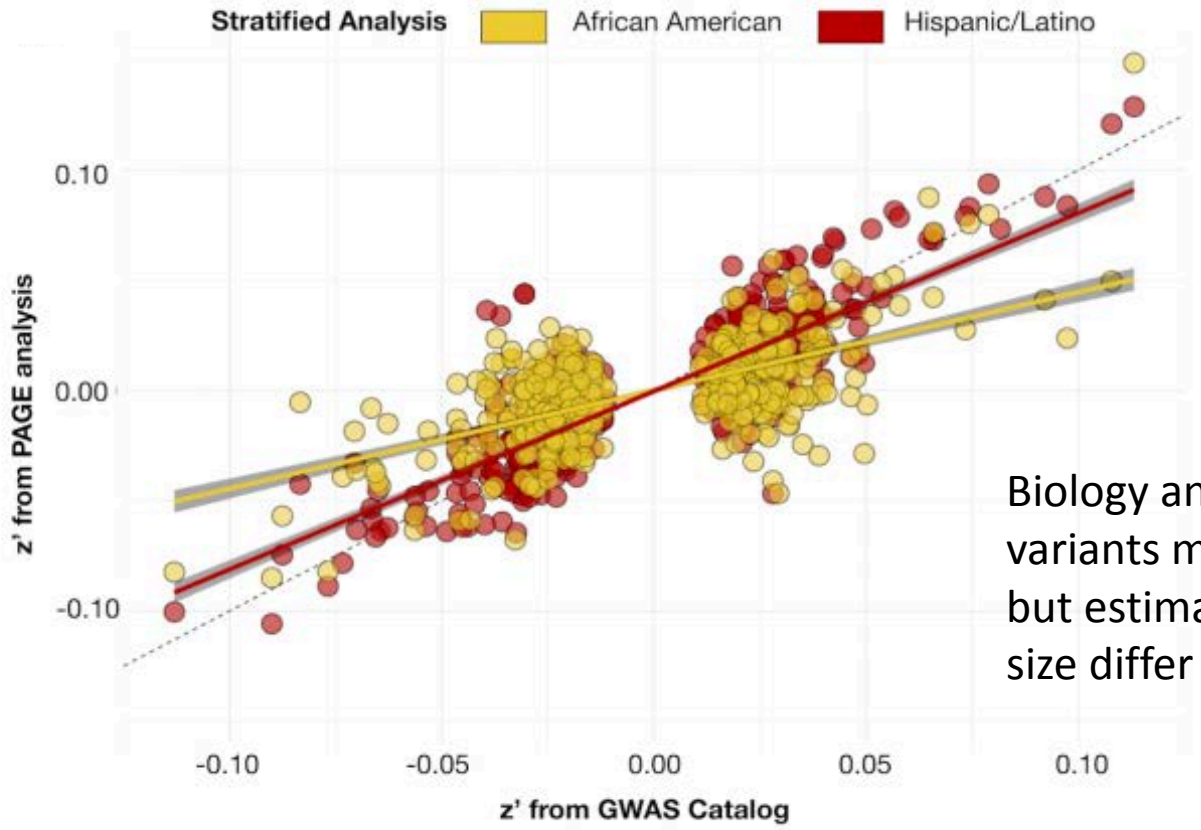


PC7 vs PC8, PAGE Hispanic/Latino variation

Gen Wojcik

Wojcik GL[#], Graff M[#], Nishimura KK[#], Tao R[#], Haessler J[#], Gignoux CR[#], Highland HM[#], Patel YM[#], … Kenny EE[$], Carlson CS[$]. Genetic diversity improves our understanding of complex traits. BioRxiv (2019)

European effect sizes are weaker in non-European populations

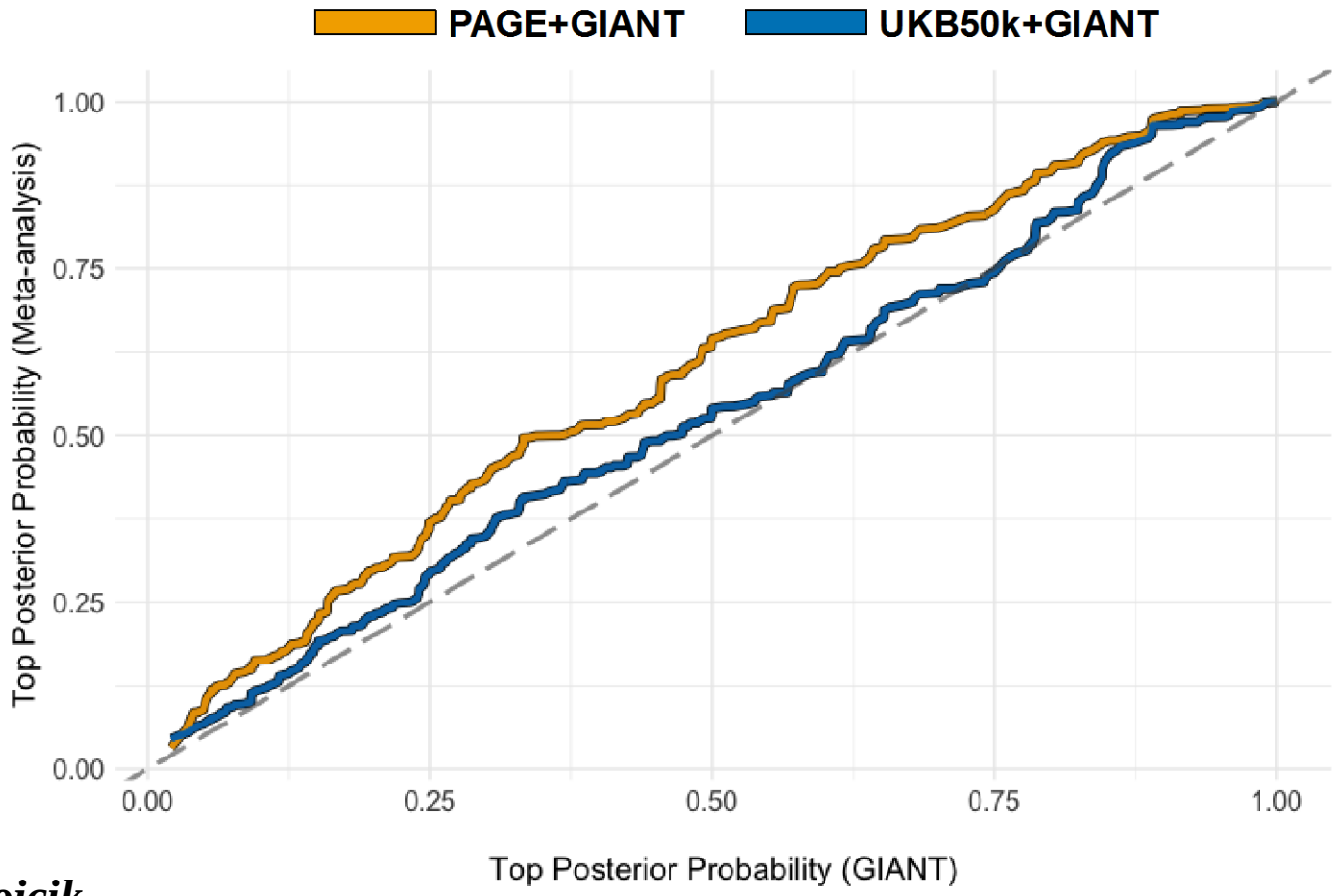


Biology and causal variants mostly shared, but estimates of effect size differ

Gen Wojcik

Wojcik GL[#], Graff M[#], Nishimura KK[#], Tao R[#], Haessler J[#], Gignoux CR[#], Highland HM[#], Patel YM[#], … Kenny EE[$], Carlson CS[$]. Genetic diversity improves our understanding of complex traits. BioRxiv (2019)

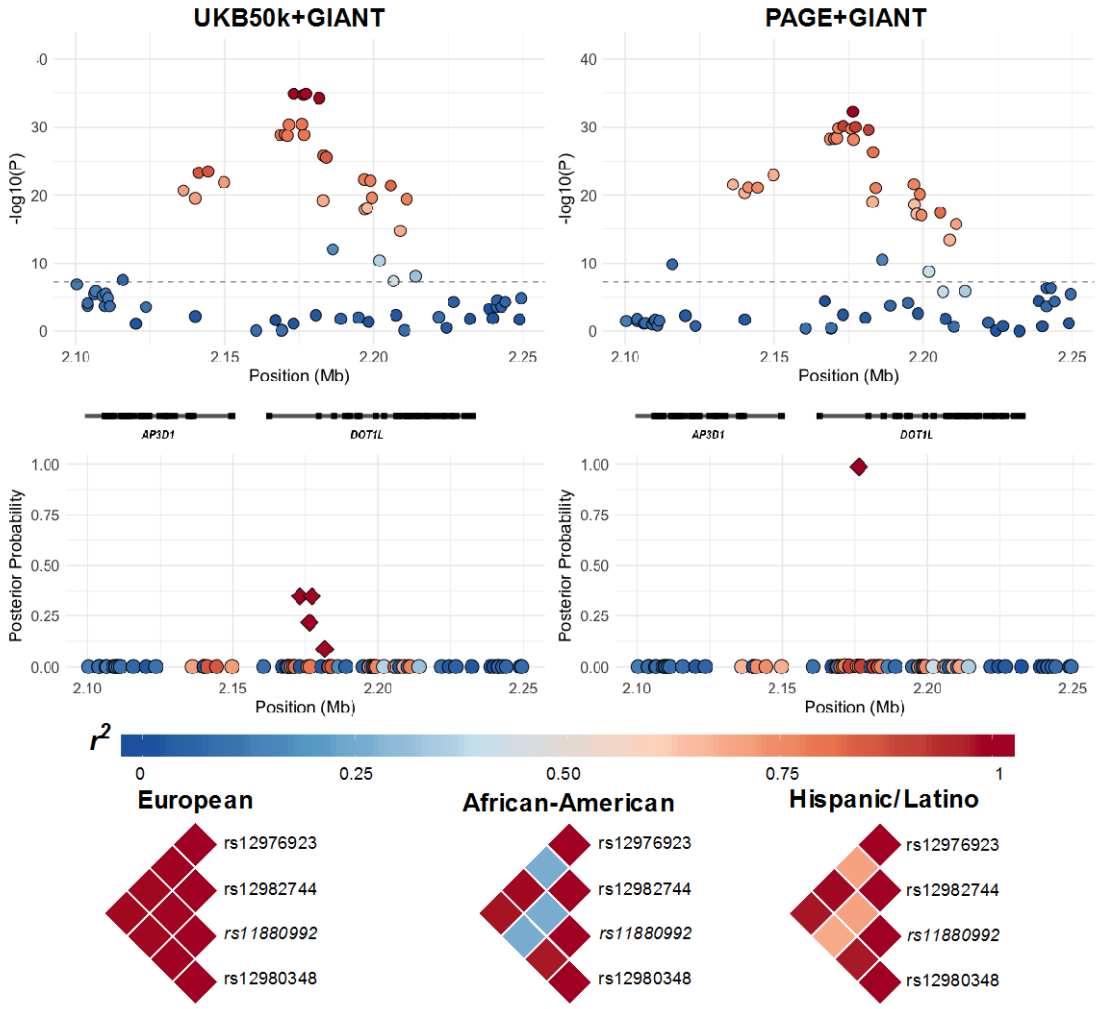
Diversity increases confidence in potential causal variants



Gen Wojcik

Wojcik GL[#], Graff M[#], Nishimura KK[#], Tao R[#], Haessler J[#], Gignoux CR[#], Highland HM[#], Patel YM[#], … Kenny EE[$], Carlson CS[$]. Genetic diversity improves our understanding of complex traits. BioRxiv (2019)

LD differences across populations can pinpoint causal variants



Gen Wojcik

Wojcik GL#, Graff M#, Nishimura KK#, Tao R#, Haessler J#, Gignoux CR#, Highland HM#, Patel YM#, … Kenny EE$, Carlson CS$. Genetic diversity improves our understanding of complex traits. BioRxiv (2019)

Environmental, selection and other factors impact PRS

Science

BREVIA

Melanesian Blond Hair Is Caused by an Amino Acid Change in TYRP1

Eimear E. Kenny[1,*], Nicholas J. Timpson[2,*], Martin Sikora[1], Muh-Ching Yee[1], Andrés Moreno-Estrada[1], Celeste Eng[3], Scott ...
+ See all authors and affiliations

biobank uk

Improving the health of future generations

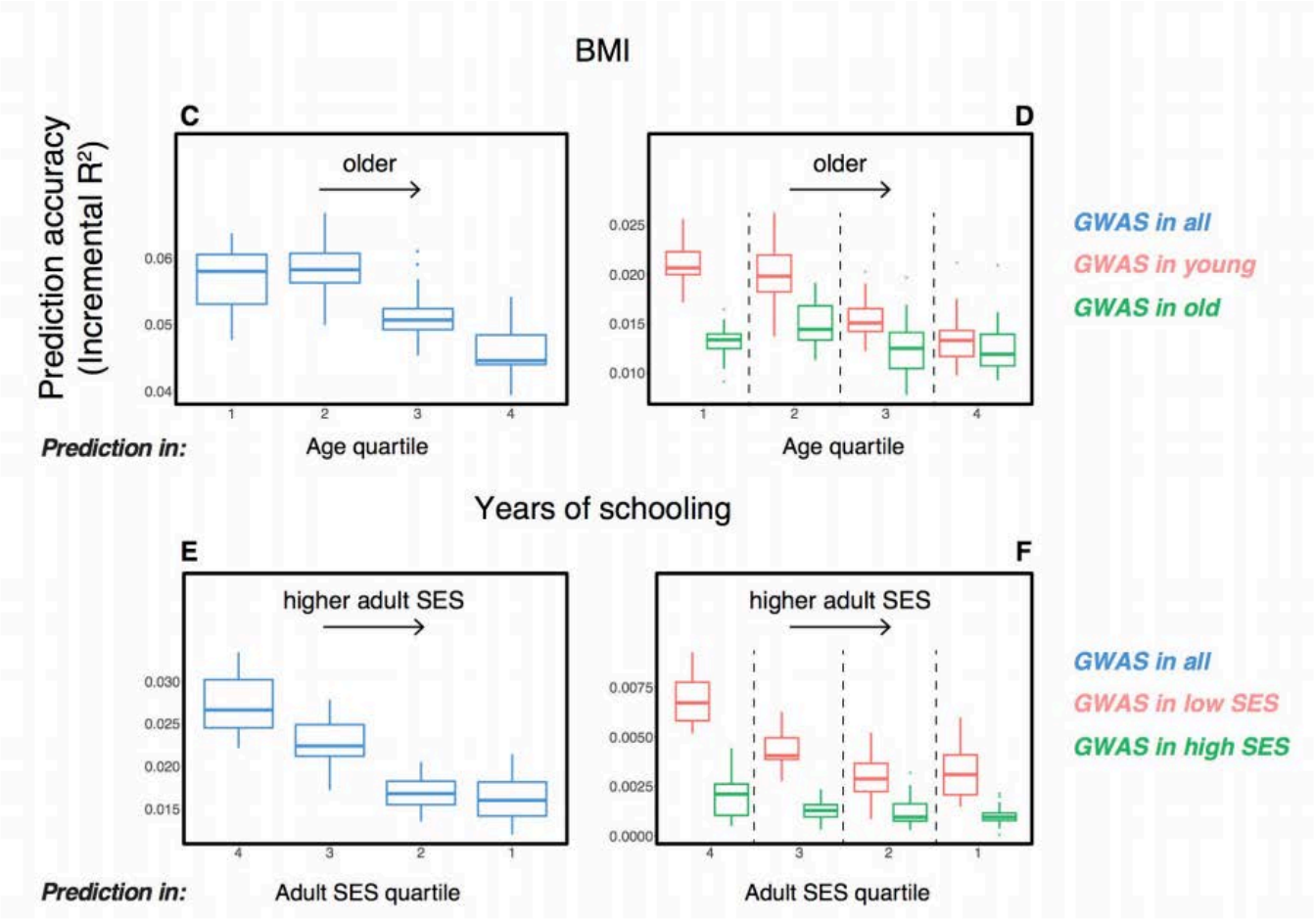
Liz Cirulli, Ph.D.
Nicole Washington, Ph.D.
Dr. Cirulli is Principal Scientist and Dr. Washington is the Associate Director of Research at Helix.
March 27, 2019

Researchers have access to new data on thousands of exomes. Here's what we found.

3. *TYRP1* variants and Blonde Hair

One of our novel findings is the association of *TYRP1* variants with blonde hair color in those of British ancestry. *TYRP1* made a big splash in 2012 when researchers found that a variant in this gene caused blonde hair in dark-skinned individuals of Melanesian ancestry from the Solomon Islands. This genetic variation is separate from the ones that are known to cause blonde hair in those of European ancestry and, until now, there has been no evidence that this gene also played a role in European hair coloring. The specific variant that causes blonde hair in Solomon Islanders, rs387907171, an arginine to cysteine substitution at amino acid position 93, was only found in 3 of the 40,648 individuals analyzed in our UK Biobank study. However, more than 30 other rare coding variants in this gene were found in 1% of the 4,671 British ancestry blonde individuals. Previous studies have shown that the Solomon Island variant is recessive,

PRS can be affected in unpredictable ways by unexpected or hidden factors



H Mostafavi, A Harpak*, D Conley, J.K. Pritchard and M Przeworski (2019) Variable prediction accuracy of polygenic scores within an ancestry group. BioRxiv*

Future directions

- PRS are gaining in accuracy, but the path to demonstrate robustness and utility is not straightforward

- Complex disease is complex, need to deepen diversity (ancestral, SES, age, etc) to understand range of genetic and environmental factors involved.

Acknowledgements

Center for Genomic Health:
Noura Abul-Husn, Clinical Director
Gillian Belbin
Gabrielle Bertier
Dean Bobo
Sinead Cullina
Sumita Kohli
Jacqueline Odgis
Jessica Rodriguez
Sabrina Suckiel
Stephane Wenric

Institute for Personalized Medicine:
Judy Cho, Director
Amanda Merkelson
Steve Ellis
Bernadette (Bibi) Liggayu
Janice Morinigo
Ben Song
Rajiv Nadukuru
Tom Kaszemacher

University of Colorado:
Chris Gignoux

Broad Institute:
Alicia Martin

Stanford University:
Genevieve Wojcik
Elena Sorokin
the Bustamante lab
Carlos Bustamante

THANK YOU
to the thousands of research participants who make this work possible.

PAGE Collaborators:
Lucia Hindorff (NHGRI)
Kari North (UNC)
Chris Carlson (Fred Hutch)
Charles Kooperberg (Fred Hutch)
Chris Haiman (MEC)
Ruth Loos (MSSM)
Tara Matise (Rutgers)
Steve Buyske (Rutgers)
Jose-Luis Ambite (USC)
Christy Avery (UNC)
Ulrike Peters (Fred Hutch)

NIH Funding:
NHGRI/U01HG109391
NHGRI/U01HG009080
NHLBI/X01HL134588
NIDDK/R01HL104608
NHLBI/R01DK110113