Preventing heart disease: with machine learning and smartphone technology

Front Row Seminar, 2023

Ali Torkamani, Ph.D.
Professor, Integrative Structural and Computational Biology
Scripps Research
Director of Genomics and Genome Informatics
Scripps Research Translational Institute
atorkama@scripps.edu
@ATorkamani

Scripps Research
Science Changing Life

Introduction to Ali

Ali Torkamani

FOLLOWING

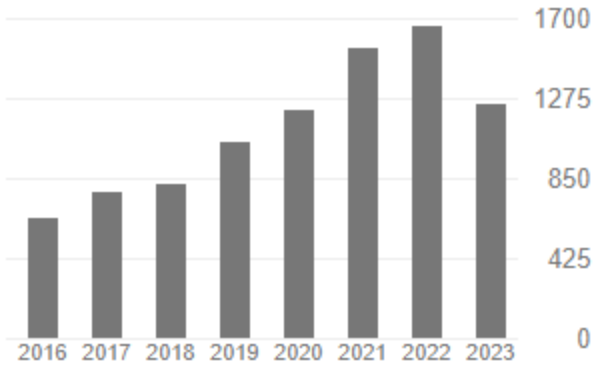
Director at SRTI, Professor at Scripps Research
Verified email at scripps.edu

Genomics Human Genetics Genome Informatics Bioinformatics Individualized Medicine

TITLE	CITED BY	YEAR
Identification of *ALK* as a major familial neuroblastoma predisposition gene YP Mossé, M Laudenslager, L Longo, KA Cole, A Wood, EF Attiyeh, ... Nature 455 (7215), 930-935	1576	2008
The personal and clinical utility of polygenic risk scores A Torkamani, NE Wineinger, EJ Topol Nature Reviews Genetics 19 (9), 581-590	1207	2018
A primer on deep learning in genomics J Zou, M Huss, A Abid, P Mohammadi, A Torkamani, A Telenti Nature genetics 51 (1), 12-18	627	2019

Cited by VIEW ALL

	All	Since 2018
Citations	12174	7586
h-index	54	44
i10-index	92	82



Free Will …… Genome Interpretation

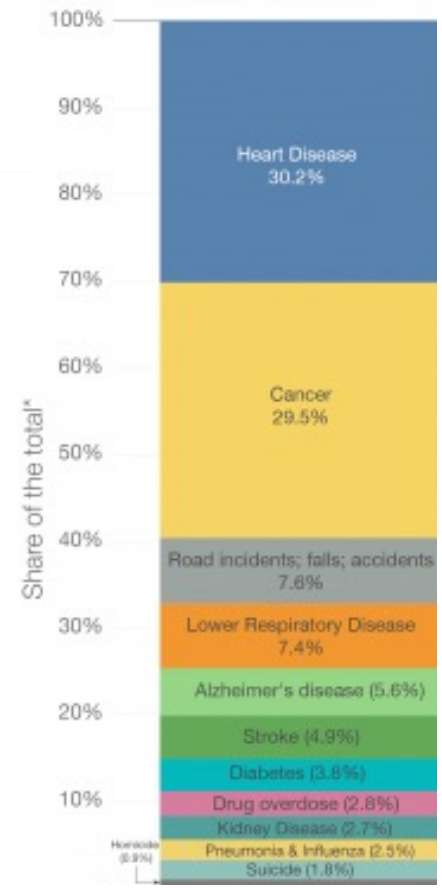
Talk Outline

- **What is Heart Disease, or Coronary Artery Disease? How to prevent it? Why are we failing?**
- **Genetic Risk and how it can be used to improve prevention.**
- **Delivery of genetic risk. MyGeneRank and PEPRS Studies.**
- **Amplifying the benefits of genetics with Machine Learning**

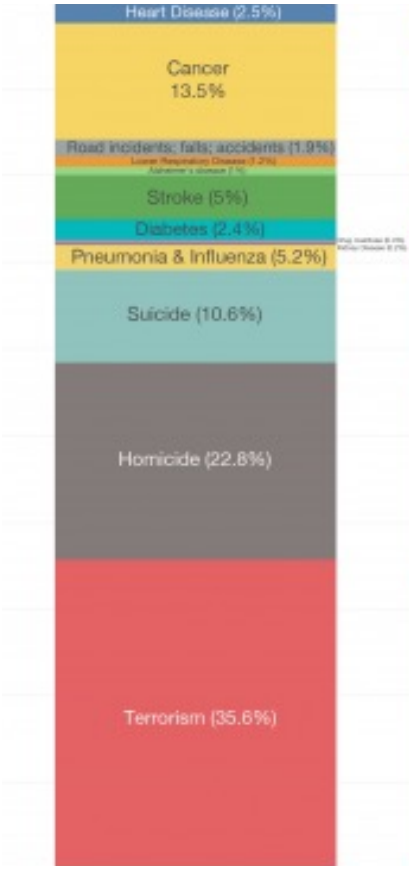
What is Heart Disease, or Coronary Artery Disease?

How to prevent it? Why are we failing?

Heart Disease: The Silent Killer



What Kills Us

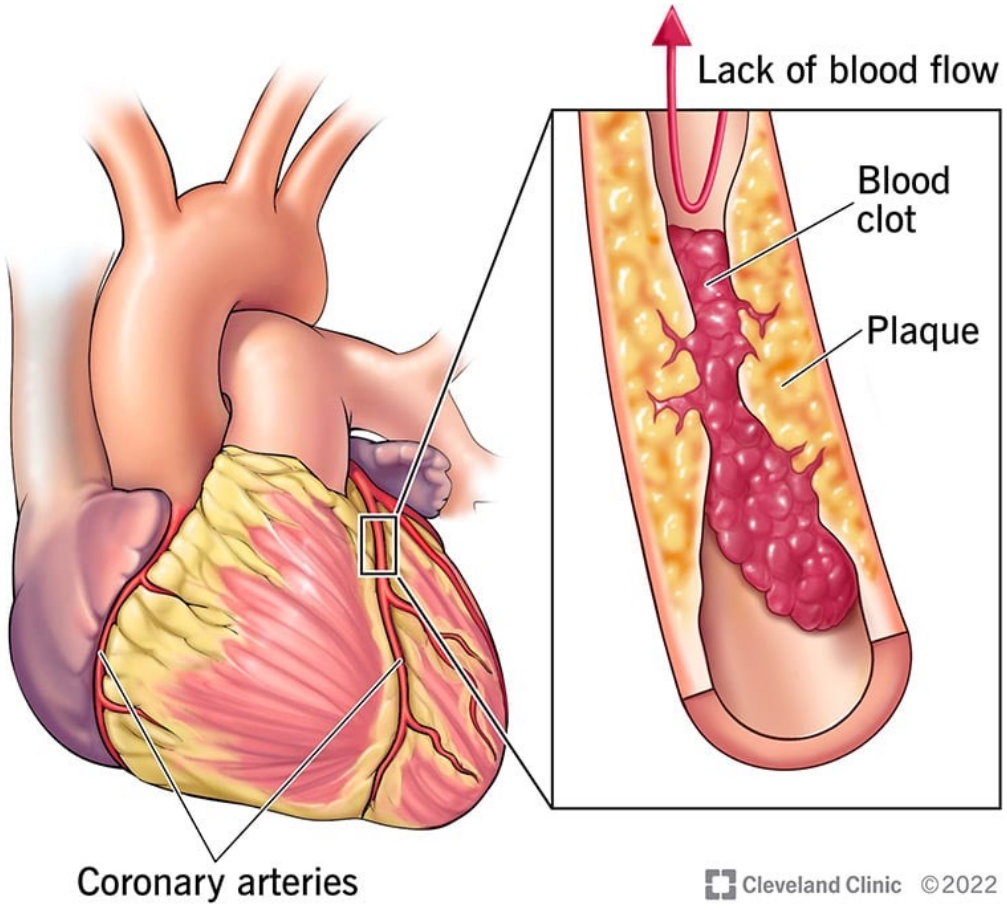


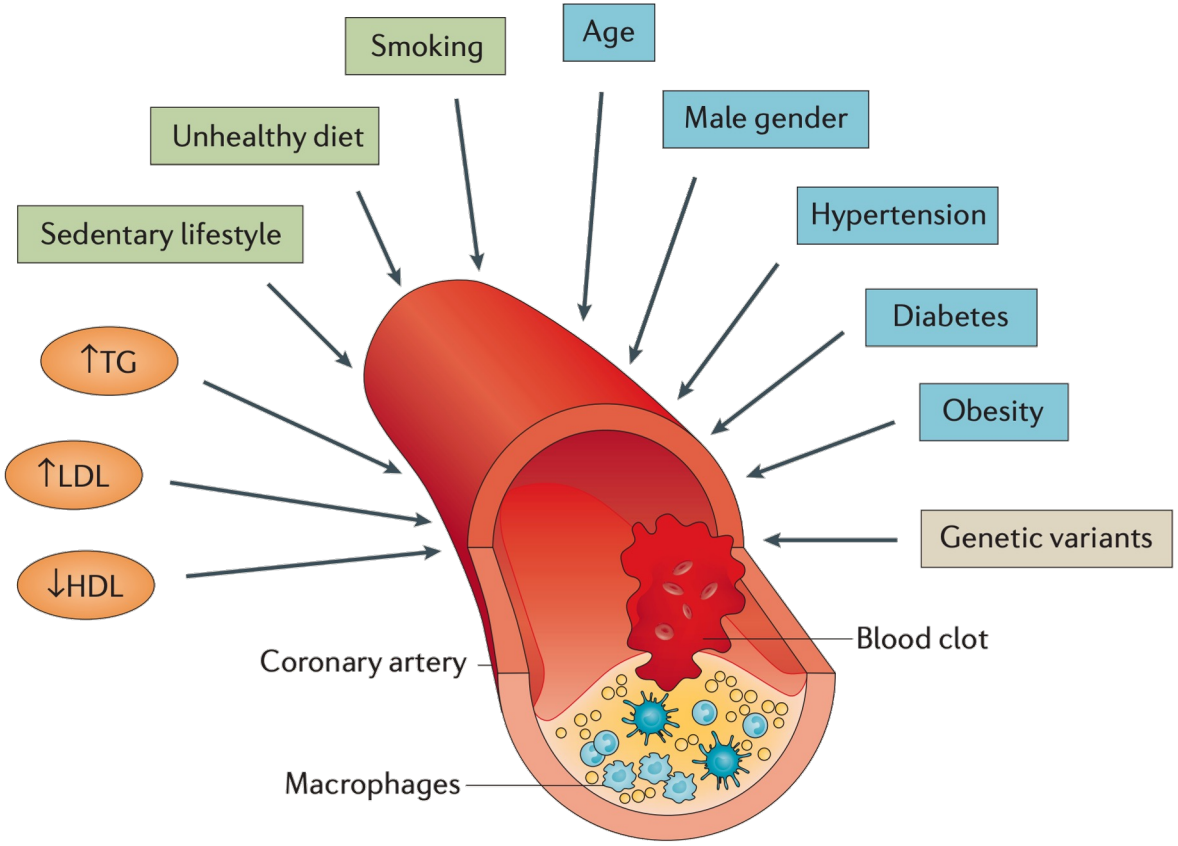
What the Media Covers

What We Google

Our world in Data

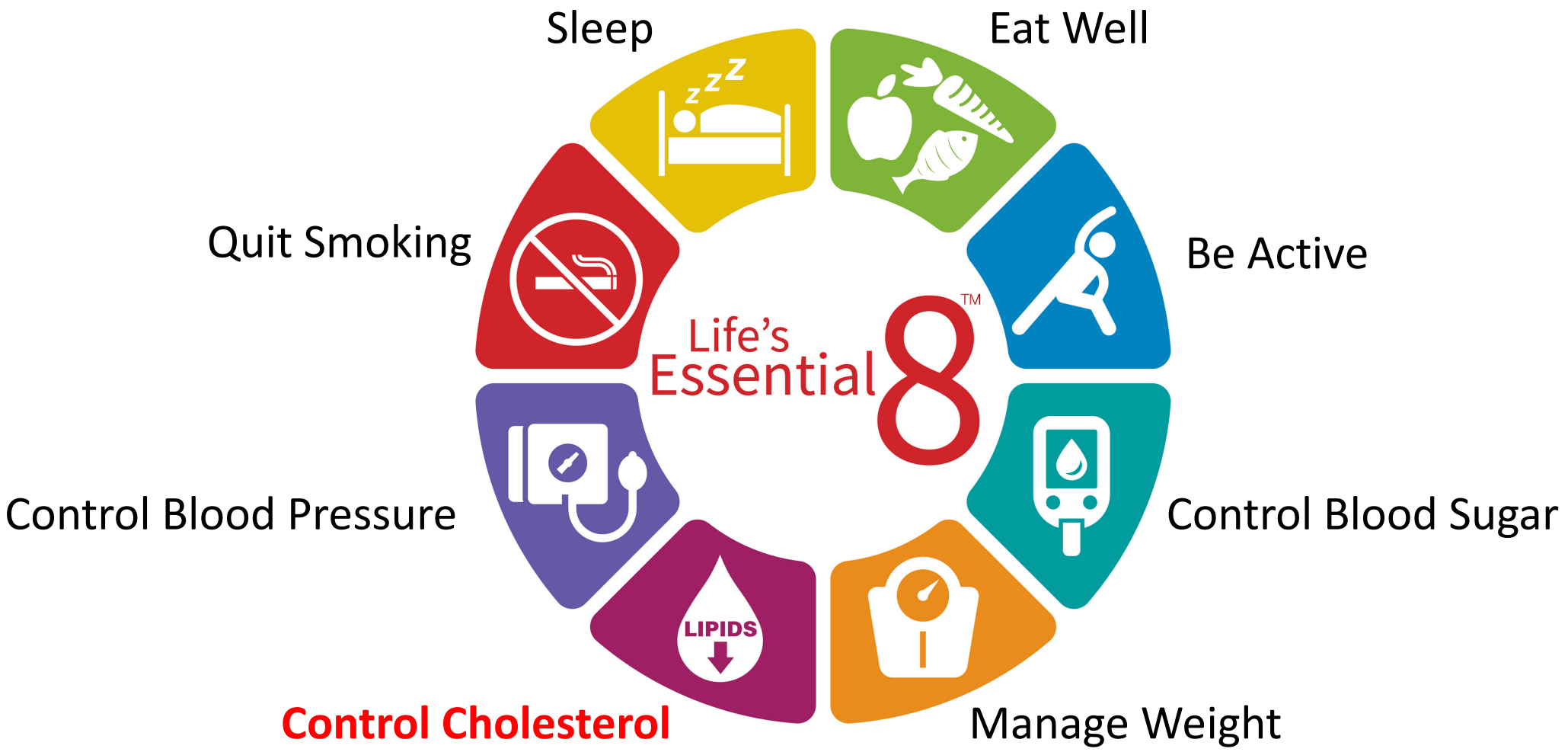
What is Coronary Artery Disease?





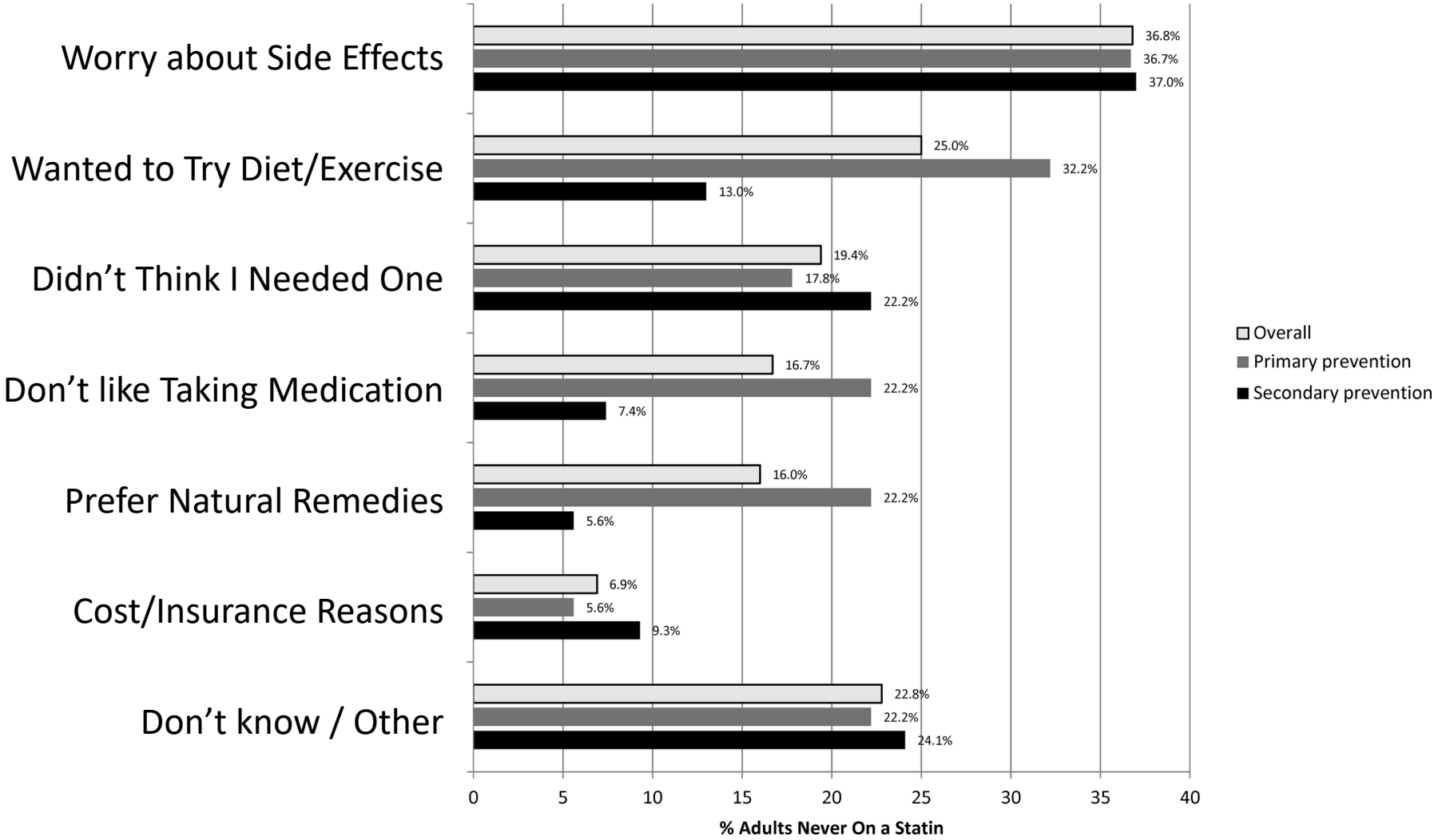
RISK FACTORS

Preventing Coronary Artery Disease



Larson, Sophia, and Eugene Yang. "Prevention Guidelines: Does one size fit all?." (2019): 2181-2183.

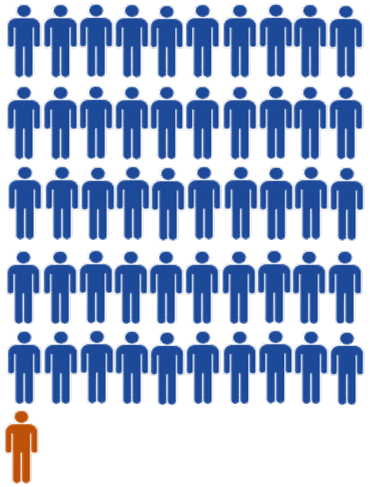
What Underlies Lack of Adherence?



Bradley et al. JAHA 2019

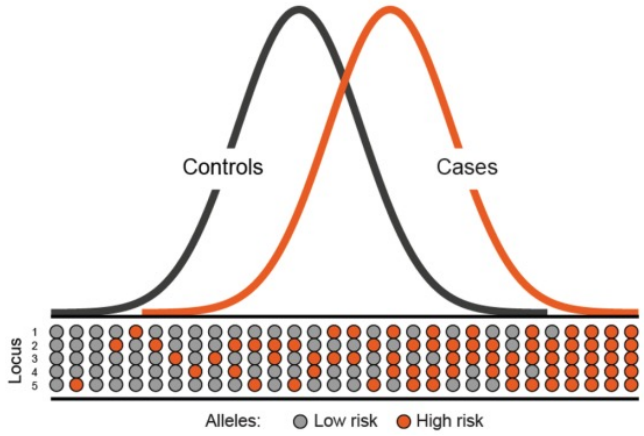
Genetics can identify those individuals who would benefit most from lipid-lowering.

ASCVD Risk (10-year risk)	Clinical Action (AHA Guidelines)
0% - 5%	Low risk
5% - 7.5%	Borderline Risk
7.5% - 20%	**Moderate intensity statin, consider risk enhancing factors**
>20%	**High-intensity statin**



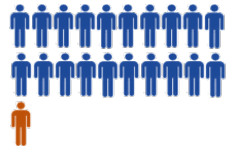
No genetics
50 : 1 treated : prevented

Genetic Risk Stratification

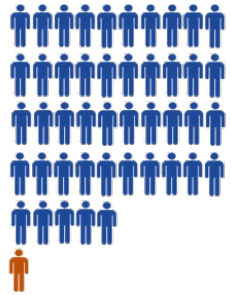


20% Low Genetic Risk
statin efficacy ↓

20% High Genetic Risk
statin efficacy ↑



high genetic risk
20 : 1 treated : prevented



average genetic risk
45 : 1 treated : prevented

low genetic risk
60 : 1 treated : prevented

statin treated

heart attack prevented

Mega et al. Lancet 2015

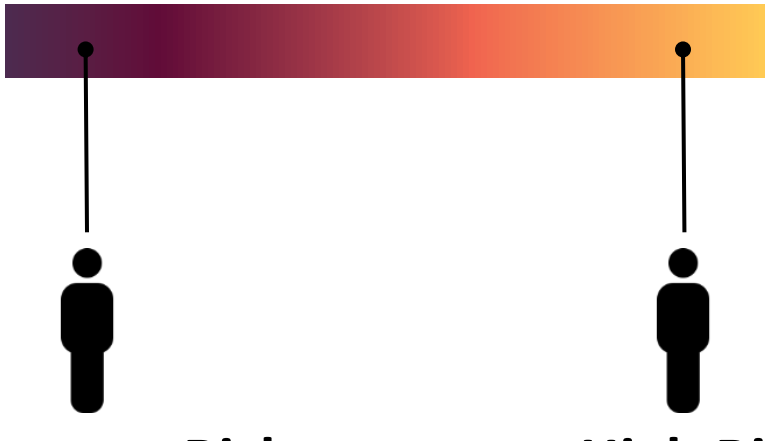
The Genetics of Coronary Artery Disease

How can genetic risk be used for prevention?

Basic Sources of Genetic Risk

Monogenic Risk

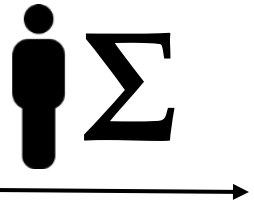
High Risk Variants
Low Heritability Explained
Binary Result



Average Risk
~99%

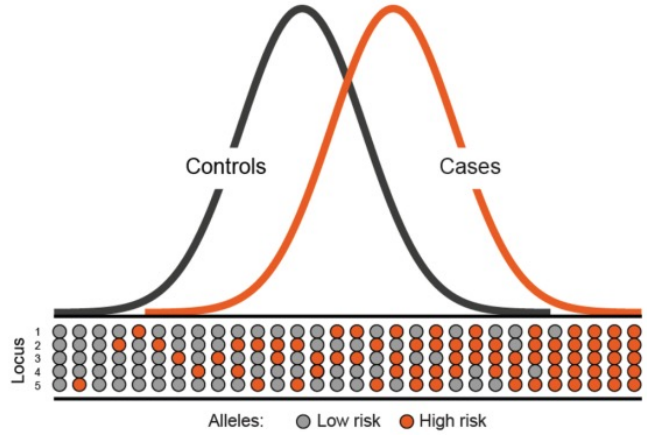
High Risk
~1%

Yes or No?



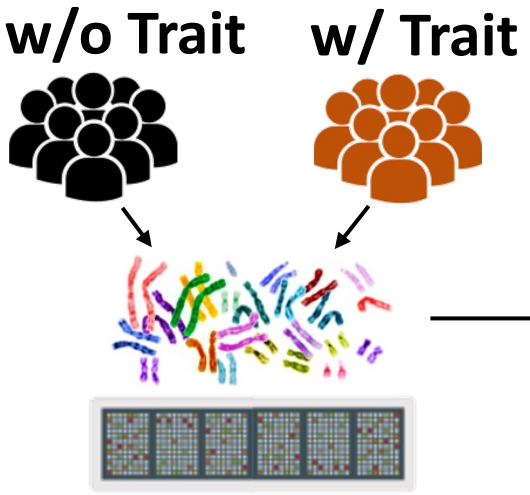
Polygenic Risk

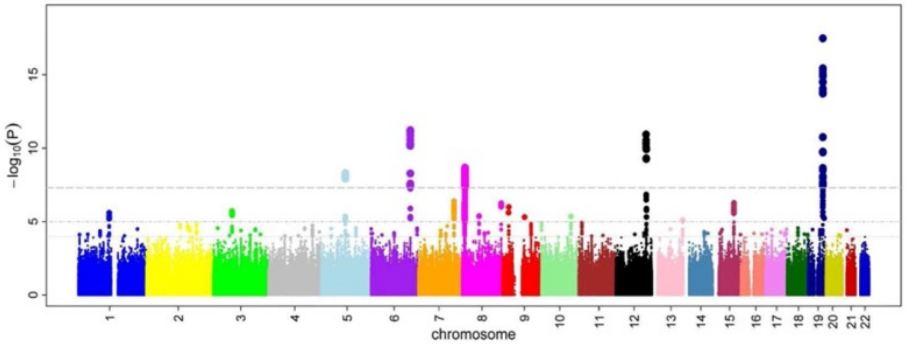
Low Risk Variants
Cumulatively Moderate Heritability
Continuous Result

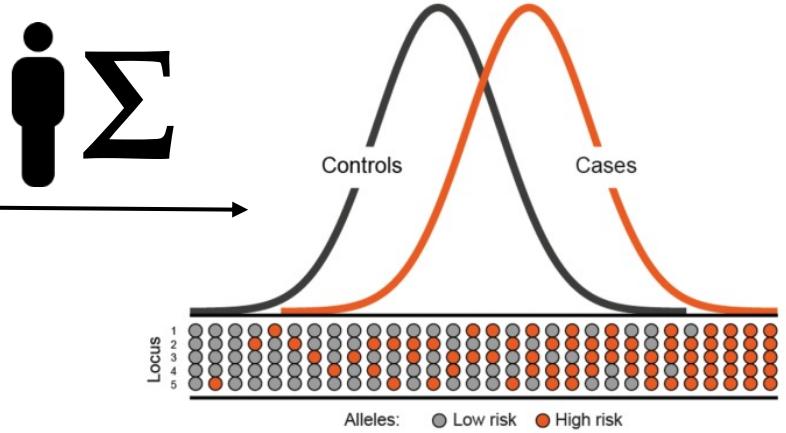


Low Risk **High Risk**

Polygenic Risk Is Defined from Data from Millions of Genomes







Design

1,000,000+ subjects

Results

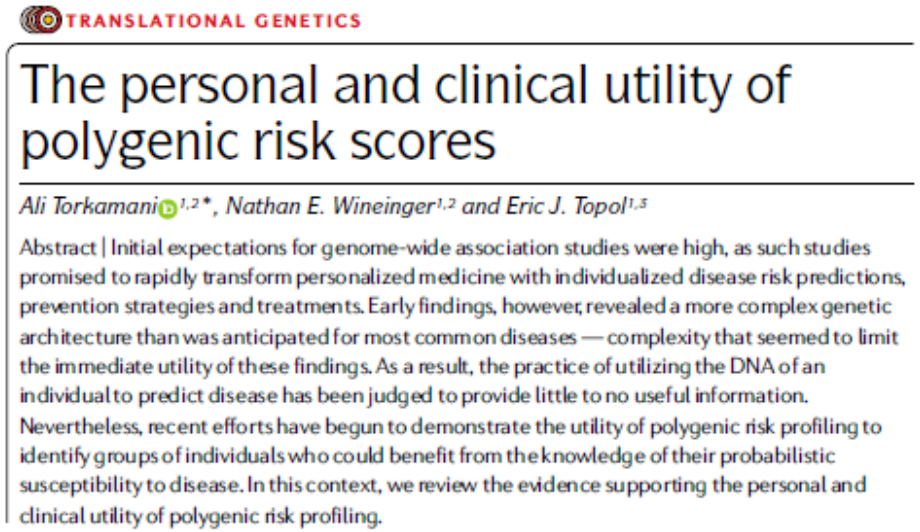
Millions of known variants

Polygenic Risk Score

Cumulative sum

Genetics is Often Not Deterministic

- **Genetically-informed Therapeutic Intervention**
  - Prioritize initiation of preventative therapy
    - e.g. lipid lowering for prevention of a first heart attack
- **Genetically-informed Disease Screening**
  - Accelerate traditional screening or initiate non-invasive screening
    - e.g. – accelerate mammography screening or circulating DNA testing
  - Improve the risk:benefit ratio of screening tests with no net benefit when applied to the general population
    - e.g. PSA testing, glaucoma screening
- **Genetically-informed Life Planning**
  - Disease progression / severity informed by PRS
    - e.g. neurodegenerative / neuropsychiatric conditions


REVIEWS

TRANSLATIONAL GENETICS

The personal and clinical utility of polygenic risk scores

Ali Torkamani[1,2], Nathan E. Wineinger[1,2] and Eric J. Topol[1,3]*

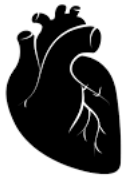
Abstract | Initial expectations for genome-wide association studies were high, as such studies promised to rapidly transform personalized medicine with individualized disease risk predictions, prevention strategies and treatments. Early findings, however, revealed a more complex genetic architecture than was anticipated for most common diseases — complexity that seemed to limit the immediate utility of these findings. As a result, the practice of utilizing the DNA of an individual to predict disease has been judged to provide little to no useful information. Nevertheless, recent efforts have begun to demonstrate the utility of polygenic risk profiling to identify groups of individuals who could benefit from the knowledge of their probabilistic susceptibility to disease. In this context, we review the evidence supporting the personal and clinical utility of polygenic risk profiling.

Read more: *Torkamani, et al. 2018*

Lower Polygenic Risk is Associated with Healthy Aging

Wellderly: 80+ yrs old, no common diseases

Decreased Risk



Coronary Artery Disease

Alzheimer's Disease

Erickson et al. Cell

Delivery of Genetic Risk

Can genetic information drive initiation and adherence to prevention?

Two Smart-Phone Approaches

MyGeneRank

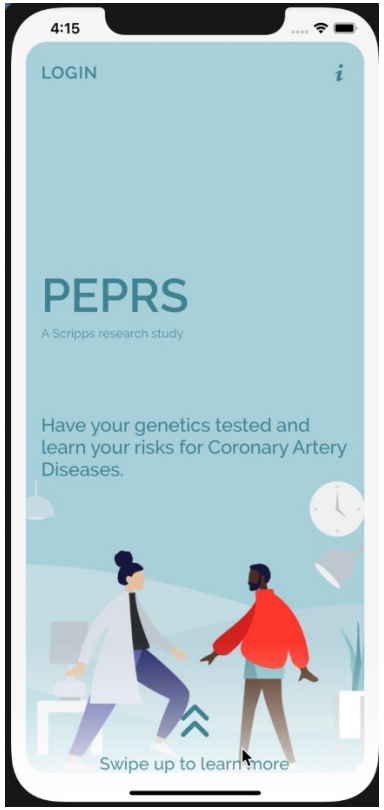
- Open to all comers
- Basic inclusion criteria
- Bring your own genetic data

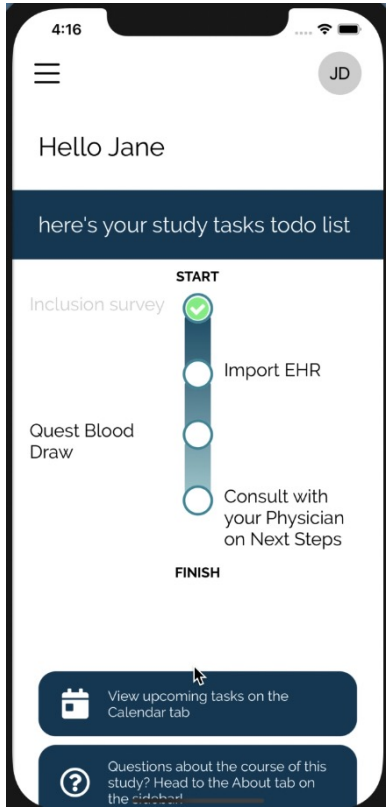
mygenerank.scripps.edu

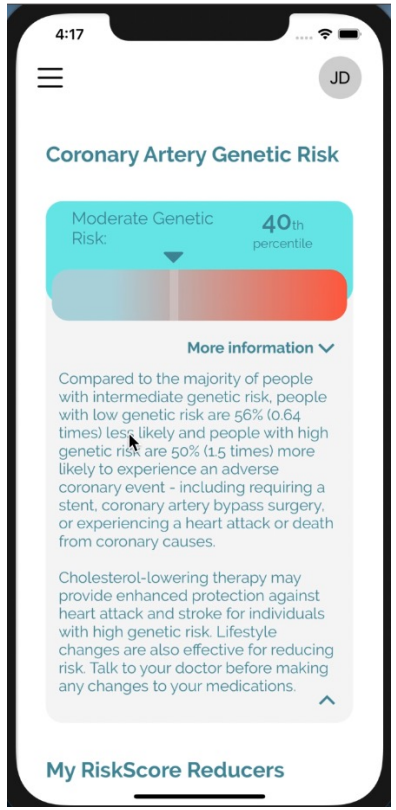
PEPRS (prospective electronic polygenic risk study)

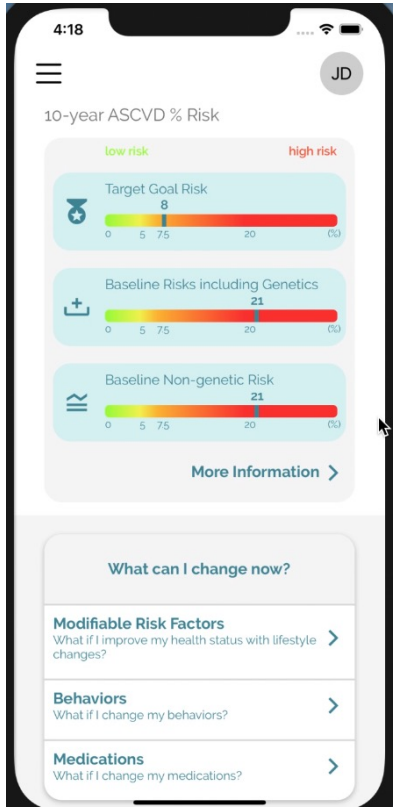
- Closed study
- Strict inclusion criteria
- De-novo data generation

First in Class App Design









Key Capabilities:

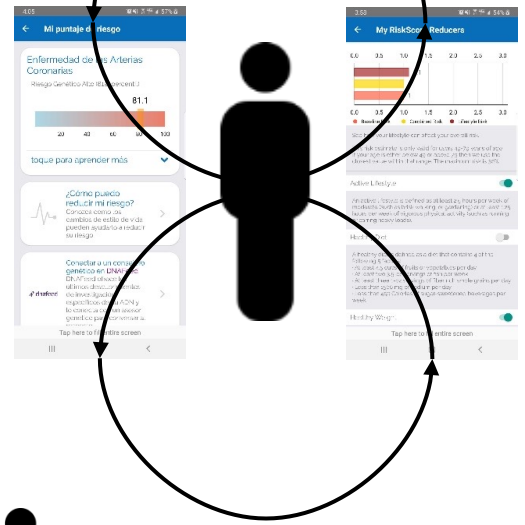
- Electronic Inclusion Criteria Screen
- Electronic Informed Consent
- Privacy Preserving Technology
  Biometric / PIN lock
- SMART FHIR EHR Data Collection
  Objective Health Data - ONC's Cures Act Final Rule
- Mobile Health Data Collection
  Objective behavioral data - Apple Health & Google Fit
- Quest Dx Patient Service Center
  API Integration for Low-Pass Sequencing
- Interactive Results Viewing
  Study Subject + Physician / Genetic Counseling Views
- Electronic Surveys
  Self-reported Health, Behavior, and Psychosocial Data
- Risk Calculation, Dynamic Interaction, and Counseling

MyGeneRank Study

Real-World Response to Polygenic Risk

MyGeneRank Experience – Participant Centric

Individual Education & Self-Efficacy



Health Data

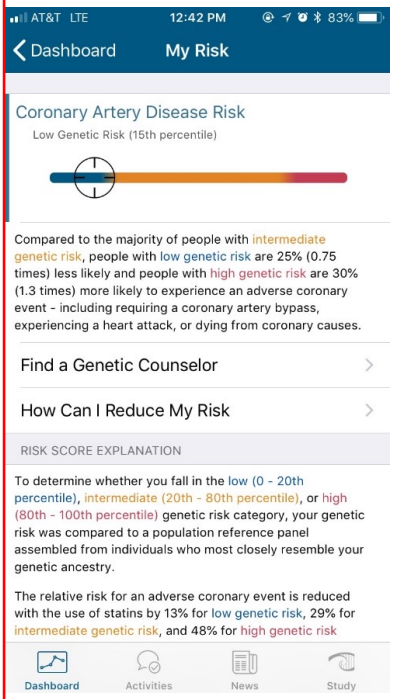
EHR Health Info Exchange, and;
Mobile Health Data

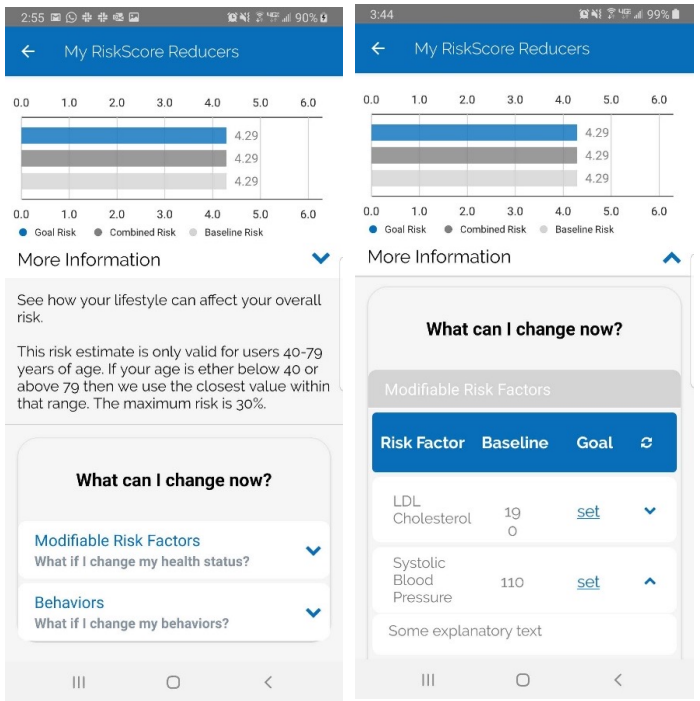
Pre-existing DNA Data, or;
De-centralized Collection

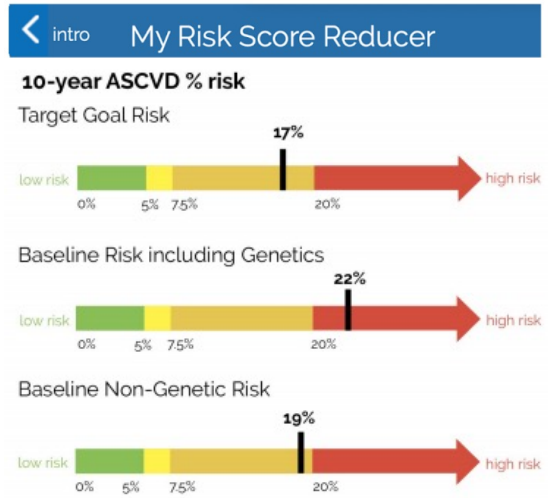
Provider Education, Confidence, & Utility

Response to CAD Polygenic Risk

Preliminary Results







Baseline

Spanish, Virtual Genetic Counseling

Improved Risk Reduction UI

Improved Source & Degree of Risk Interpretability

mygenerank.scripps.edu

MyGeneRank Study Flow and Participants

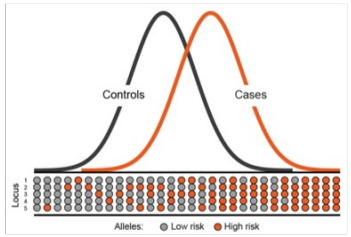
Download MyGeneRank → **Digital Informed Consent** → **Link Genetic Data: 23andMe API** → **In App Baseline Health Survey**

3,800
valid participants

3,484 total
2,928 before December 2019
556 after December 2019

In App Return of Results → **In App Psychosocial Survey** → *6-Months* → **In App Follow-up Health Survey** → **Email-based Follow-up Health Survey**

Polygenic Risk Stratification



20% Low Genetic Risk
statin efficacy ↓

20% High Genetic Risk
statin efficacy ↑

1,053 total
727 before December 2019
326 after December 2019

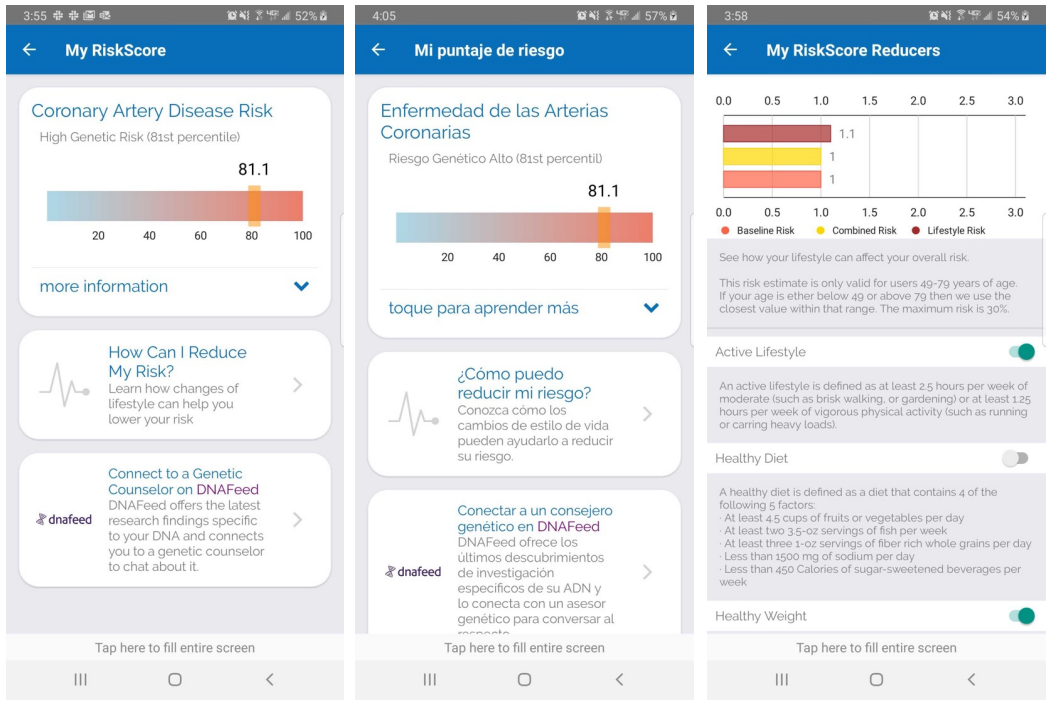
253 total
210 before December 2019
43 after December 2019

459 total
0 before December 2019
459 after December 2019

Read more: *Muse, et al. 2022 npj Digital Med*

MyGeneRank Outcomes

MyGeneRank



>2X RATE OF STATIN INITIATION

- Among individuals identified as at high genetic risk of developing coronary artery disease vs. individuals identified as low risk

10 YEARS EARLIER

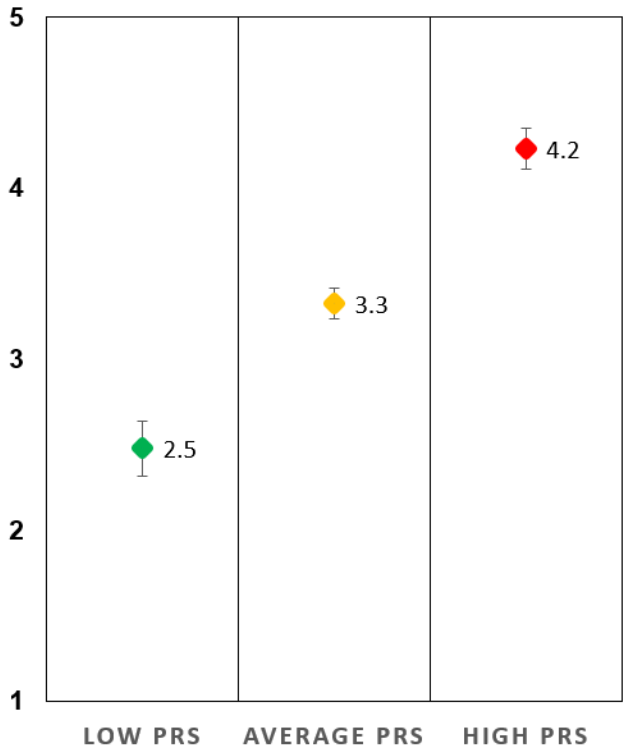
- The average age of the individuals initiating statins was 55 years old in the high genetic risk group vs 65 years old in the low genetic risk group.

Muse, et al. 2022 npj Digital Med

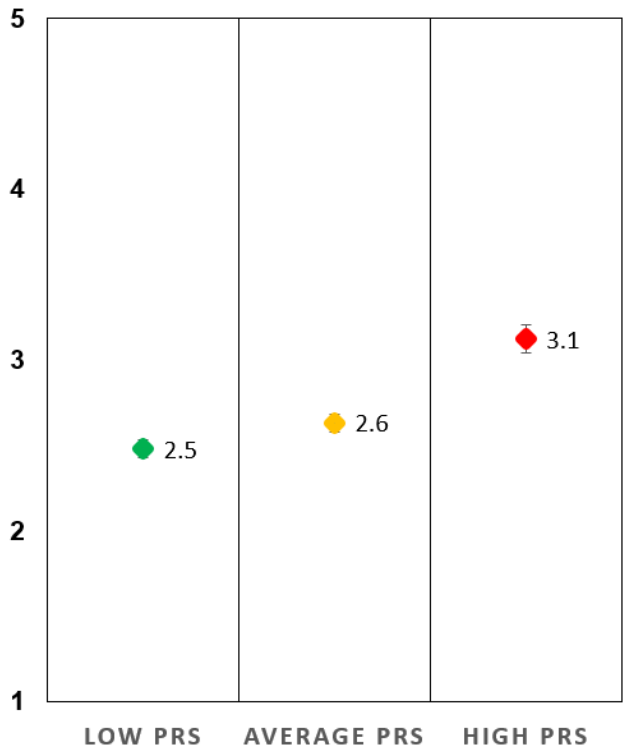
Combined initiation rate of 20% (n = 19 of 95) of high genetic risk vs 7.9% (n = 8 of 101) of low genetic risk individuals. P-value = 0.002

Participant Reactions

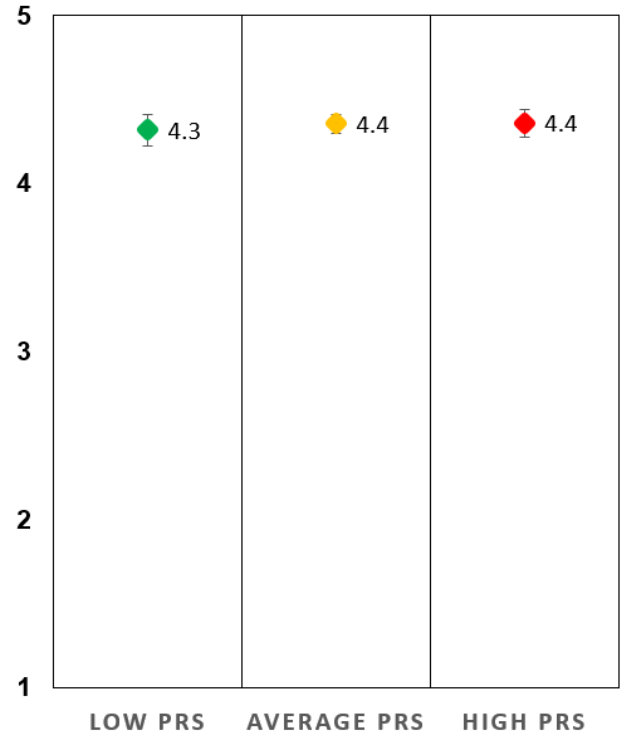
My genetics make it more likely that I will develop Coronary Artery Disease.



I worry a lot about developing Coronary Artery Disease.



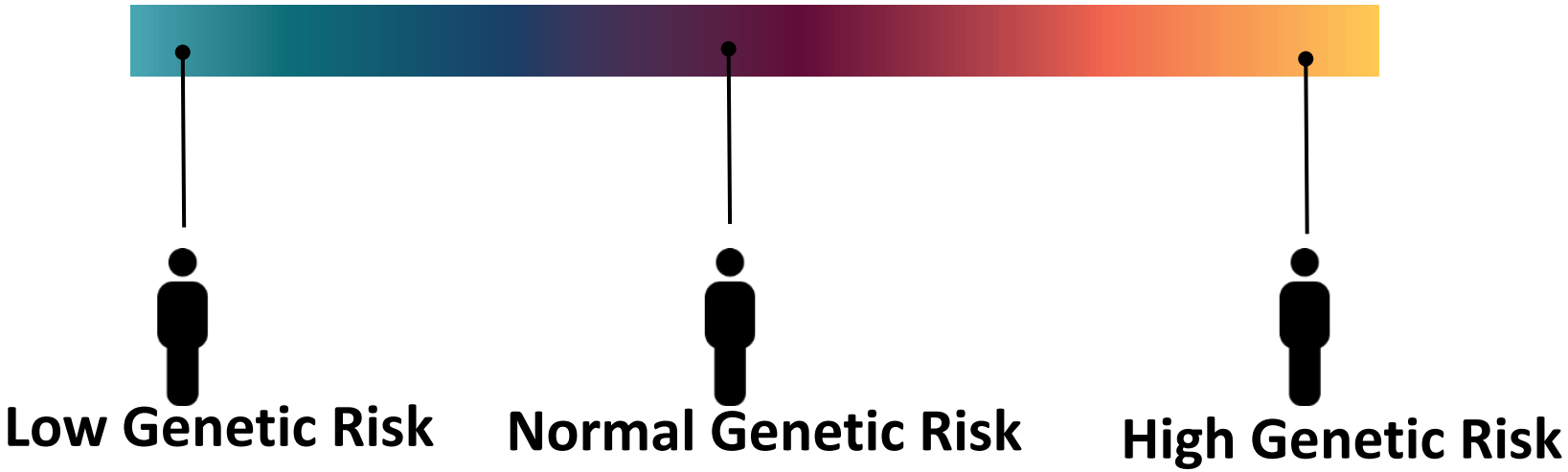
I am able to reduce my risk for developing Coronary Artery Disease.



n = 300+

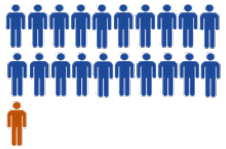
Likert scores: strongly agree (5) – neutral (3) – strongly disagree (1) mean score w/ standard error

We are ignoring useful genetic risk information!

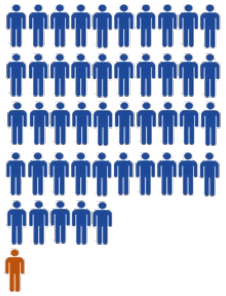


Low Genetic Risk Individuals

<1% familial hypercholesterolemia
10 : 1 treated : prevented

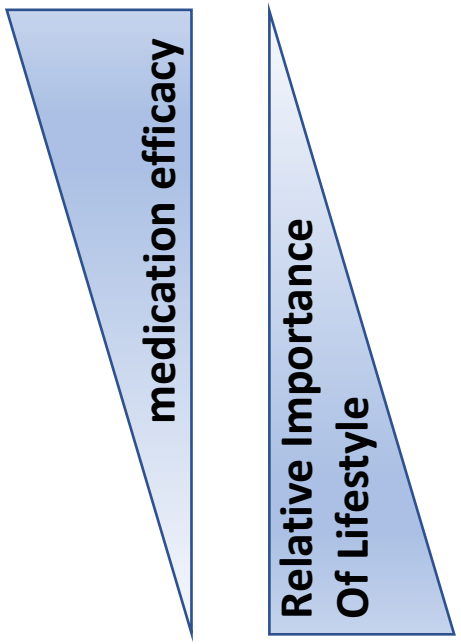


20% high polygenic risk
20 : 1 treated : prevented



60% intermediate polygenic risk
45 : 1 treated : prevented

20% low polygenic risk
60 : 1 treated : prevented



Building the Evidence for Broad Adoption

PEPRS (Prospective Electronic Polygenic Risk Study)

Key Questions:

Does polygenic risk information change physician clinical-decision making?

Does change in preventive behaviors in response to polygenic risk persist?

Are health outcomes improved as a result?

PEPRS Studies and Funding Opportunities

Phase 1 –
1,000 Participants

Local e-Recruitment

Physician Outcomes:
- *Perceived Utility*
- *Confidence in Use*
- *Actions Attributable to Use*
- *Establish Physician Education*

Randomized by Nature:
Degree of CAD Risk

Soft Patient Outcomes
- *Statin Initiation / Intensification*
- *Statin Adherence*
- *LDL Levels*
- *Test Study Logistics*

VoLoFoundation

Phase 2 –
10,000 Participants

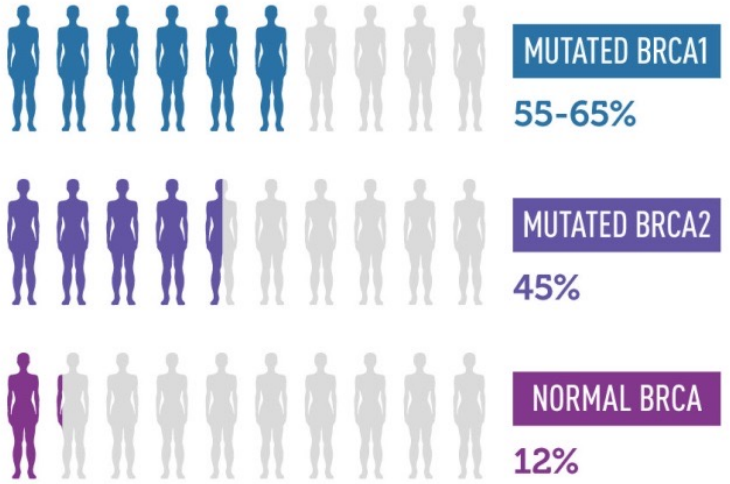
Remote e-Recruitment

Rapid Scale-Up:
- *Multi-modal Campaign*
- *Research Network Resources*
- *Long-Term Follow-Up*

Randomized Study Arm:
CAD vs Glaucoma

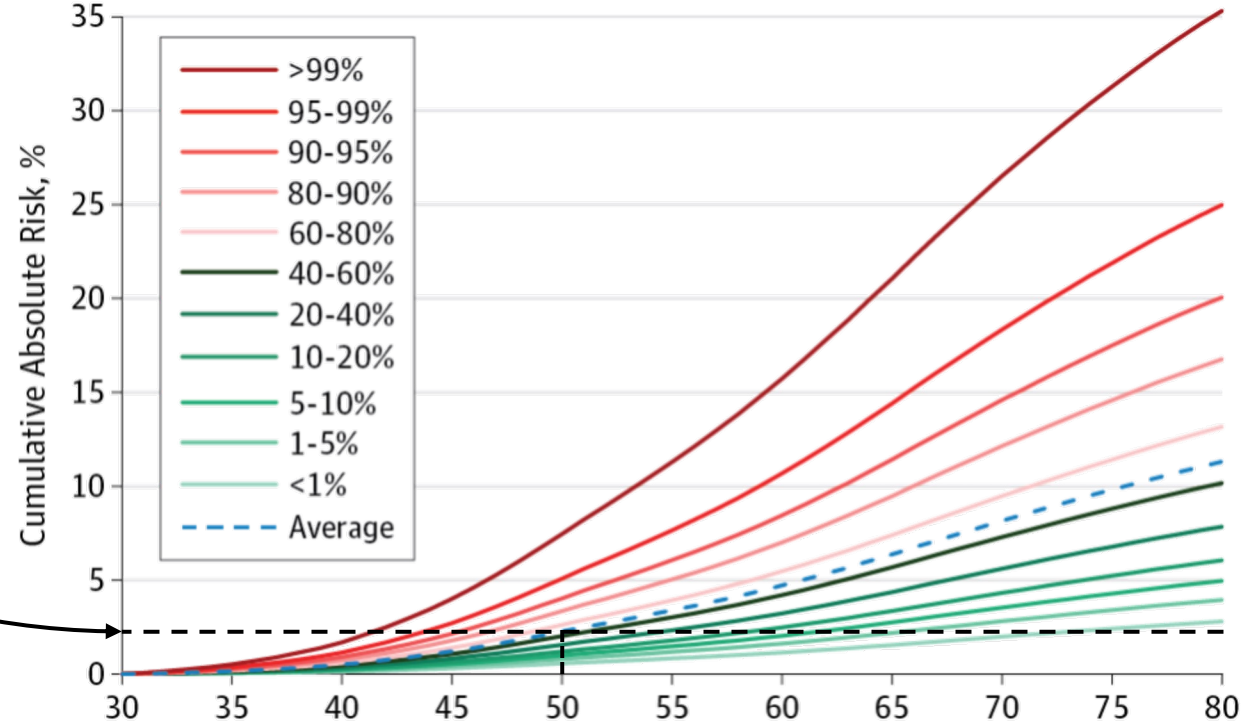
Hard Patient Outcomes
- *Heart Attacks & Death*
- *Glaucoma Diagnosis*

MyGeneRank Expansion Opportunity: Breast Cancer Risk Stratification



5% - 10% of population

90% - 95% of population



USPSTF recommends the initiation of biennial screening mammography for women at 50 years of age. In ten years:

- **41.6% false positive recall**
- **5% false positive biopsy**

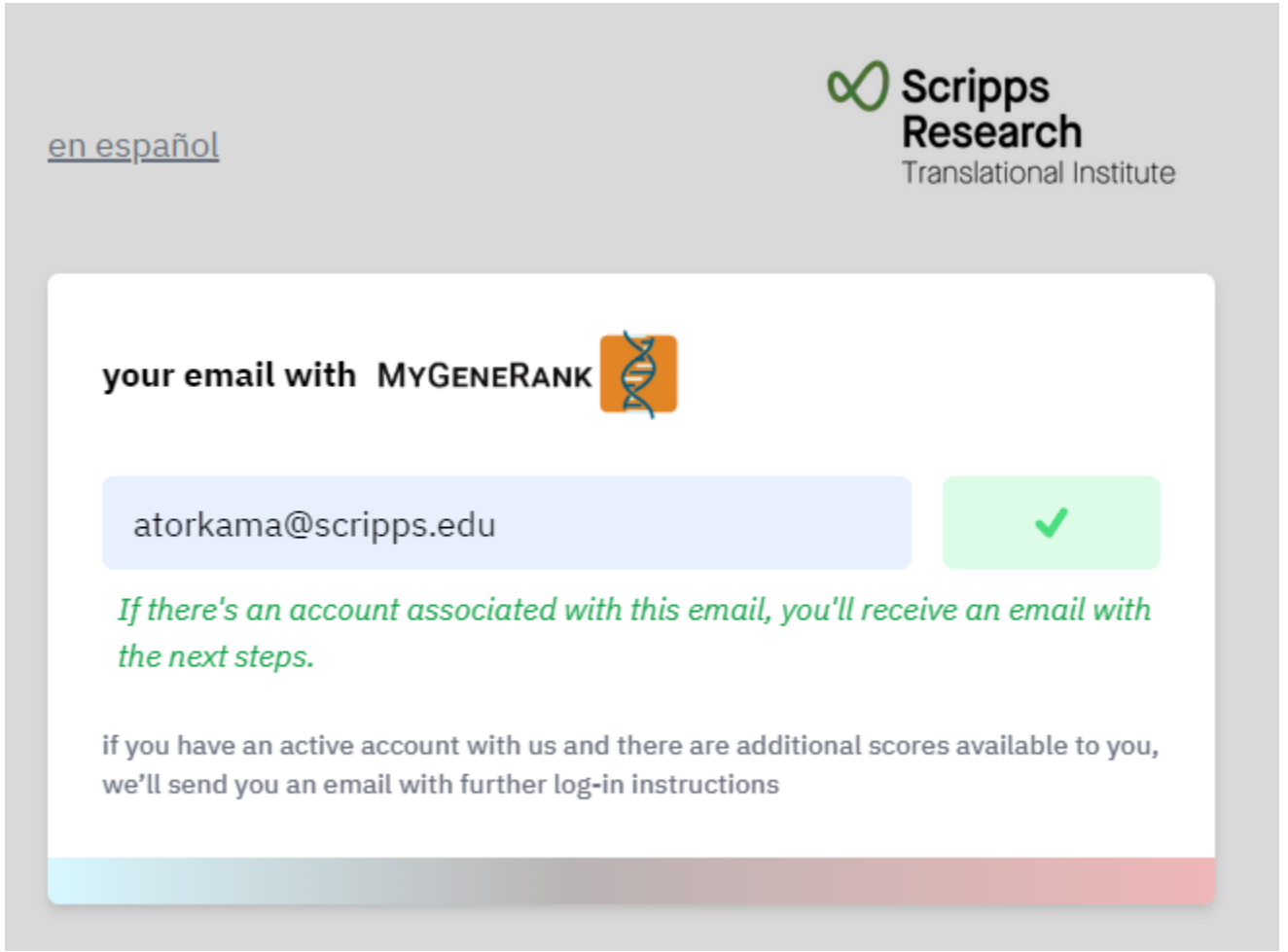
Risk-based initiation of screening mammography would allow:

- **Accelerated screening: 16%** of the population at **40 years** of age have risk that is **higher** than that of an average **50-year-old**
- **Delayed screening: 32%** of the population at **50 years** of age have risk that is **lower** than that of an average **40-year-old**

(Screening for breast cancer: US Preventive Services Task Force recommendation statement. Ann. Intern. 2016).

(Maas et al. JAMA Oncology 2018)

MyGeneRank Expansion Coming Soon

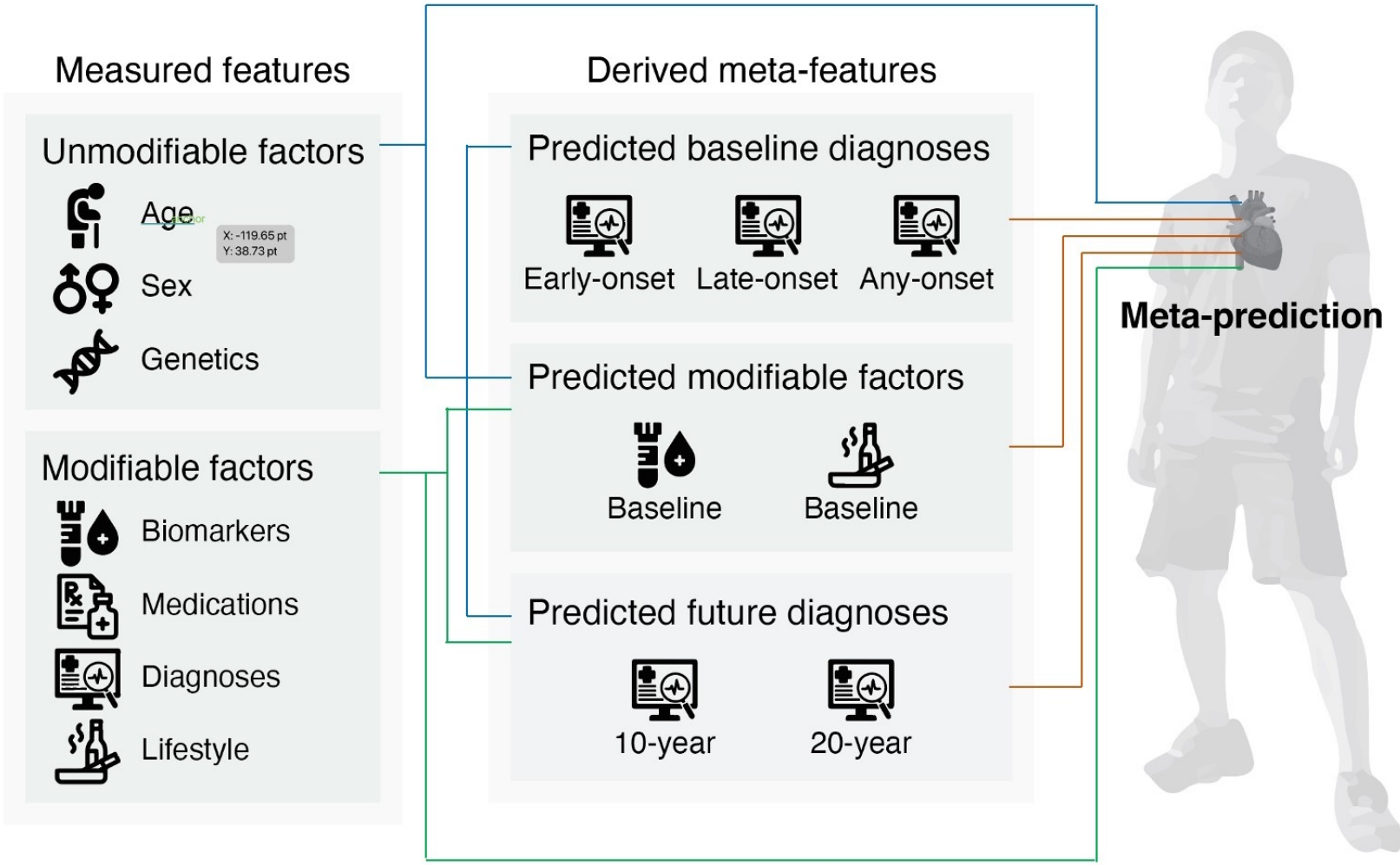


Amplifying the Benefits of Genetics with Machine Learning

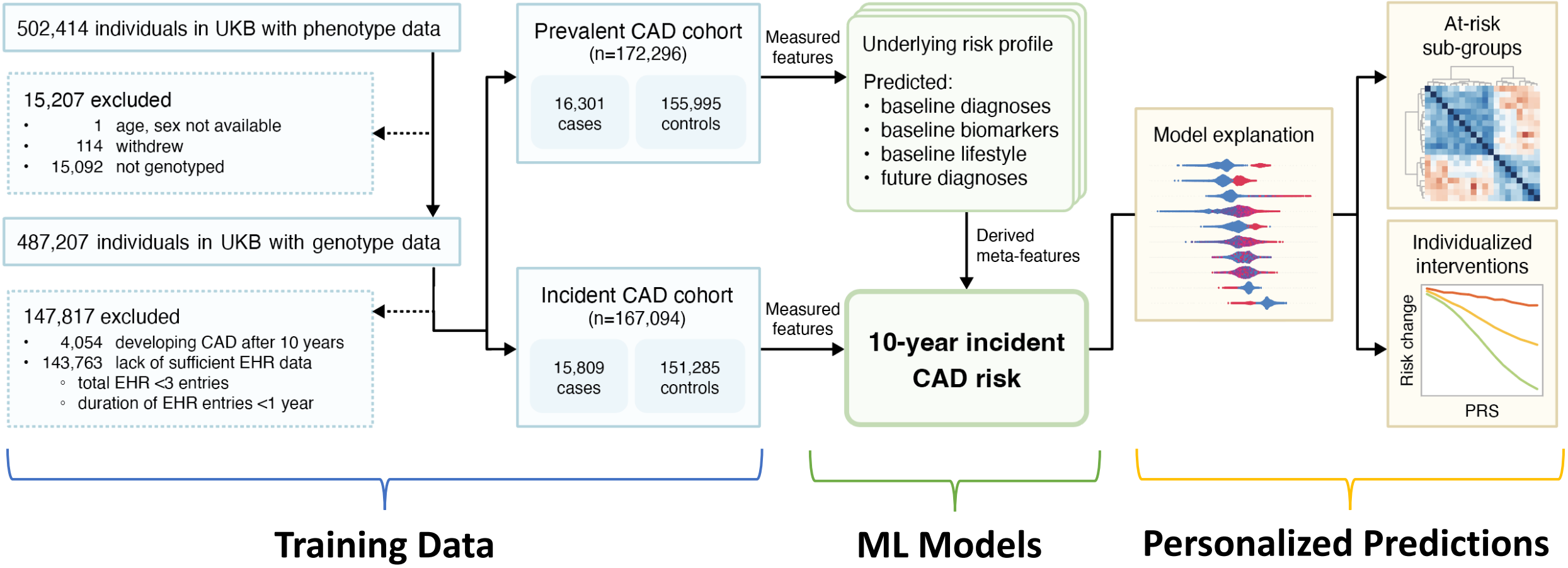
Scripps Research

Science Changing Life

Genetically Informed Meta-Prediction



Training on >500K Genomes + Phenomes



Feature curation


Evan D. Muse


Sang Eun Lee


Jun-Bean Park


Ahmed Khattab

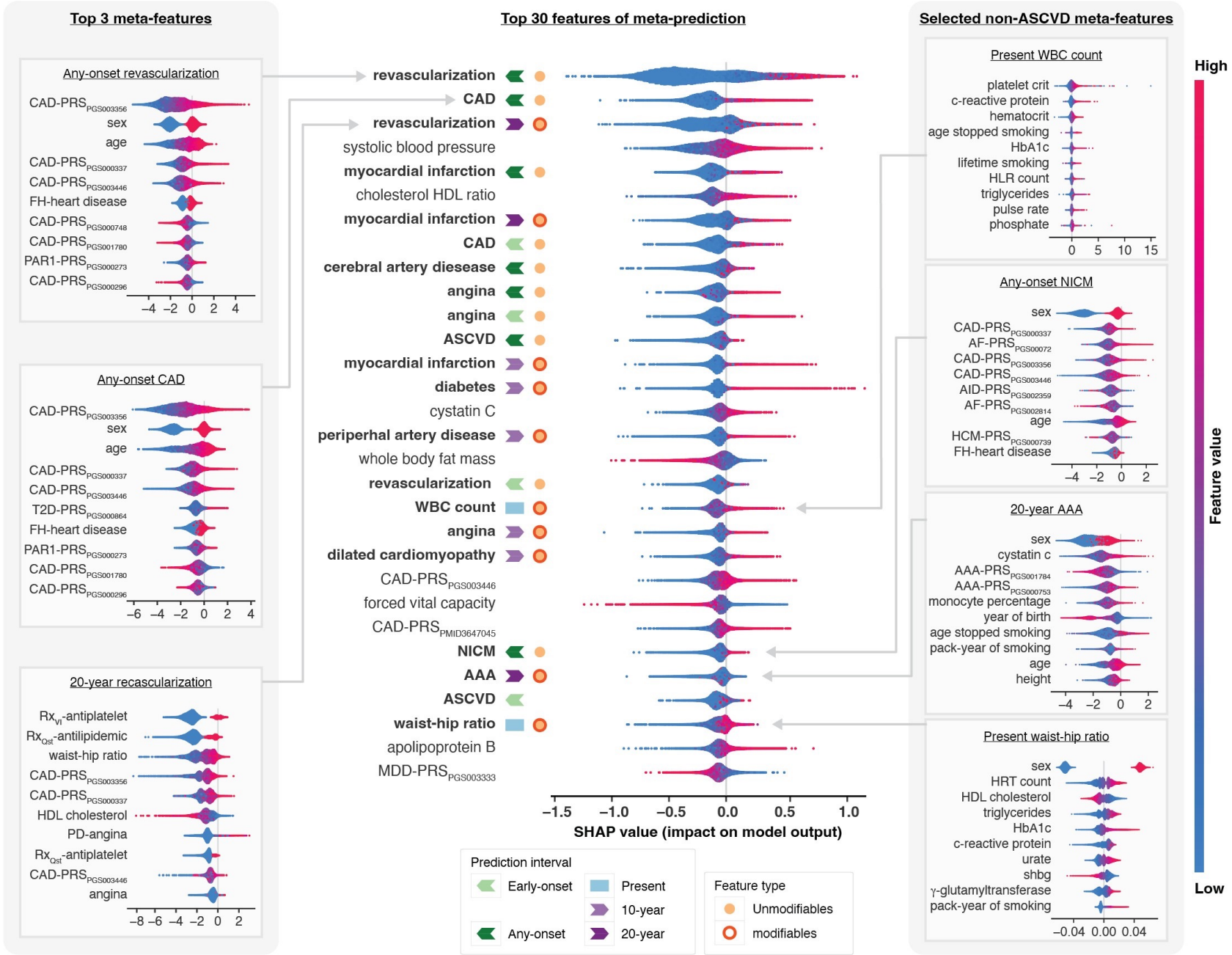

Corneliu Henegar

Type	Count	Example
Unmodifiable		
Sociodemographic	*36*	Age, Sex, Ethnic background, Smoking status, Income
Genetic ancestry	5	AFR, AMR, EAS, EUR, SAS
Family history	37	Heart disease, Stroke, High blood pressure, Diabetes, Alzheimer's disease, Parkinson's disease
PRS	1093	<< non-UKBB-derived PRS from PGS Catalog† >>
Modifiable		
Medications	24	Antihypertensive, Antiplatelet, Antidiabetic, Insulin, Lipid regulating, Statin, Steroids, etc.
Physical measurements	22	Standing/seated/sitting height, Weight, Waist/hip circumference, Systolic/diastolic blood pressure, etc.
Biomarker	63	Albumin, Alkaline phosphatase, Alanine aminotransferase, Apolipoprotein A, Apolipoprotein B, Aspartate Aminotransferase, Direct bilirubin, Urea, Calcium, Total Cholesterol, Creatinine, Cystatin C, Glucose, HbA1c, HDL cholesterol, LDL cholesterol, Total bilirubin, Triglycerides, SHBG, etc.
Diagnosis	31	Coronary artery disease, Atrial fibrillation, Atherosclerotic cardiovascular disease, Angina, Abdominal aortic aneurysm, Stroke, Ischemic stroke, Myocardial infarction, Revascularization, Heart failure, Nonischemic/dilated/hypertrophic cardiomyopathy, Peripheral artery disease, Preterm delivery, Chronic kidney disease, Migraine, Rheumatoid arthritis, Systemic lupus, Fetus intrauterine growth retardation, Gestation hypertension, Gestation diabetes, Polycystic ovary syndrome, Type 1&2 diabetes, Mental illness, Erectile dysfunction, etc.
Diet	17	daily vegetable/ fruit/ fish/ fresh meat/ whole grain/ refined grain/ processed meat intake compliance, etc

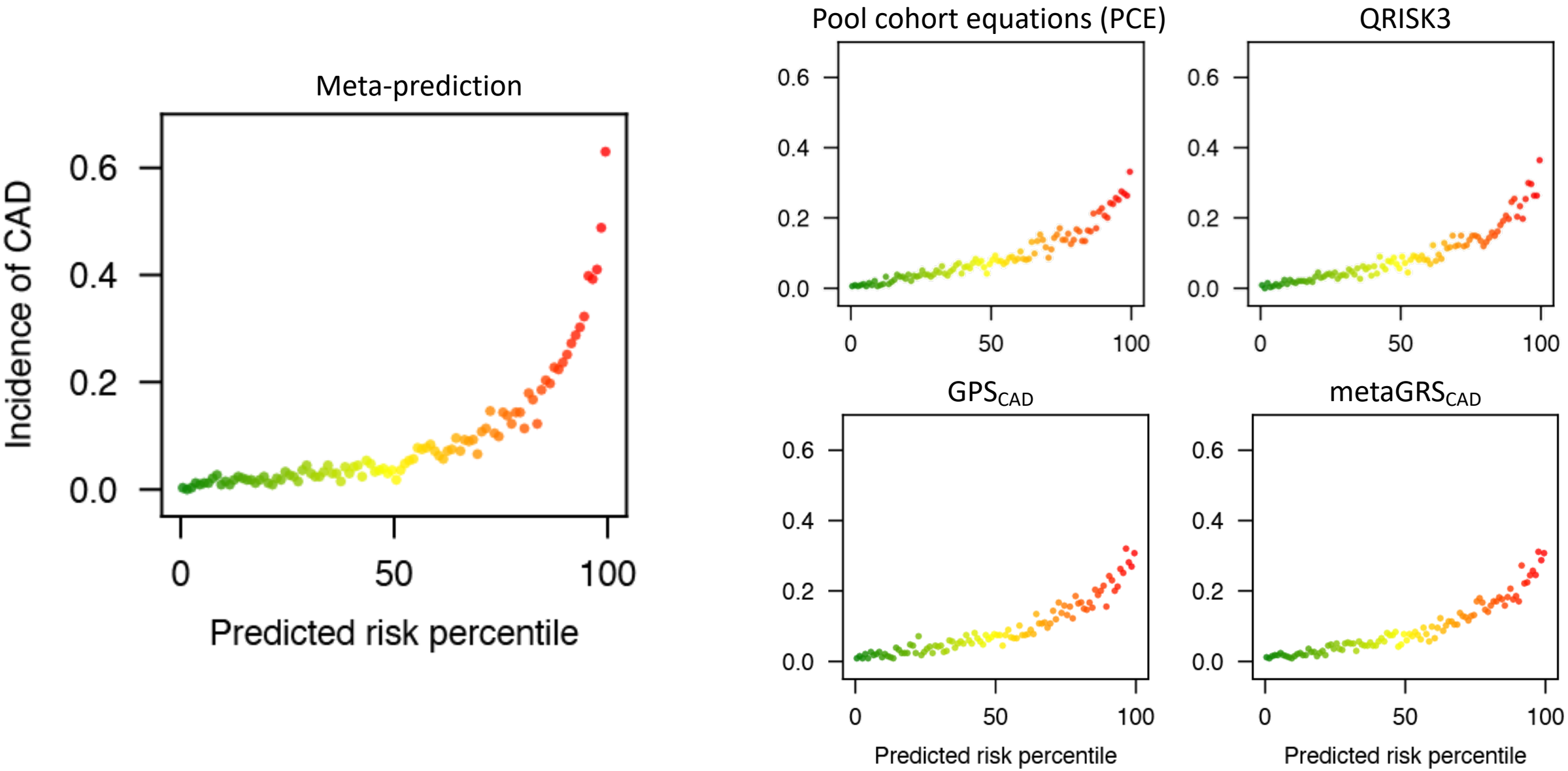
†The Polygenic Score (PGS) Catalog (https://www.pgscatalog.org/)

Most Important For Prediction?

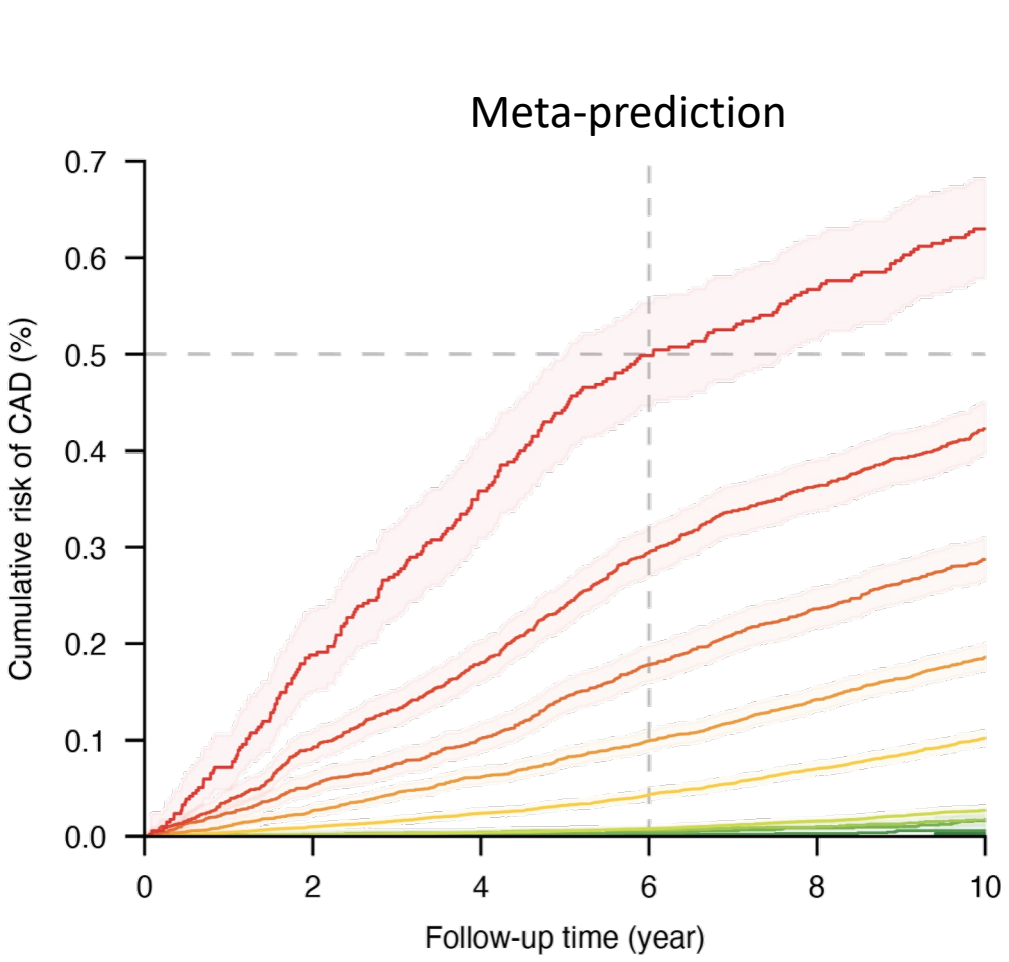
Models Predicting Past Events from Genetics

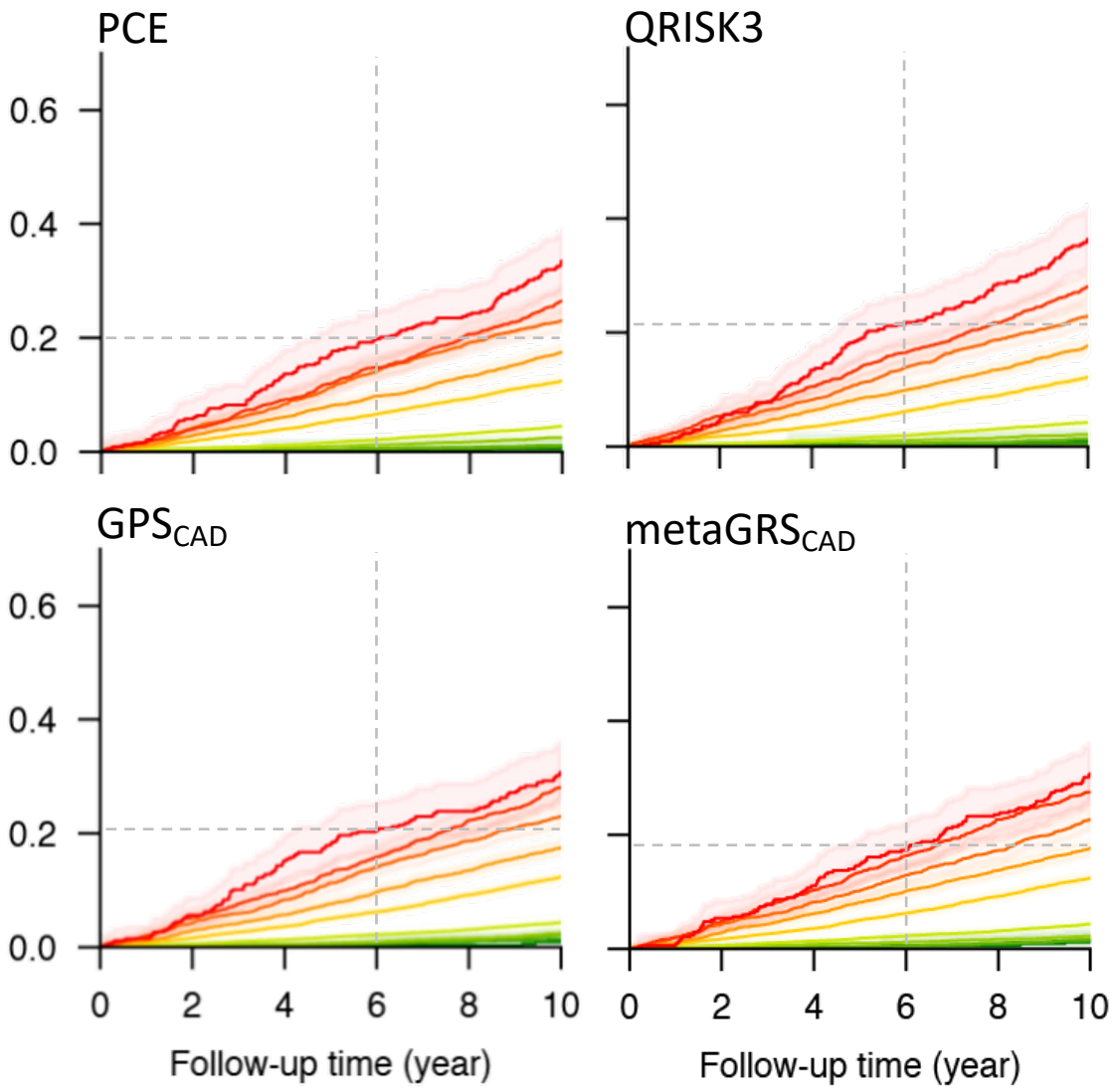


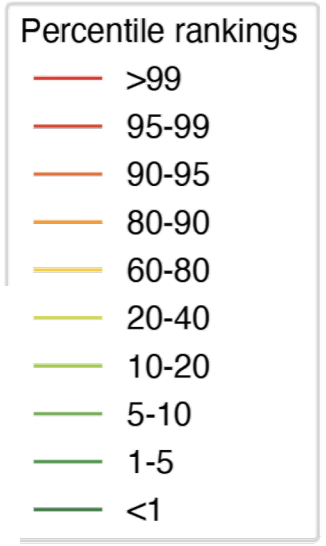
Superior Risk Stratification



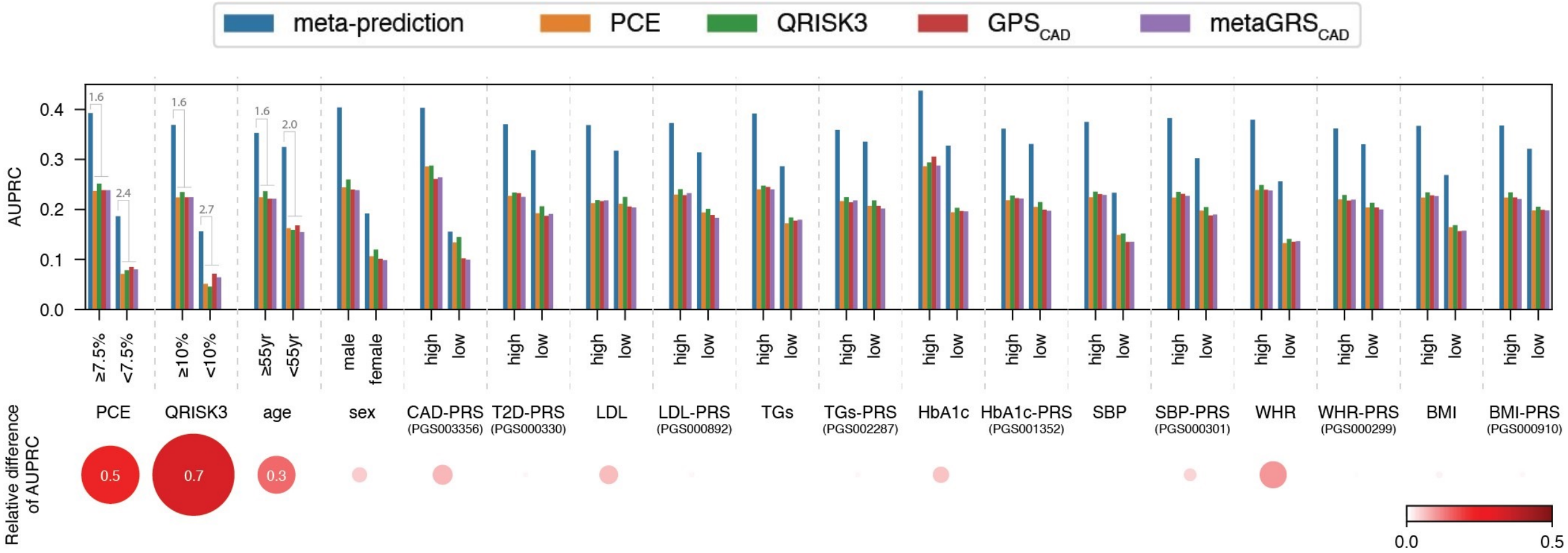
Superior Stratification of Risk Trajectories



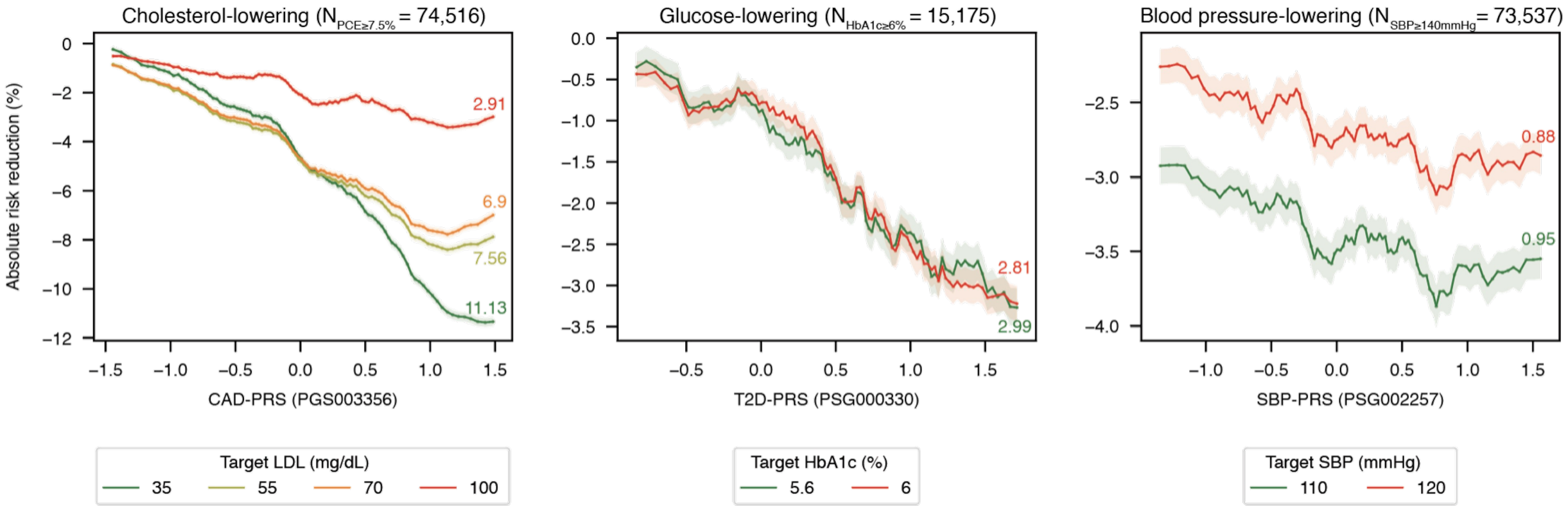




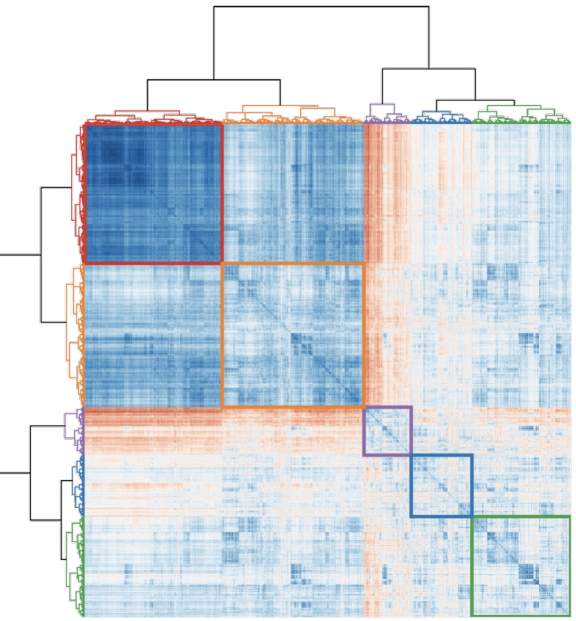
Our Model Captures Hidden Risk in “Low Risk” Individuals

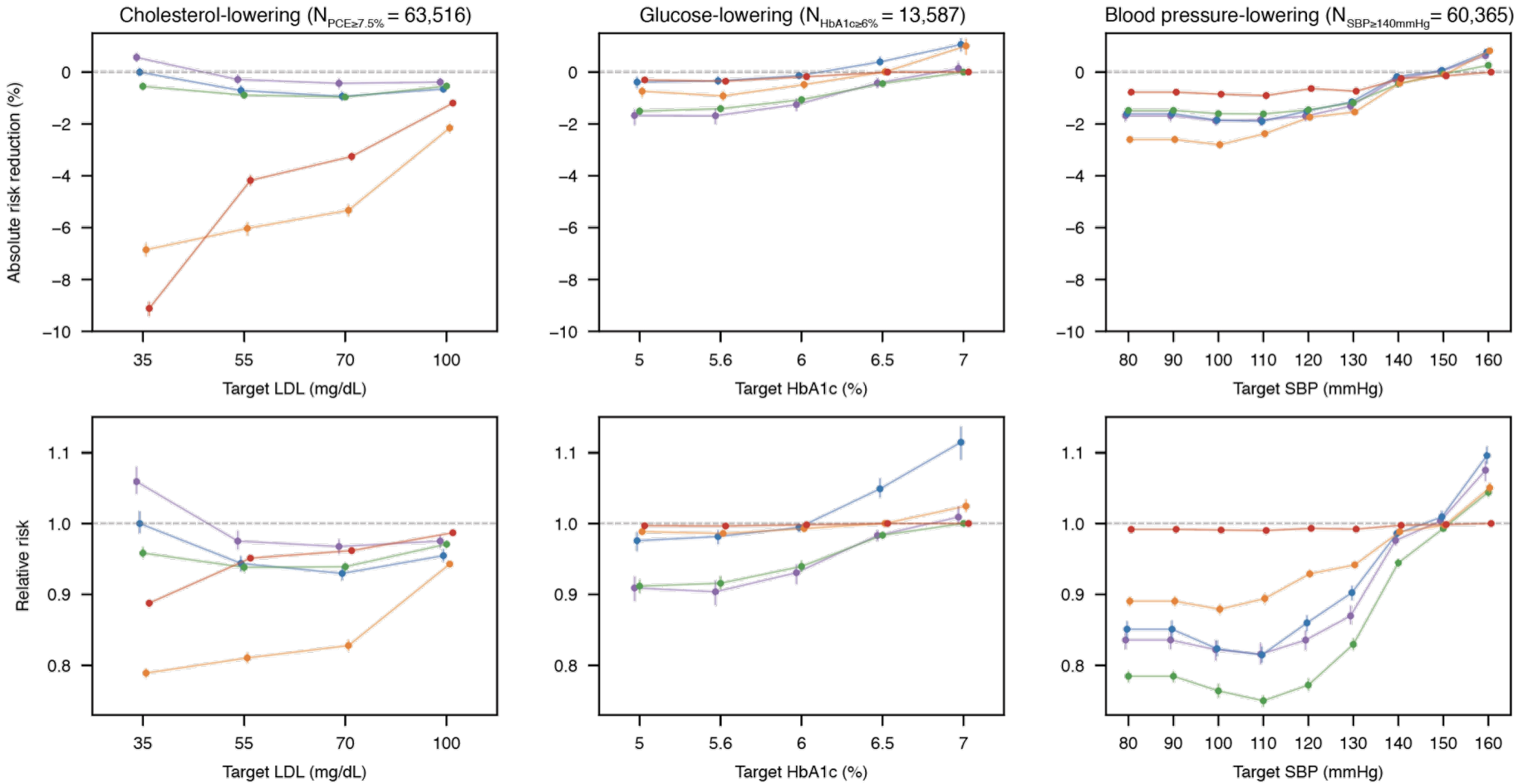


When Predicting Your Risk, Genetic Risk Mediates Benefit of Clinical Interventions



Identification of Sub-Groups with Unique Benefit Profiles from Prevention





RECAP

- **Coronary Artery Disease is deadly and we don’t care**
- **Genetic Risk can find the right people who should care**
- **Genetic Risk can convince the right people to care**
- **We can do even better with Machine Learning**

Acknowledgements

Scripps Team

Kai-Yu Chen	Graduate Student
Shang-Fu Chen	**Graduate Student**
Raquel Dias PhD	Senior Scientist
Doug Evans	Bioinformatic Analyst
Brianna Fernandez	Programmer
Hossein Javedani Sadaei	Data Scientist
Ahmed Khattab	Graduate Student
Will Liu	Programmer
Salvatorre Logueric io	Sr. Scientist
Sang Eun Lee MD	Visiting Physician Scholar
Evan Muse MD PhD	Physician Scientist
Jun-Bean Park MD	Visiting Physician Scholar
Elias Salfati PhD	Senior Scientist
Emily Spencer PhD	Associate Director
Juan Antonio Raygoza Garay	Bioinformatic Analyst
Eric Topol MD	Director
Sarah Topol	Study Coordinator
Nathan Wineinger PhD	Associate Professor

Funding

Center For Translational Sciences

Stowers Family Foundation

Volo Foundation

Kessler Family

Interns

Hector Aguilar
Arjun Ghoshal
André Nicolás León
Raymond Lee
Nolan Mejia
Ernesto Prado Montes de Oca
Neha Pubbi
Paris Zhang
…many others

Contact:
atorkama@scripps.edu
@ATorkamani