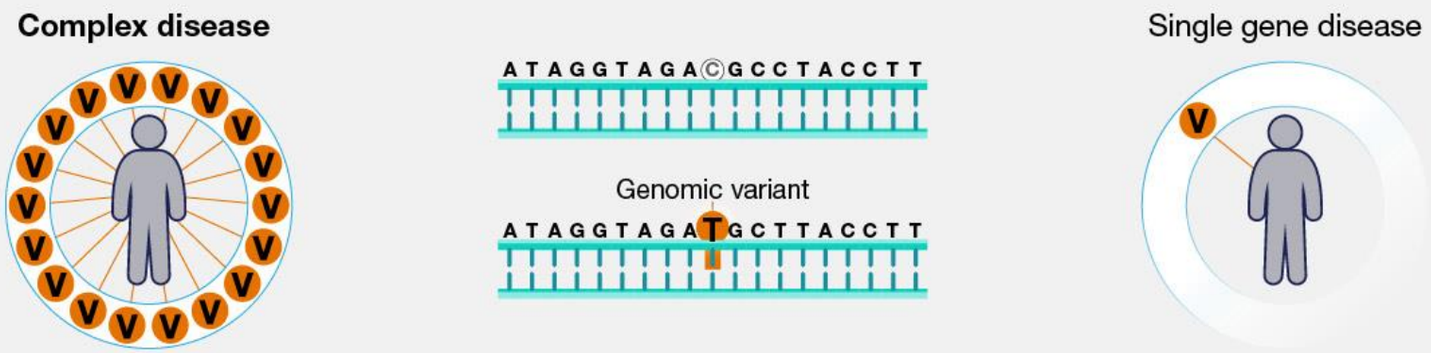
Polygenic risk scores for complex diseases

Evangelos Evangelou, PhD
Assc. Prof., University of Ioannina Medical School
Affiliated Scientist, BRI-FORTH

vangelis@uoi.gr

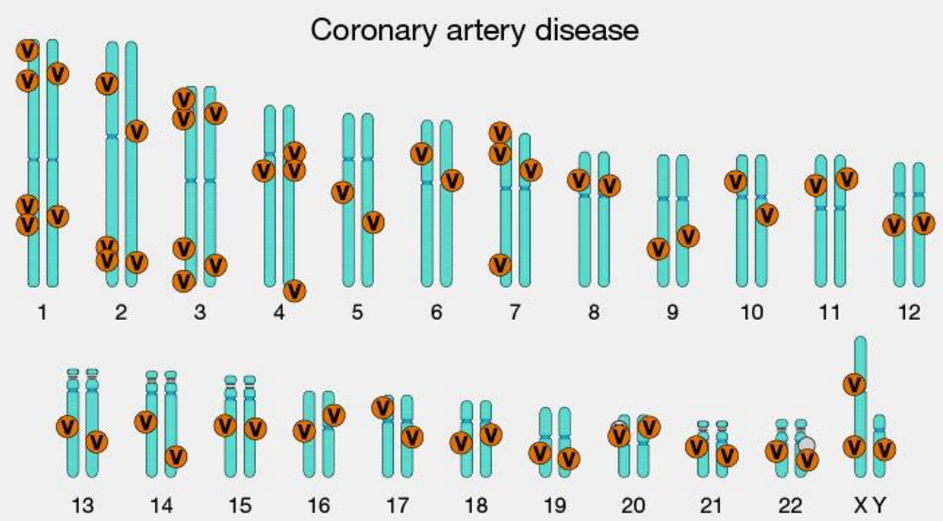
13th FORTH Retreat, July 2022, Crete, GR

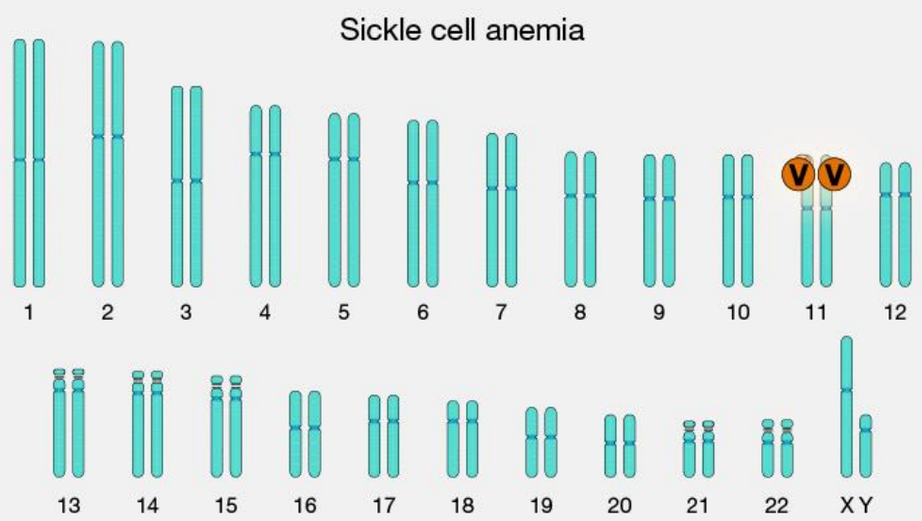
Complex vs single gene diseases



Many genomic variants (polygenic) interacting with environmental factors

One genomic variant involved





Quality Control

MACH-IMPUTE

90M 1KG Project-300M TopMed

Association testing

Published Genome-Wide Associations as of July 2019

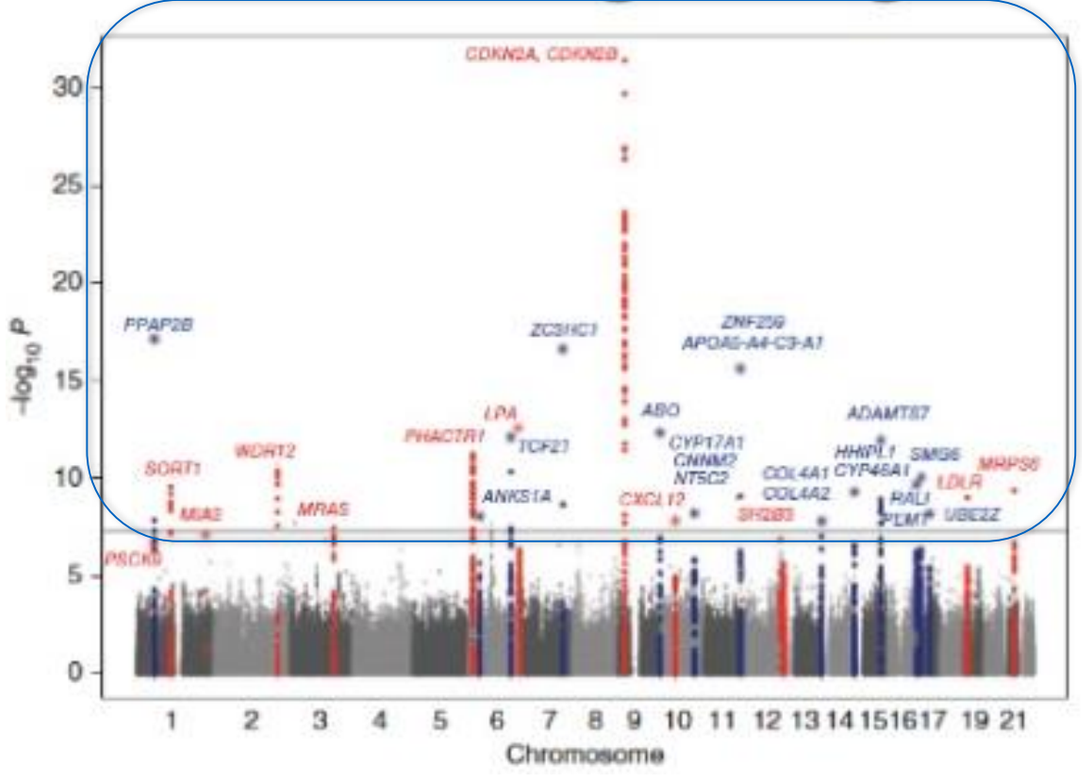
p≤5X10-8 for 17 trait categories

National Human Genome Research Institute

EMBL-EBI

NHGRI-EBI GWAS Catalog
www.ebi.ac.uk/gwas

Polygenic Risk Score



- Large GWAS studies
  → robust genetic variants associated with complex traits/diseases

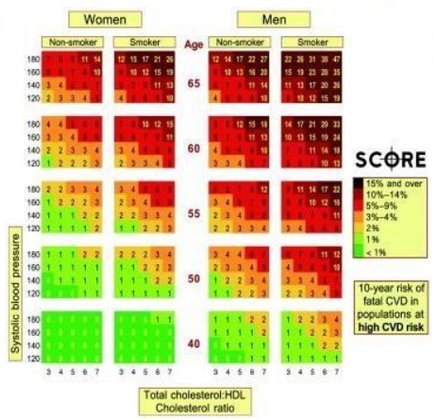
- Cumulative effect of genetic variants in relation to trait/disease→ polygenic risk score

$$PRS_j = \sum_i^N \beta_i * dosage_{ij}$$

Risk prediction

- Risk prediction is widely used for clinical practice e.g. in cardiology

- Various risk scores have so far been developed



New England Journal of Medicine 2009

Genetic Risk Prediction — Are We There Yet?

Peter Kraft, Ph.D., and David J. Hunter, M.B., B.S., Sc.D., M.P.H.

A major goal of the Human Genome Project was to facilitate the identification of inherited genetic variants that increase or decrease the risk of comp diseases. The completion of t International HapMap Project a the development of new metho for genotyping individual D samples at 500,000 or more l have led to a wave of discover through genomewide associati studies. These analyses have id tified common genetic varia that are associated with the r of more than 40 diseases and l man phenotypes. Several com nies have begun offering dire to-consumer testing that uses

tests of genetic predisposition to important diseases would have major clinical, social, and economic ramifications. But the great ma-

est relative risks are almost certainly overrepresented in the first wave of findings from genome-wide association studies, since

polymorphisms [SNPs]) and novel a virtual explosion of novel genet complex human diseases. In th advances have been remarkably many novel genetic associations (MI) and cardiovascular risk fac pressure, diabetes, and obesity. A studies has always been to prov biology of CVD. However, a high these discoveries has been to use usher in a new era of personalized genetic information into risk pre

Genetic Cardiovascular Risk Prediction
Will We Get There?

George Thanassoulis, MD; Ramachandran S. Vasan, MD

Circulation 2010

Major advances in genetics, including the sequencing of the human genome in 2001[1,2] and the publication of the HapMap in 2005,[3] have paved the way for a revolution in our understanding of the genetics of complex diseases, including cardiovascular disease (CVD). A results and failure to replicate put ciations, high-throughput technolo than 500 000 genetic markers k

these factors, a number of risk prediction algorithm scores have been developed, including the Framingham risk score, that provide an estimate of the 10-year risk (and recently, the 30-year risk) of CVD.[6–9] Generally speaking, the metrics

Clinical Utility of Genetic Variants for Cardiovascular Risk Prediction
A Futile Exercise or Insufficient Data?

Emanuele Di Angelantonio, MD, MSc, PhD; Adam S. Butterworth, MSc, PhD

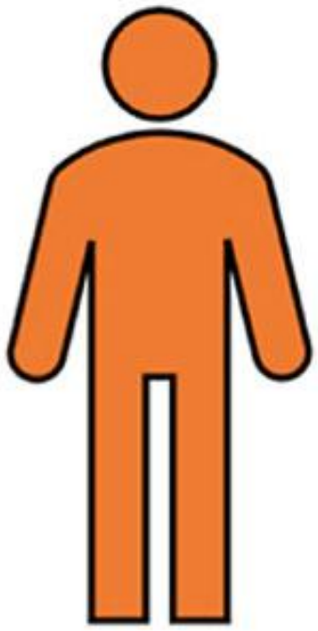
Circ Cardiovasc Genet. 2012

Estimation of an individual's cardiovascular disease (CVD) risk usually involves measurement of risk factors correlated with risk of CVD to identify people who may especially benefit from preventive action, such as lifestyle advice or pharmacologic agents.[1] Since the Framingham Risk Score was first developed, several other risk-prediction algorithms have been proposed, each involving a core set of the same established risk factors (ie, age, sex, smoking, blood pressure, and total cholesterol), but differing in their inclusion of various other characteristics (eg, ethnicity or presence of diabetes mellitus).[2] The challenge in recent years has been to improve existing CVD risk-prediction models by including additional information to the traditional risk factors generally included in risk scores. Several additional soluble biochemical factors have been advocated for inclusion, but contradictory evidenc been reported on the incremental predictive gain afford these markers, and there is divergence of expert opinion

Until a few years ago, genetic epidemiologic studies of CVD were predominantly candidate gene studies involving focused investigation of relatively few genetic variants based on plausible biological hypotheses. Many of these studies had anticipated identification of variants that are common in populations with moderate-to-large effects on disease risk. However, the combination of the low prior odds of the variants selected for study, inadequate power (ie, small sample size), and overliberal declarations of significance, resulted in the reporting of many seemingly positive findings that remain unreplicated or directly refuted.[7] In recent years, genome-wide association studies (GWAS) have demonstrated that so-called hypothesis-free global-testing methods can advance discovery and understanding of genetic variants in relation to chronic

our knowledge of the genetic architecture of vascular disease,

Familial combined hyperlipidemia



~ 1 in 50
people

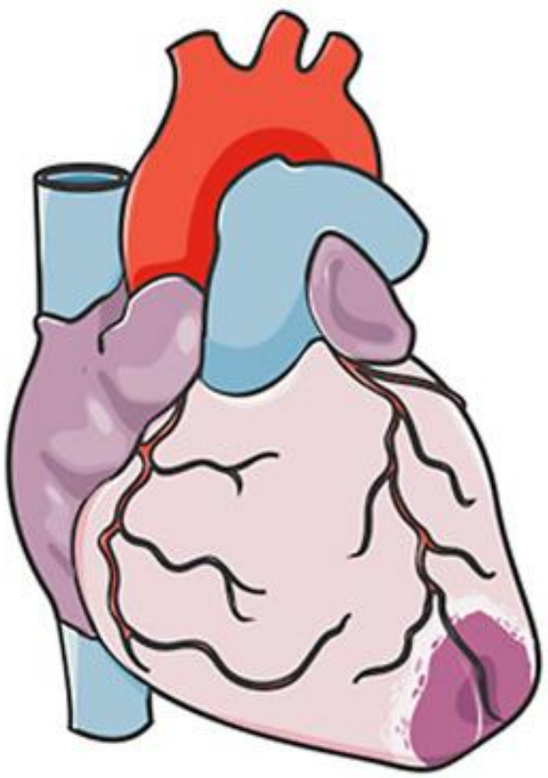
Polygenic
Co-morbidities
Environment

Familial hypercholesterolemia

~ 1 in 300
people

Monogenic

Similar high risk
of coronary artery
disease



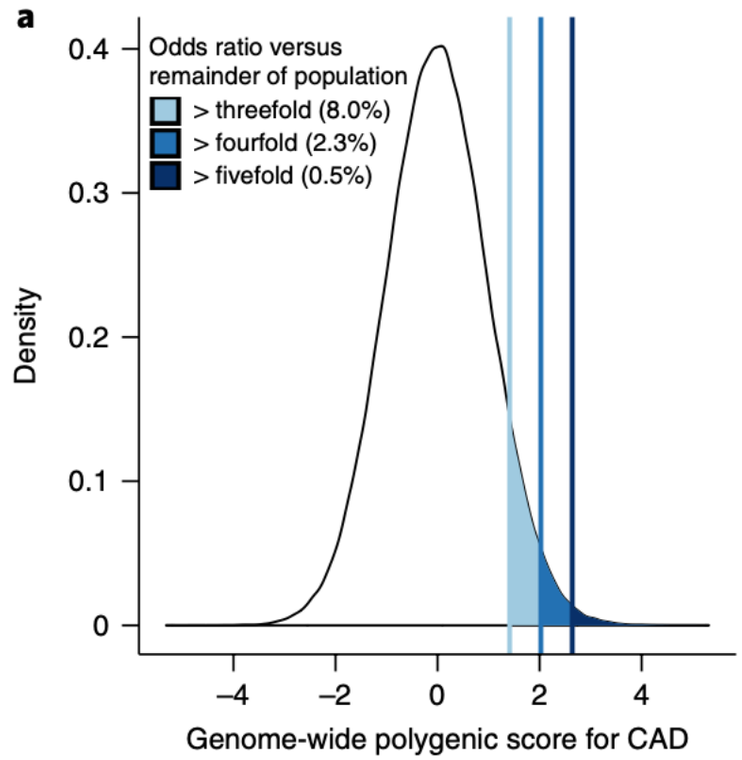
nature genetics

LETTERS

https://doi.org/10.1038/s41588-018-0183-z

Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations

Amit V. Khera[1,2,3,4,5], Mark Chaffin[4,5], Krishna G. Aragam[1,2,3,4], Mary E. Haas[4], Carolina Roselli[4], Seung Hoan Choi[4], Pradeep Natarajan[2,3,4], Eric S. Lander[4], Steven A. Lubitz[2,3,4], Patrick T. Ellinor[2,3,4] and Sekar Kathiresan[1,2,3,4]*



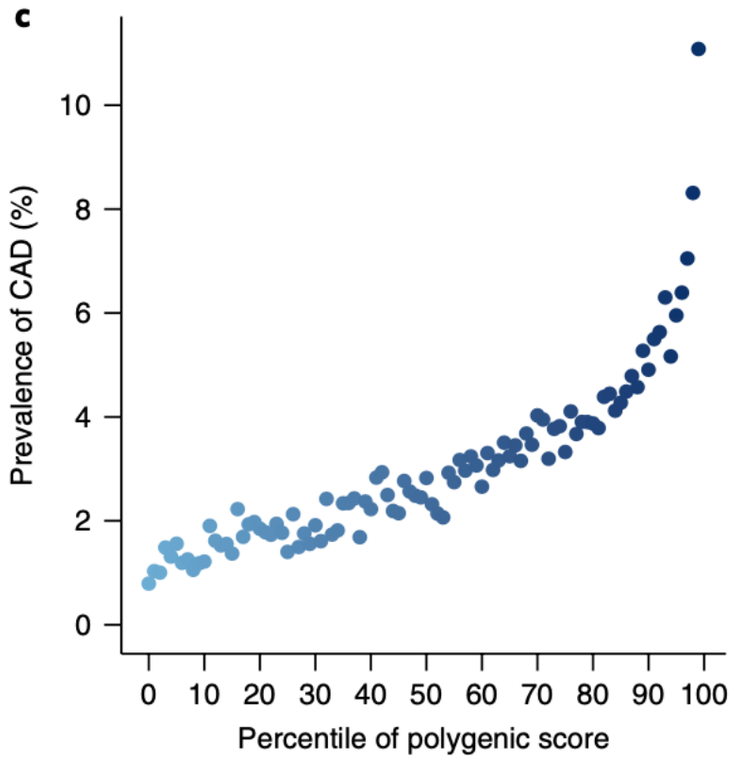


Table 2 | Proportion of the population at three-, four- and fivefold increased risk for each of the five common diseases

High GPS definition	Individuals in testing dataset (*n*)	% of individuals
Odds ratio ≥3.0		
CAD	23,119/288,978	8.0
Atrial fibrillation	17,627/288,978	6.1
Type 2 diabetes	10,099 288,978	3.5
Inflammatory bowel disease	9,209 288,978	3.2
Breast cancer	2,369/157,895	1.5
Any of the five diseases	57,115/288,978	19.8

What has changed?

- 2005-2007: Initial GWAS for complex diseases
- 2008-2010: Proof of concept for polygenic risk scores but limited utility
- 2010-2018: Common variants explain majority of disease heritability
- 2018: Ability to identify clinically meaningful increases in risk

- Large GWAS→ increases precision for effect estimates

- Algorithms to combine large sets of variants

- Large biobanks for validation and testing

Huge data resources

FINNGEN

biobank uk

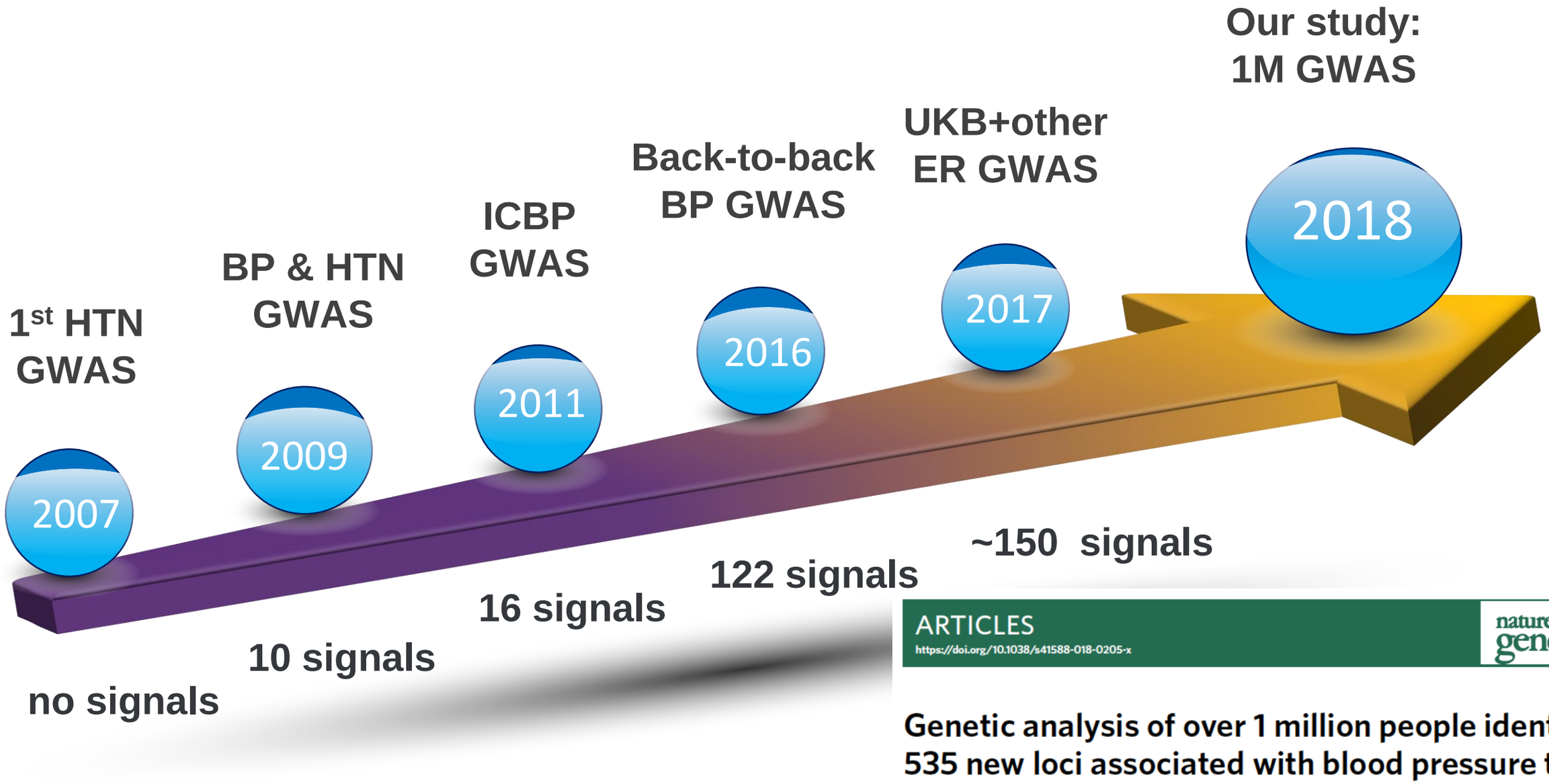
23andMe

MVP

Tailor-made Medical Treatment Program
(BioBank Japan: BBJ)

CHINA KADOORIE BIOBANK
中国慢性病前瞻性研究

Unravelling the genetic architecture of BP



ARTICLES
https://doi.org/10.1038/s41588-018-0205-x
nature genetics

Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits

High blood pressure is a highly heritable and modifiable risk factor for cardiovascular disease. We report the largest genetic association study of blood pressure traits (systolic, diastolic and pulse pressure) to date in over 1 million people of European ancestry. We identify 535 novel blood pressure loci that not only offer new biological insights into blood pressure regulation but also highlight shared genetic architecture between blood pressure and lifestyle exposures. Our findings identify new biological pathways for blood pressure regulation with potential for improved cardiovascular disease prevention in the future.

Evangelou E et al. Nat Genet; 50(10):1412-1425

Analysis

Systolic BP **Diastolic BP** **Pulse Pressure**

Adjust for medication use: +15 mmHg and 10mmHg to mean SBP and DBP
PP=SBP-DBP

Two-stage analysis
Follow-up SNPs with $P < 1 \times 10^{-6}$ for any BP trait
(with concordant direction of effect for UKB vs ICBP)

Independent Replication meta-analysis
→ Lookups of sentinel SNPs
in MVP (N=220,520) and EGCUT (N=28,742)
→ combined meta-analysis (**N=1,006,863**)

(i) genome-wide significant ($P < 5 \times 10^{-8}$) in combined meta
(ii) $P < 0.01$ in replication meta-analysis
(iii) concordant direction of effect discovery vs replication

One-stage analysis
Follow-up SNPs with $P < 5 \times 10^{-9}$ for any BP trait
(with concordant direction of effect for UKB vs ICBP)

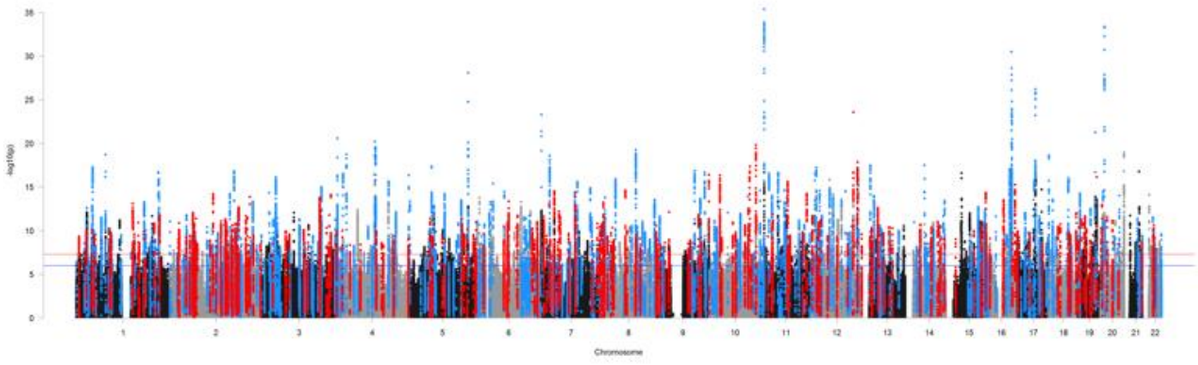
→ **UKB-ICBP Internal Replication**
of sentinel SNPs

(i) $P < 0.01$ in UKB GWAS
(ii) $P < 0.01$ in ICBP GWAS meta-analysis

Secondary analyses of identified signals

Summary results

- **535** novel loci identified



- **92** loci replicated for first time within 2-stage design

- Support for all **274** previously published loci

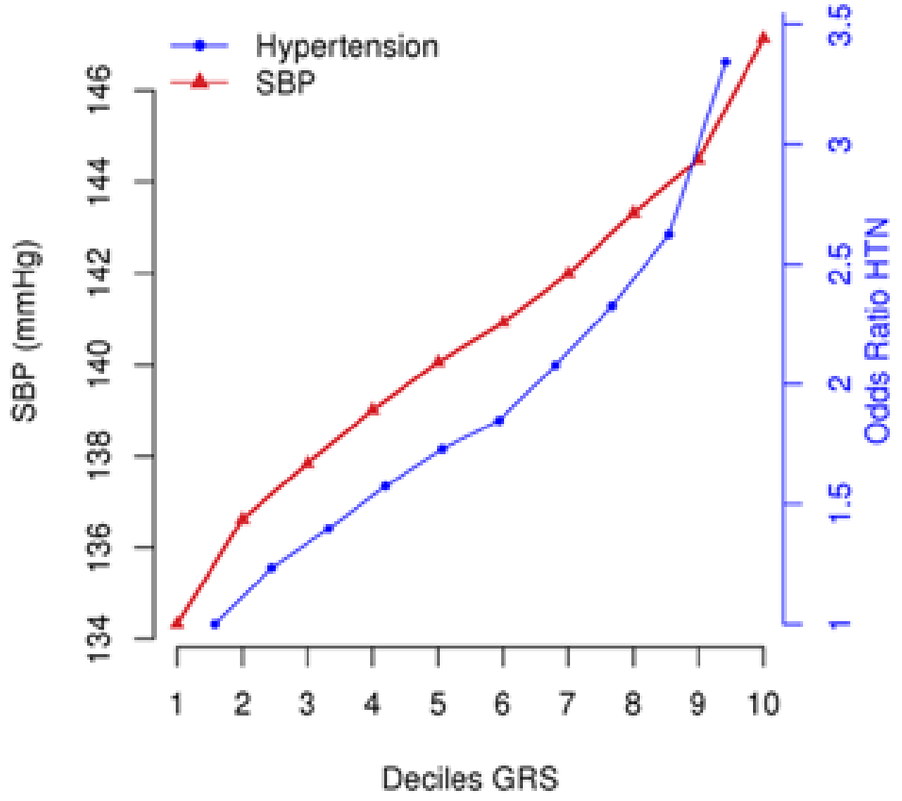
- **163** independent secondary signals from conditional analysis

Now, over 1,000 independent BP signals

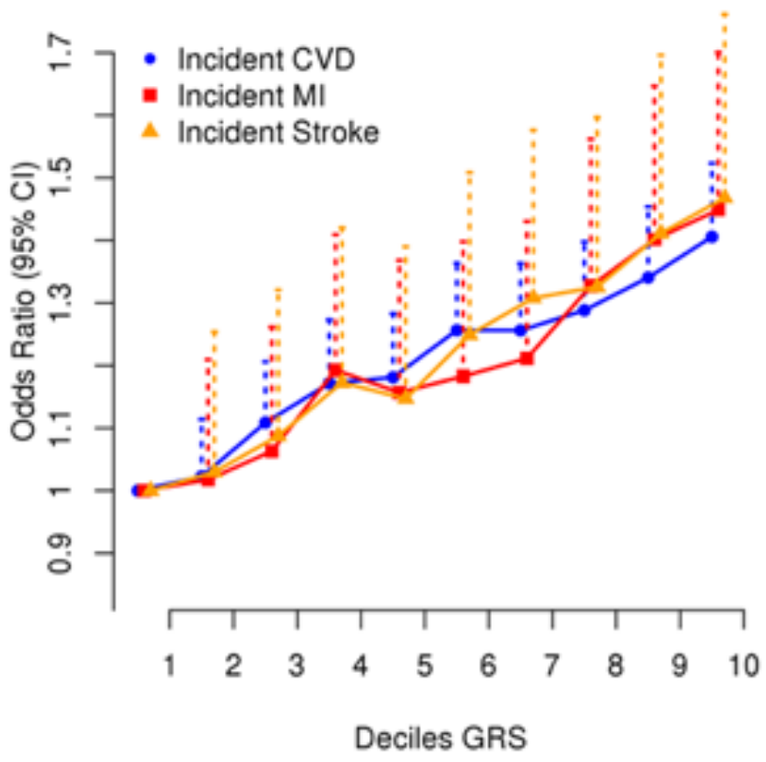
Polygenic Risk Scores

We calculated wPRS combining all novel and published BP loci in UKB (N~390K) and an independent cohort (Airwave: N~14K)

Risks for SBP, HTN and CVD were calculated comparing top vs bottom 10% of GRS



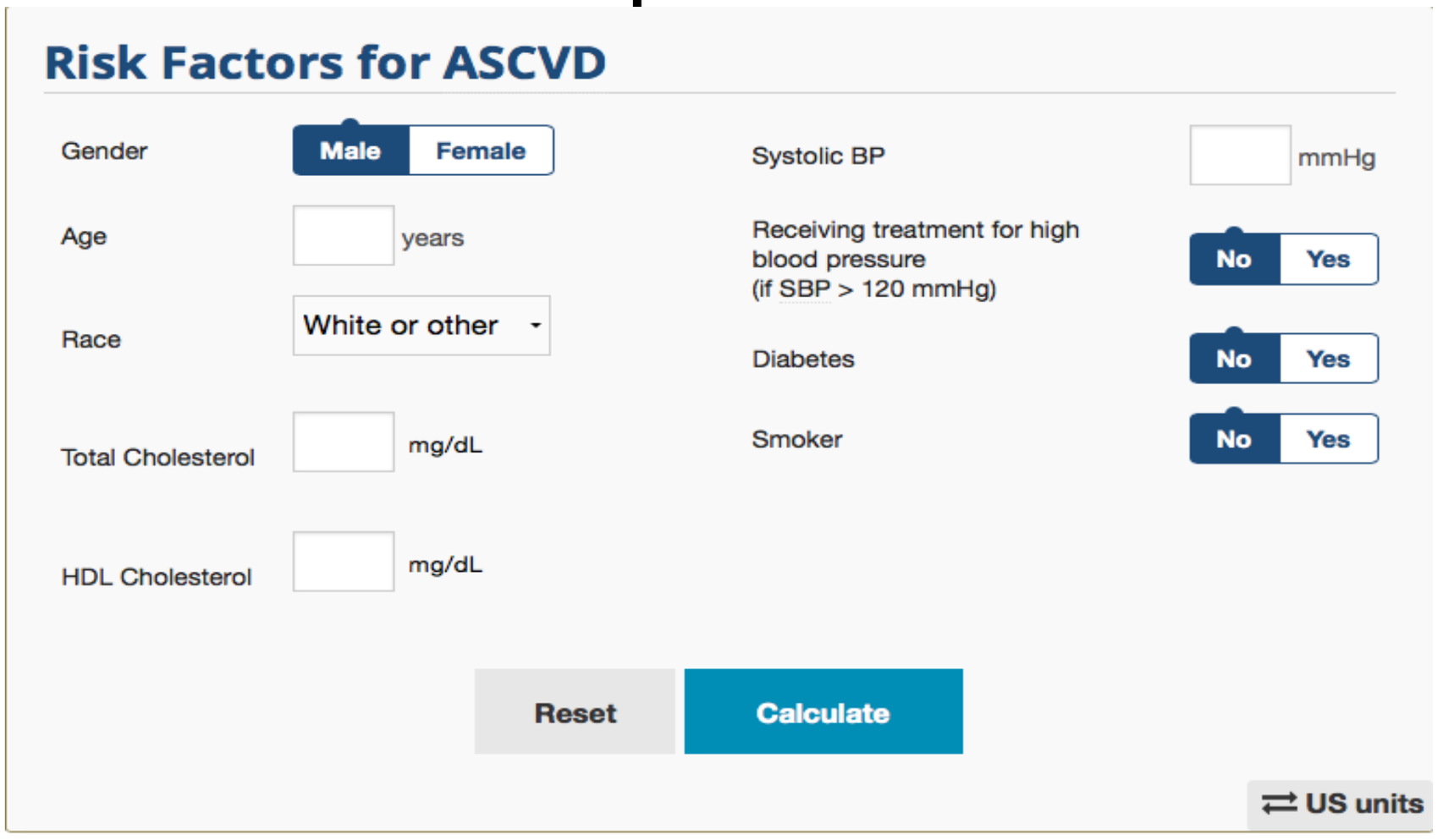
Increase of 12.85 mmHg ($P<1x10^{-300}$)
OR=3.34 for HTN ($P<1x10^{-300}$)
In unrelated UKB samples



Increased risk (OR=1.52; $P=7.7x10^{-6}$) of all cardiovascular **(CVD) outcomes** in UKB HES data

Risk prediction for CVD in general populations

Prevention strategies currently based on well established risk prediction models



Genetic risk prediction in CVD

- We now know hundreds of genetic variants that influence disease risk
- Can they be useful in disease risk prediction?

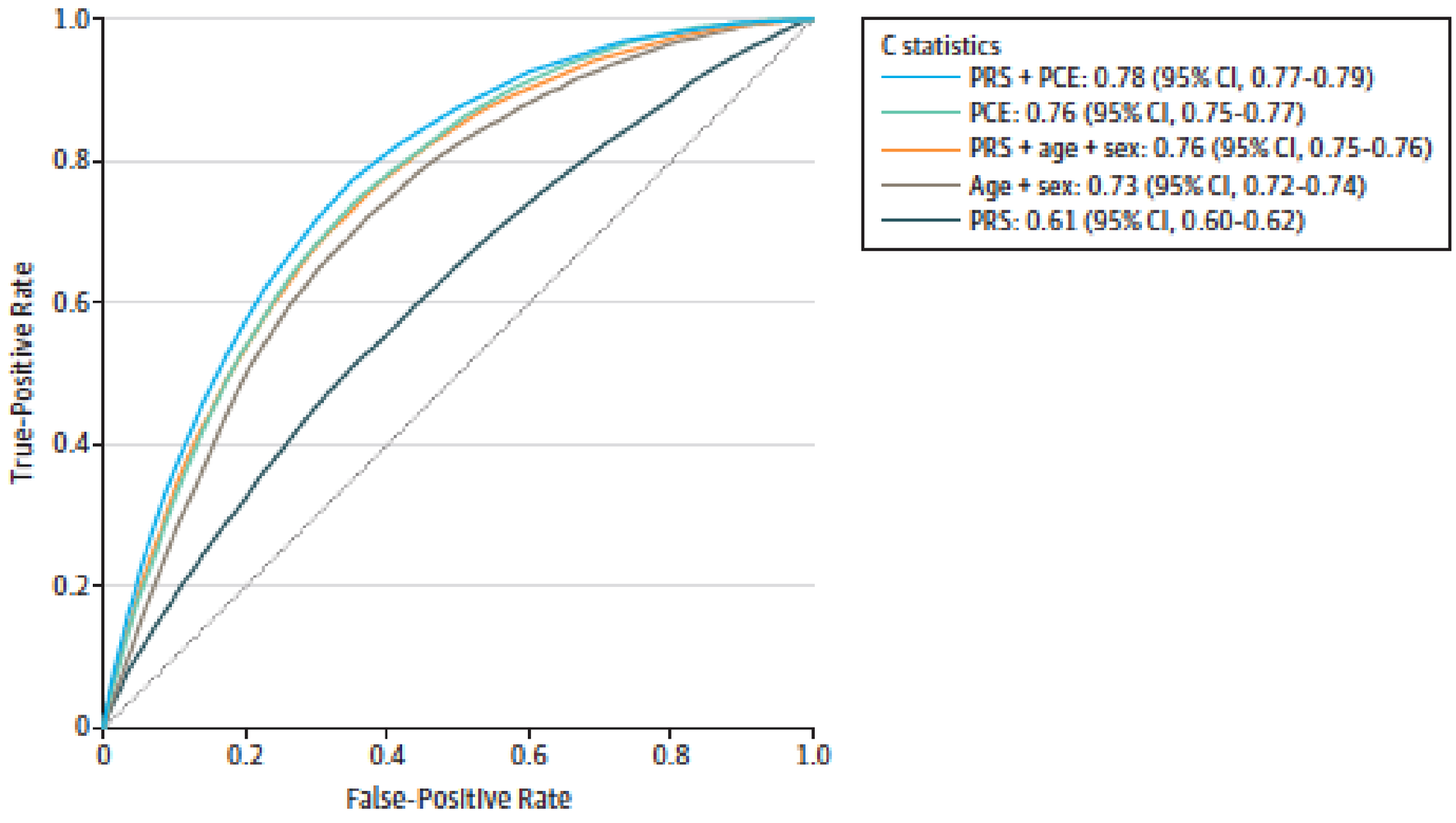
Predictive Accuracy of a Polygenic Risk Score-Enhanced Prediction Model vs a Clinical Risk Score for Coronary Artery Disease

Joshua Elliott, MBBS, MSc; Barbara Bodinier, MSc; Tom A. Bond, MBBS; Marc Chadeau-Hyam, PhD; Evangelos Evangelou, PhD; Karel G. M. Moons, PhD; Abbas Dehghan, MD, PhD; David C. Muller, PhD; Paul Elliott, MBBS, PhD; Ioanna Tzoulaki, PhD

JAMA | **Original Investigation**

PRS and pooled cohorts equation

Figure 3. Receiver Operator Characteristic Curves and C Statistics for Different Models in Cohort Analyses of 352 660 Participants Aged 40 to 69 Years Old Over a Mean of 8 Years of Follow-up With 6272 Incident Coronary Artery Disease (CAD) Events



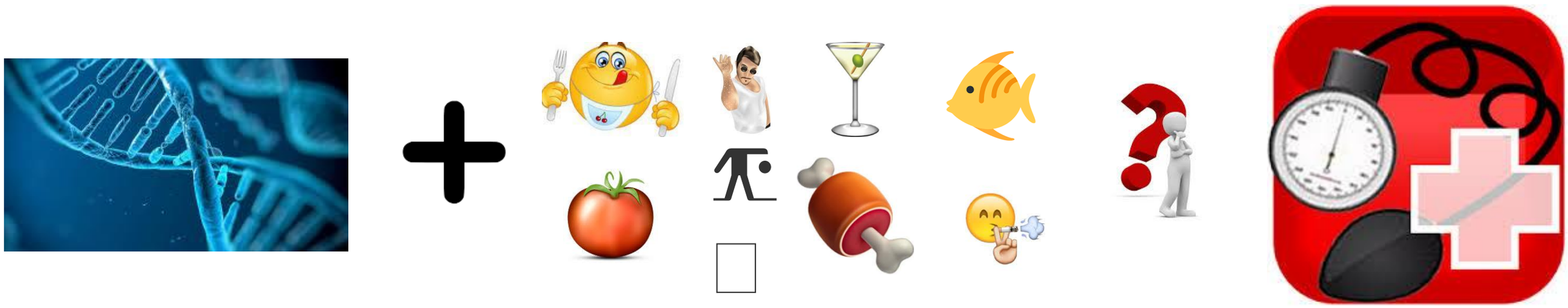

Your genes
Your environment & lifestyle
YOU!

Circulation

ORIGINAL RESEARCH ARTICLE

Genetic Predisposition to High Blood Pressure and Lifestyle Factors

Associations With Midlife Blood Pressure Levels and Cardiovascular Events



Aim: To investigate the extent to which lifestyle factors could offset the effect of an adverse BP genetic profile and its effect on CVD risk

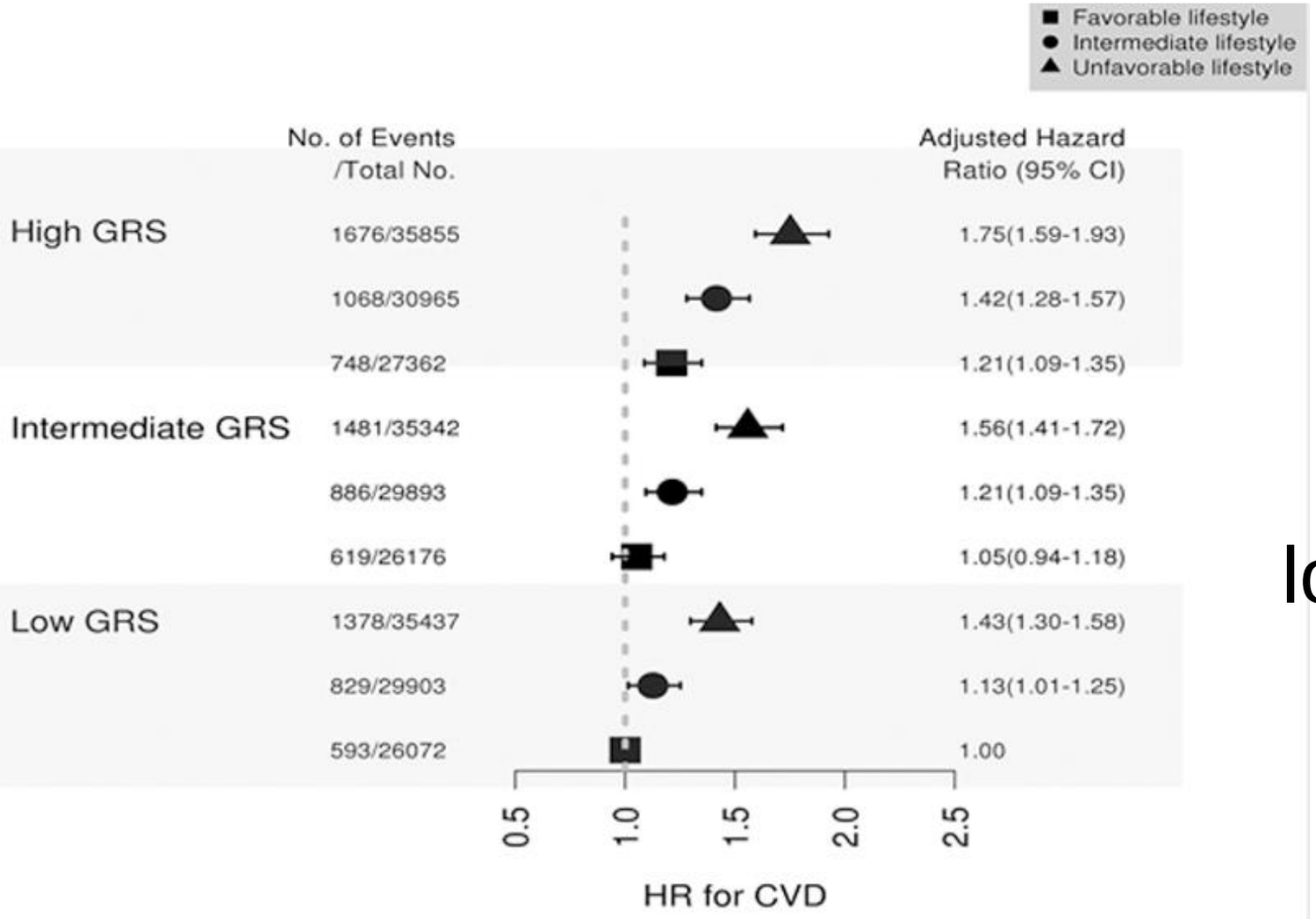
Favorable lifestyle
Intermediate lifestyle
Unfavorable lifestyle

Predicted SBP (mmHg)

110
120
130
140
150
160
170

Low Genetic Risk
Intermediate Genetic Risk
High Genetic Risk

Lifestyle and Genetic Risk Categories



30%, 31% AND 33% modification benefit between favorable and unfavorable lifestyle at low, intermediate and high genetic risk

Raha Pazoki. Circulation. Genetic Predisposition to High Blood Pressure and Lifestyle Factors, Volume: 137, Issue: 7, Pages: 653-661, DOI: (10.1161/CIRCULATIONAHA.117.030898)

© 2017 American Heart Association, Inc.

The Polygenic Score (PGS) Catalog

An open database of polygenic scores and the relevant metadata required for accurate application and evaluation.

Search the PGS Catalog

Examples: breast cancer, glaucoma, BMI, EFO_0001645

Explore the Data

In the current PGS Catalog you can **browse** the scores and metadata through the following categories:

Polygenic Scores

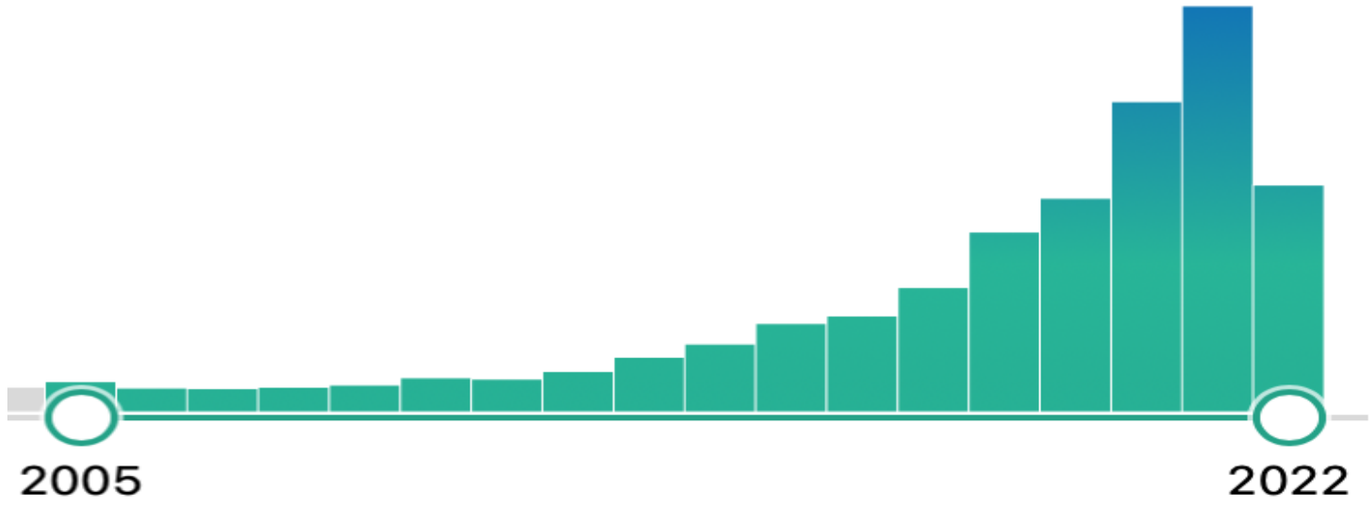
2,628

Traits

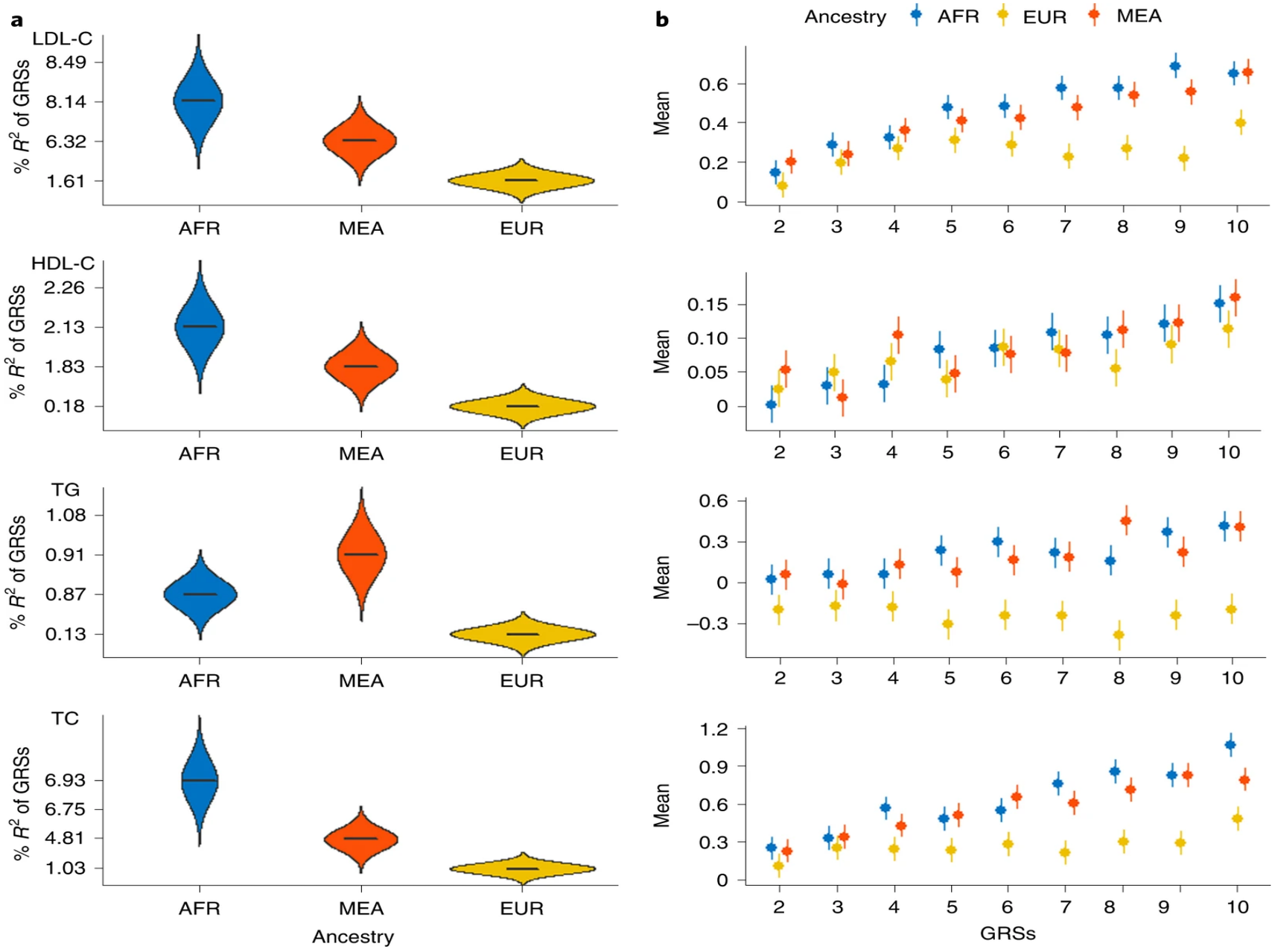
549

Publications

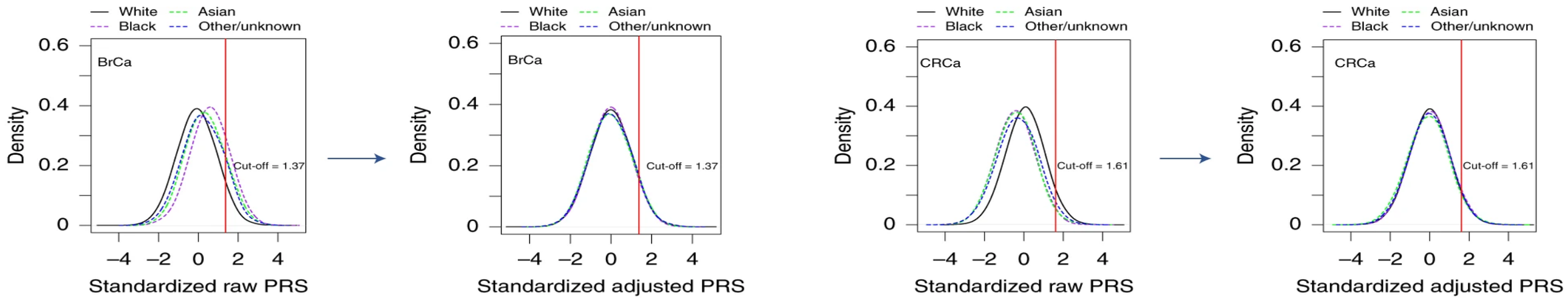
323



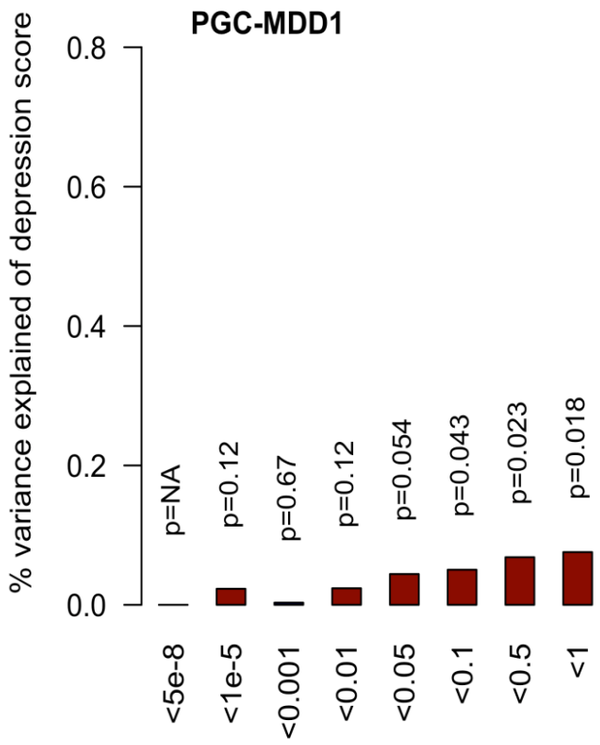
Transferability of PRS in other ancestries



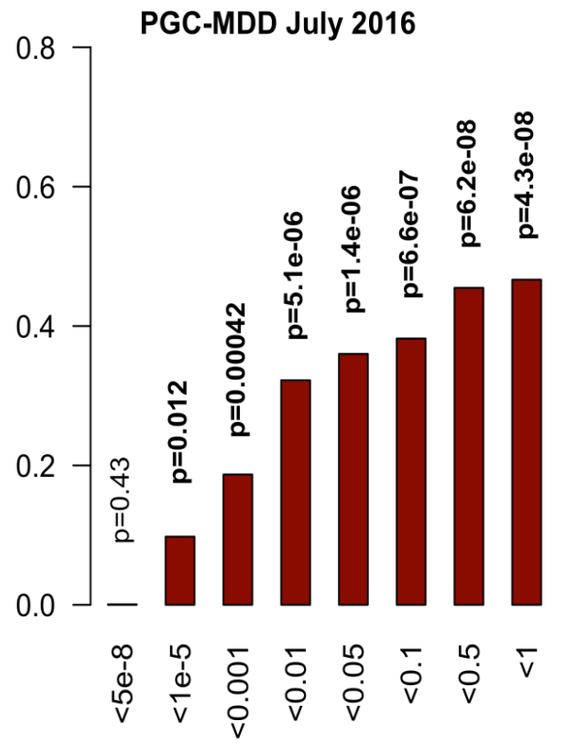
PRS by reported race before and after adjustment for population structure



Power of PRS analysis increases with GWAS sample size



PGC-MDD1: N=18k
max variance explained = 0.08%,
p=0.018

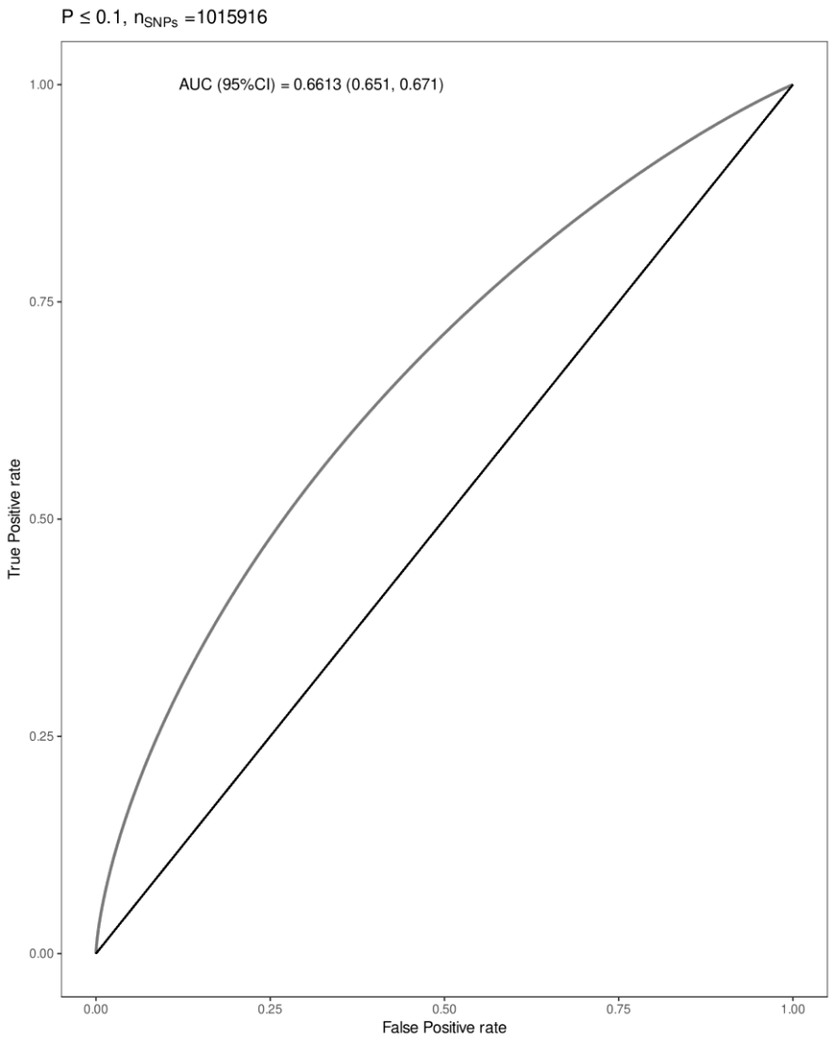


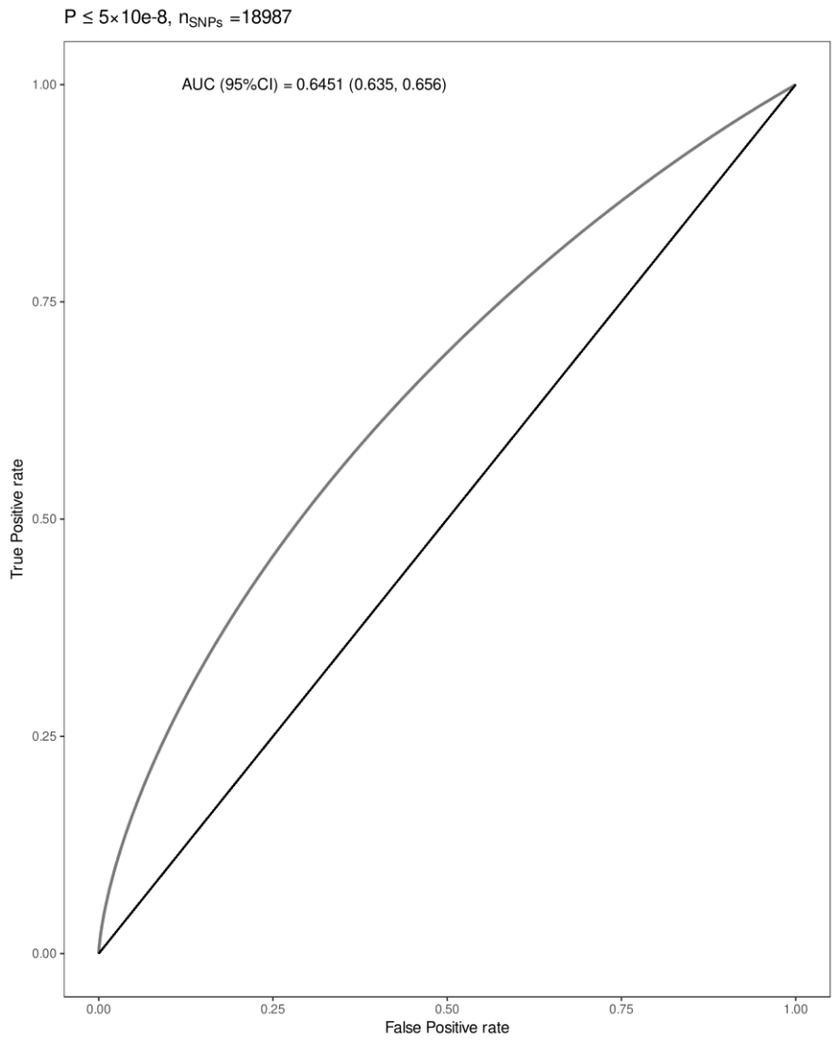
PGC-MDD2: N=163k
max variance explained =0.46%,
p= 5.01e-08

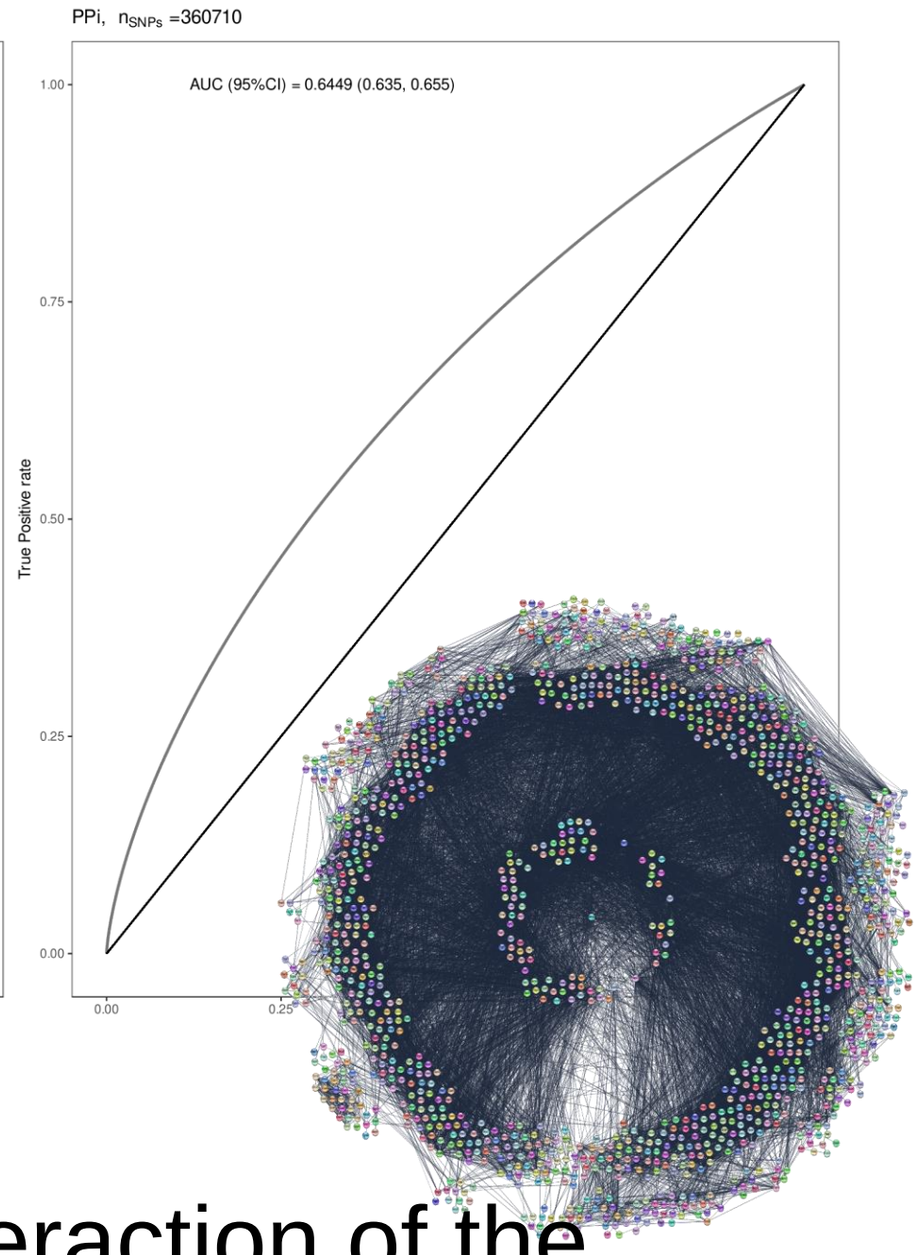
Colodro-Conde L,
Couvy-Duchesne B, et al, (2017)
Molecular Psychiatry

Preliminary work

- Application of European-based PRS to other populations
- Methods for PRS derived from multivariate analysis
- PRS for various traits and diseases
  - We proposed first GRS for melanoma
- PRS and lifestyle factors
- Drug-related PRS/Network-related PRS



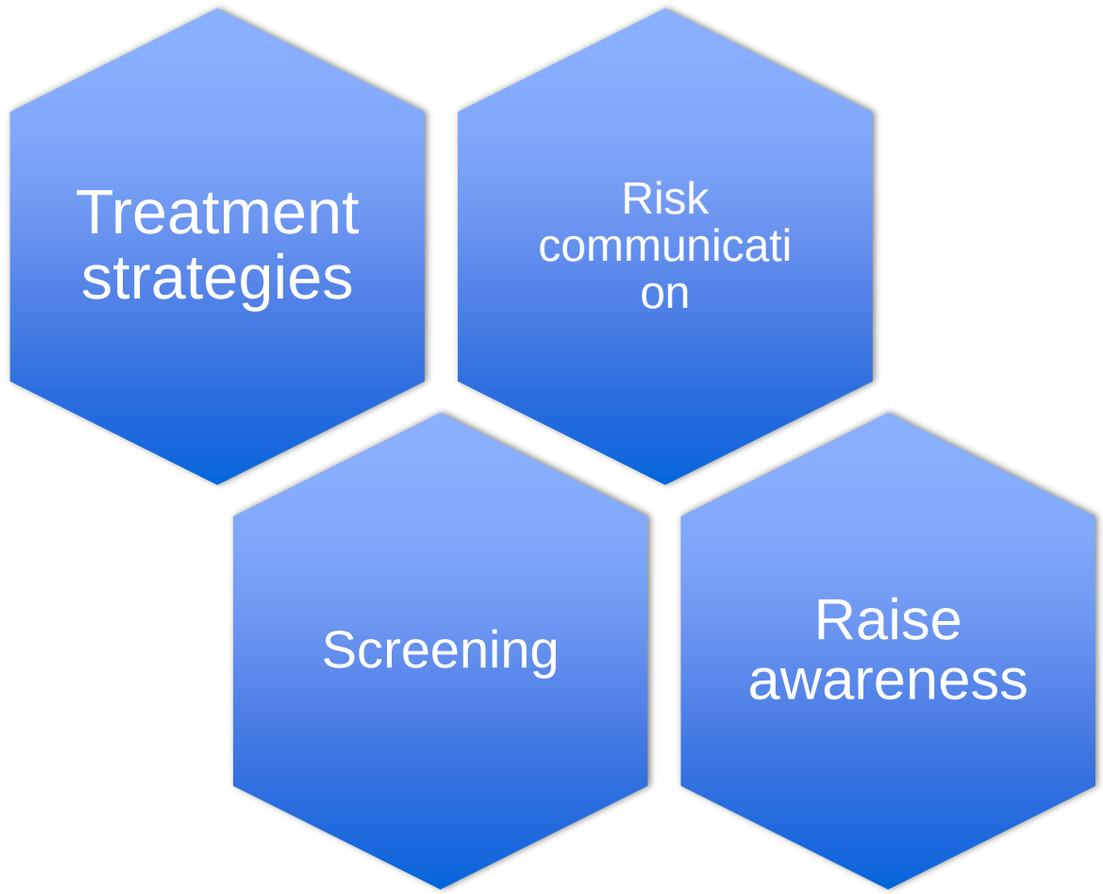




Protein-Protein Interaction of the 1575 PSORS-interacting genes

Polygenic risk scores in disease risk prediction

- ✓ Genetic variants are set at birth → attractive predictors of disease
- ✓ No measurement error, early information, inexpensive and easily obtained
- ❖ Association with disease often modest, limited predictive power
- ❖ Only useful when there are preventive strategies/ screening/treatment
- ❖ Non-modifiable



MIT Technology Review

VOL. 121 NO. 2 MARCH/APRIL 2018 US $6.99/CAN $7.99

3-D metal printing

Babel-fish earbuds

The sensing city

AI for everyone

Dueling neural networks

Materials' quantum leap

Zero-carbon natural gas

Perfect online privacy

Artificial embryos

and

Genetic fortune-telling

10 Breakthrough Technologies 2018

DISPLAY UNTIL 05/1/2018

Forecasts of genetic fate just got a lot more accurate

DNA-based scores are getting better at predicting intelligence, risks for common diseases, and more.

BY ANTONIO REGALADO

A. Musas, C. Antonatos, F. Koskeridis, G. Ntritsos, V. Bellou, I. Tzoulaki, K. Tsilidis, E. Ntzani