Analysis And Application Of Cohort Data

Cardiovascular Disease Precision Medicine Driven By Population Genomics

2025/10/14
International Training Course
on Biological Data Integration,
Sharing and Application

汪敏先 PhD.
研究员、博士生导师
应用发展部部长
国家生物信息中心

中国科学院北京基因组研究所（国家生物信息中心）
BEIJING INSTITUTE OF GENOMICS CHINESE ACADEMY OF SCIENCES / CHINA NATIONAL CENTER FOR BIOINFORMATION

INNOVATION TOGETHER
Wang lab
A data-driven
precision medicine lab

Personal Profile

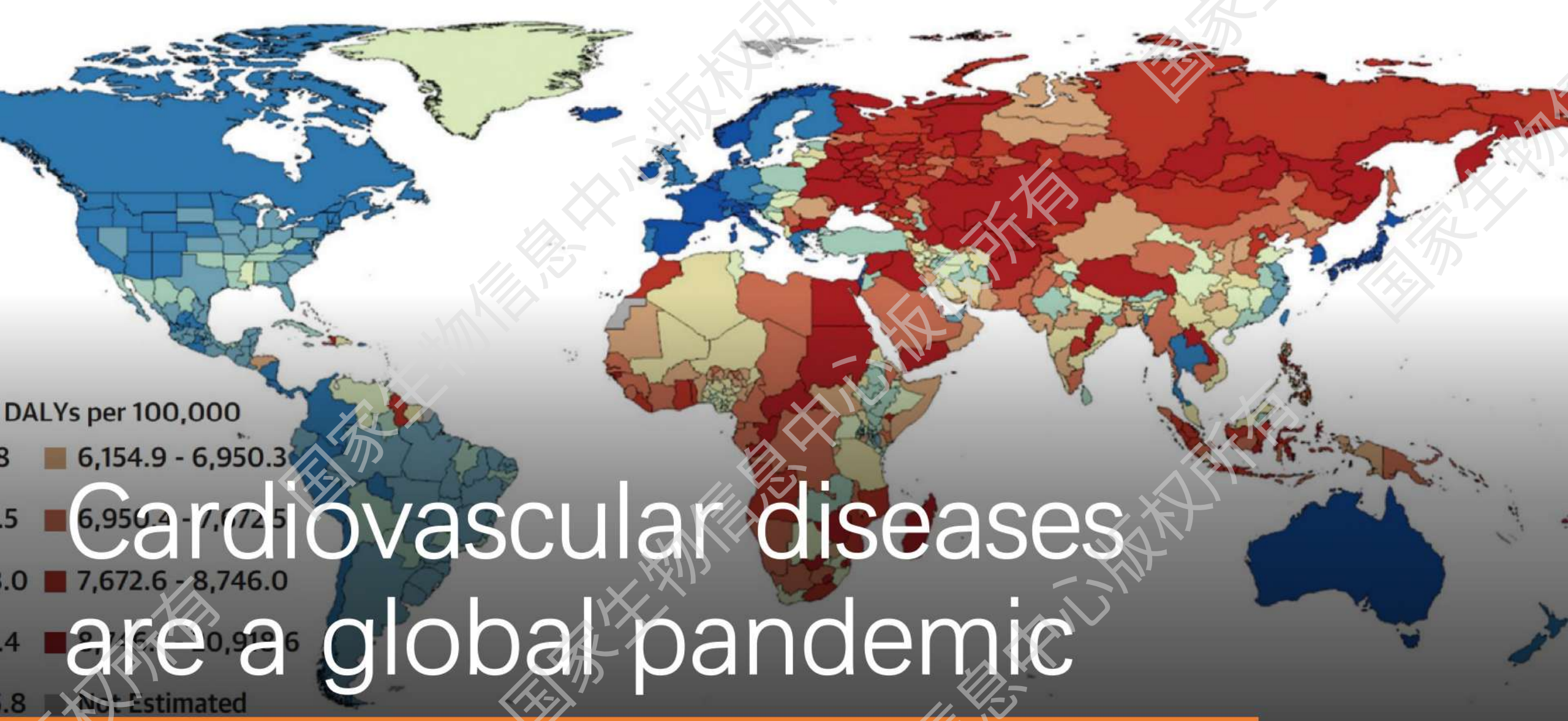
Jilin University
2005 – 2009
BSc
Animal Science

CAS-PICB
2009 – 2015
PhD
Computational Biology

Harvard Medical School
2015 – 2018
Postdoc
Medical Genetics

Broad Institute of MIT and Harvard
Massachusetts General Hospital
2018.01 – 2021.10
Computational Biologist

CAS-CNCB
2021.10 –
Principal Investigator
PhD Supervisor

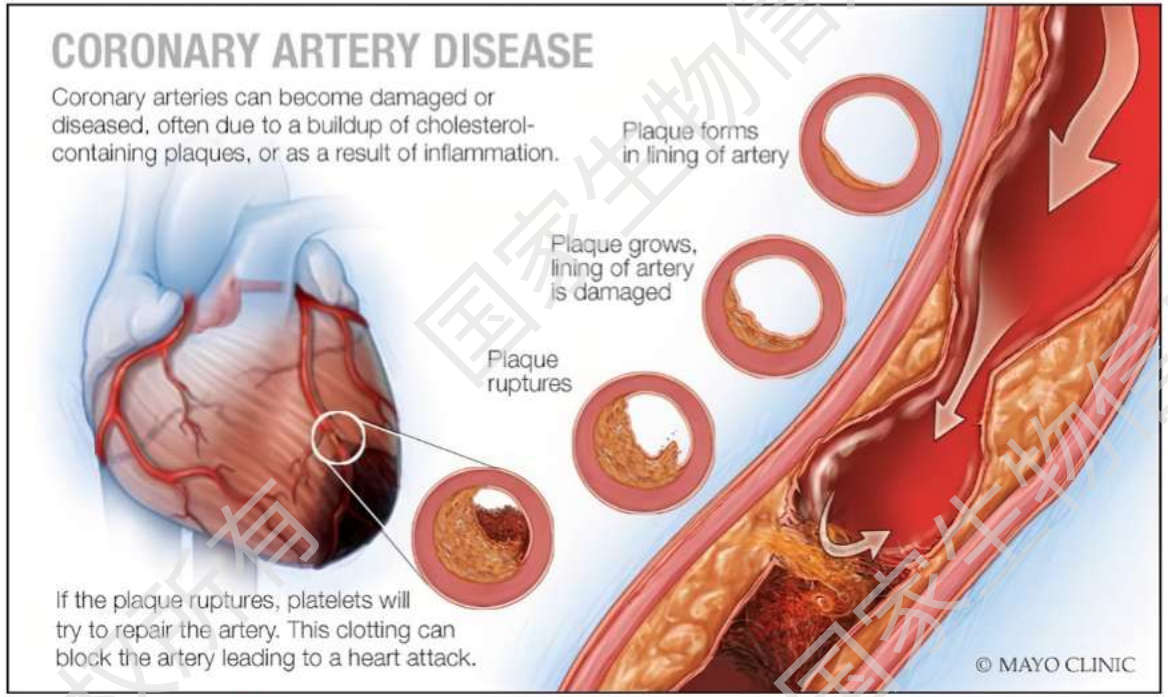


Cardiovascular diseases are a global pandemic

- ~1/2 of total death
- Coronary artery disease: ~1/4 death

Coronary Artery disease and atrial fibrillation

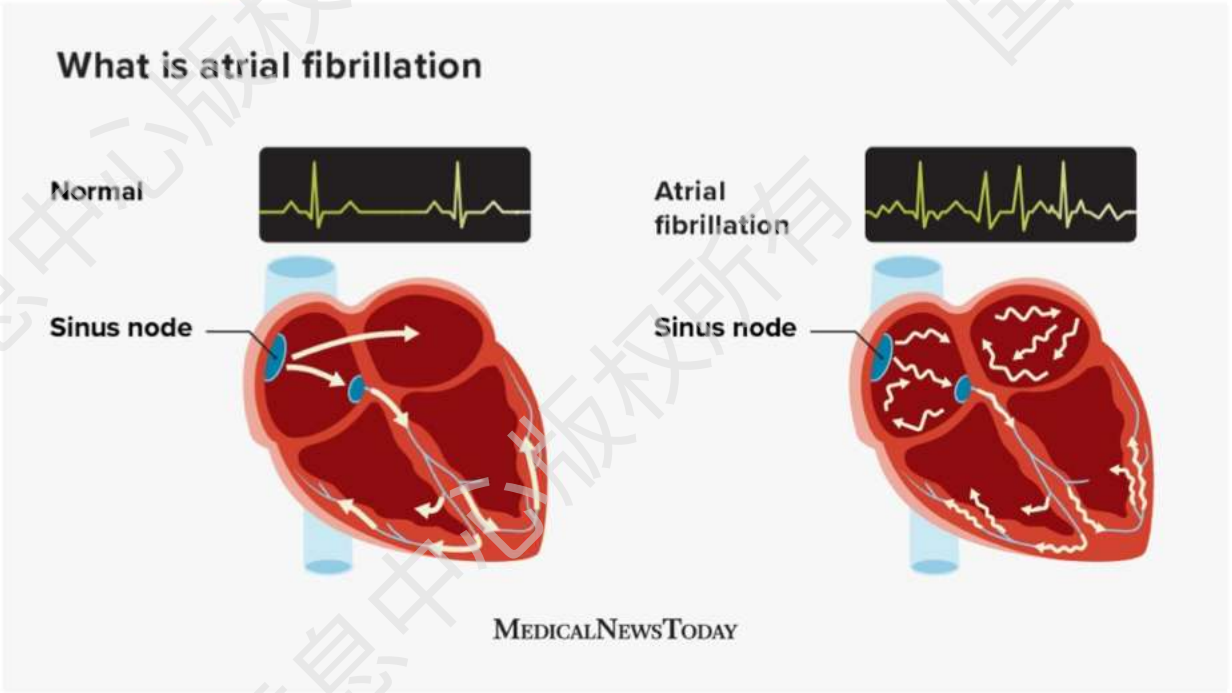
- **CAD: Atherosclerosis in the coronary arteries reduces blood flow to the heart**



Myocardial infarction

chest pain and tightness

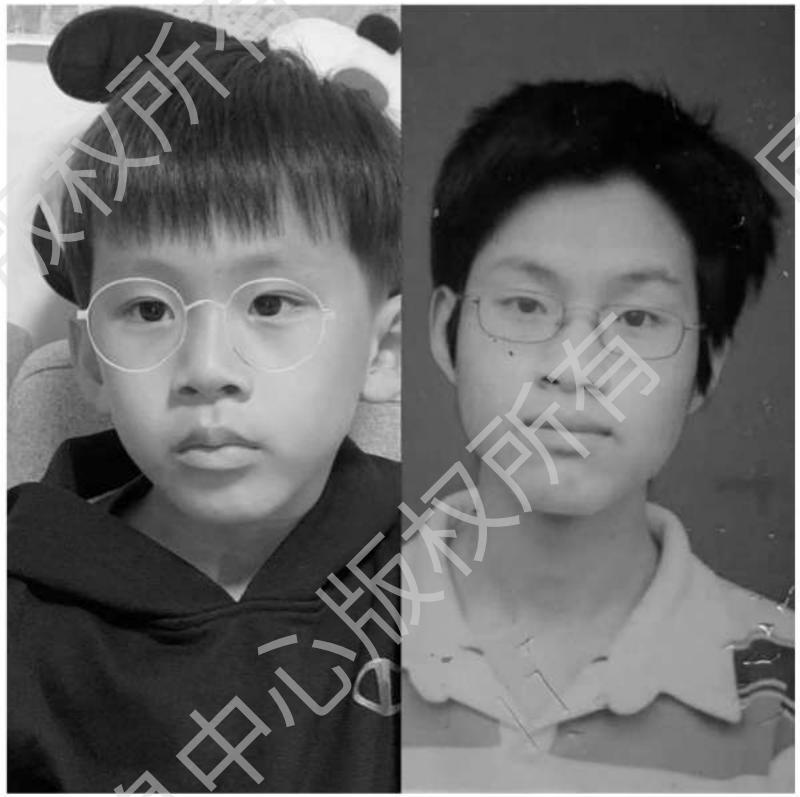
- **AF: Rapid, disordered electrical activity in the atria causes an irregular heartbeat**



Stroke incidence in AF patients: 17.5%

Stroke incidence in general population: 0.35%

Genes and Phenotypes



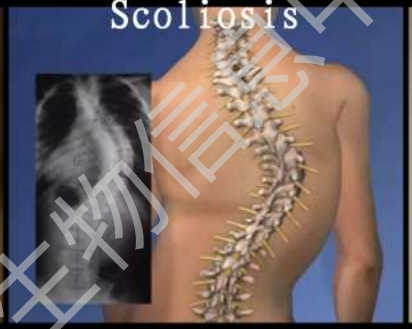
Aaron at 7 Me at 18

You reap what you sow: Genes have a strong determinative effect on phenotypes

Genes and Human Disease and Health

ALS

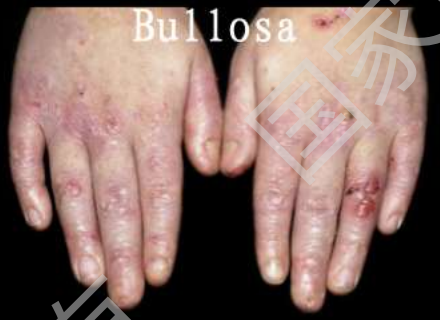
Congenital Scoliosis



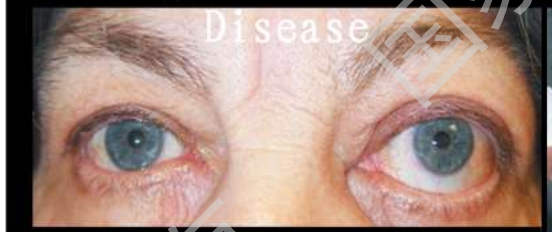
Gaucher Disease

Laron Syndrome

Epidermolysis Bullosa



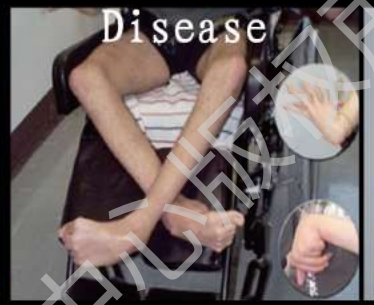
IgG4-Related Disease



Marfan Syndrome

ALS

Wilson Disease

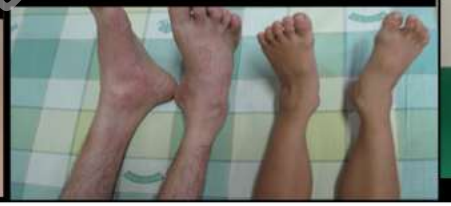


Albinism

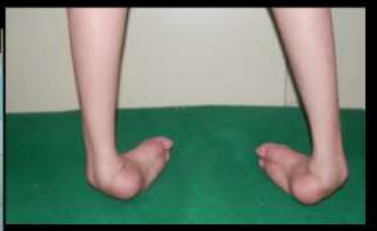
Hereditary Angioedema

Congenital Myasthenic Syndromes

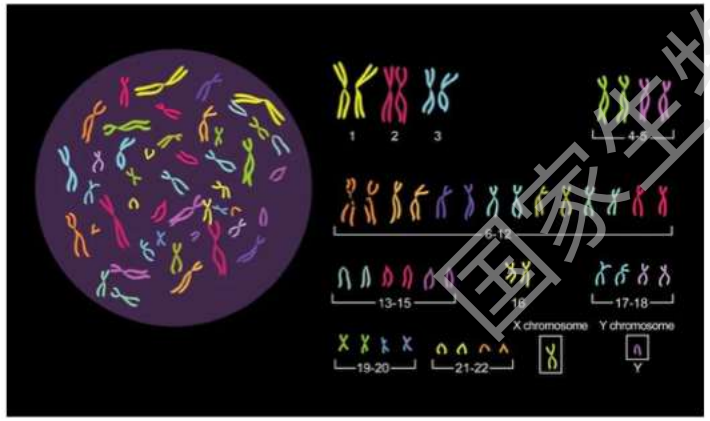
Hereditary Spastic Paraplegia



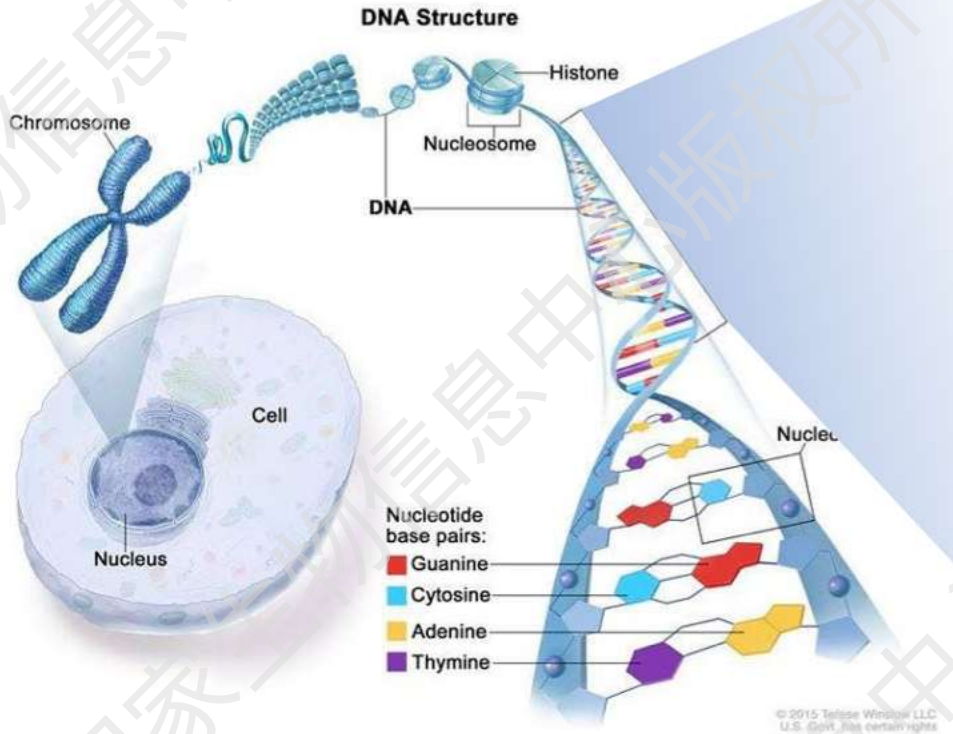
CMT



A gene is a functional sequence within the genome



Human Chromosome Map



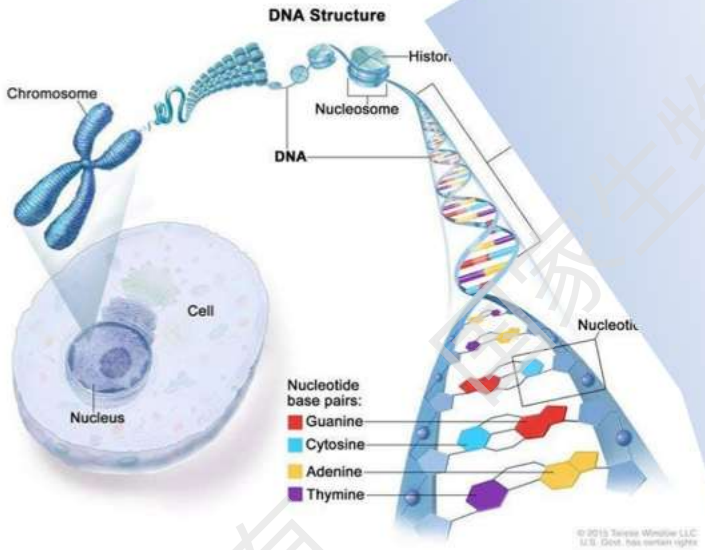

```
1    acccccccac cccaggccct aatgggccag gcggcagggg ttgagaggta
51   ggggagatgg gctctgagac tataaagcca gcgggggccc agcagccctc
101  AGCCCTCCAG GACAGGCTGC ATCAGAAGAG GCCATCAAGC AGgtctgttc
151  caagggcctt tgcgtcaggt gggctcagga ttccagggtg gctggacccc
201  aggccccagc tctgcagcag ggaggacgtg gctgggctcg tgaagcatgt
251  gggggtgagc ccagggggcc caaggcaggg cacctggcct tcagcctgcc
301  tcagccctgc ctgtctccca gATCACTGTC CTTCTGCCAT GGCCCTGTGG
351  ATGCGCCTCC TGCCCCTGCT GGCGCTGCTG GCCCTCTGGG GACCTGACCC
401  AGCCGCAGCC TTTGTGAACC AACACCTGTG CGGCTCACAC CTGGTGGAAG
451  CTCTCTACCT AGTGTGCGGG GAACGAGGCT TCTTCTACAC ACCCAAGACC
501  CGCCGGGAGG CAGAGGACCT GCAGGgtgag ccaactgccc attgctgccc
551  ctggccgccc ccagccaccc cctgctcctg gcgctcccac ccagcatggg
601  cagaaggggg caggaggctg ccacccagca gggggtcagg tgcacttttt
651  taaaaagaag ttctcttggt cacgtcctaa aagtgaccag ctccctgtgg
701  cccagtcaga atctcagcct gaggacggtg ttggcttcgg cagccccgag
751  atacatcaga gggtgggcac gctcctccct ccactcgccc ctcaaacaaa
801  tgccccgcag cccatttctc caccctcatt tgatgaccgc agattcaagt
851  gttttgttaa gtaaagtcct gggtgacctg gggtcacagg gtgccccacg
901  ctgcctgcct ctgggcgaac accccatcac gcccggagga gggcgtggct
951  gcctgcctga gtgggccaga cccctgtcgc caggcctcac ggcagctcca
1001 tagtcaggag atggggaaga tgctggggac aggccctggg gagaagtact
1051 gggatcacct gttcaggctc ccactgtgac gctgccccgg ggcgggggaa
1101 ggaggtggga catgtgggcg ttggggcctg taggtccaca cccagtgtgg
1151 gtgacctcc ctctaacctg ggtccagccc ggctggagat gggtgggagt
1201 gcgacctagg gctggcgggc aggcgggcac tgtgtctccc tgactgtgtc
1251 ctcctgtgtc cctctgcctc gccgctgttc cggaacctgc tctgcgcggc
1301 acgtcctggc agTGGGGCAG GTGGAGCTGG GCGGGGGCCC TGGTGCAGGC
1351 AGCCTGCAGC CCTTGGCCCT GGAGGGGTCC CTGCAGAAGC GTGGCATTGT
1401 GGAACAATGC TGTACCAGCA TCTGCTCCCT CTACCAGCTG GAGAACTACT
1451 GCAACTAGAC GCAGCCCGCA GGCAGCCCCA CACCCGCCGC CTCCTGCACC
1501 GAGAGAGATG GAATAAAGCC CTTGAACCAg ccctgctgtg ccgtctgtgt
1551 gtcttggggg ccctgggcca agccccactt cccggcactg ttgtgagccc
1601 ctcccagctc tctccacgct ctctgggtg
```

Human Insulin Gene

Four bases of genome:
A Adenine
G Guanine
C Cytosine
T Thymine

- **The human DNA molecule is approximately 2.2 meters long**
- **The human genome contains 3.2 billion BP**

A gene is a functional sequence within the genome

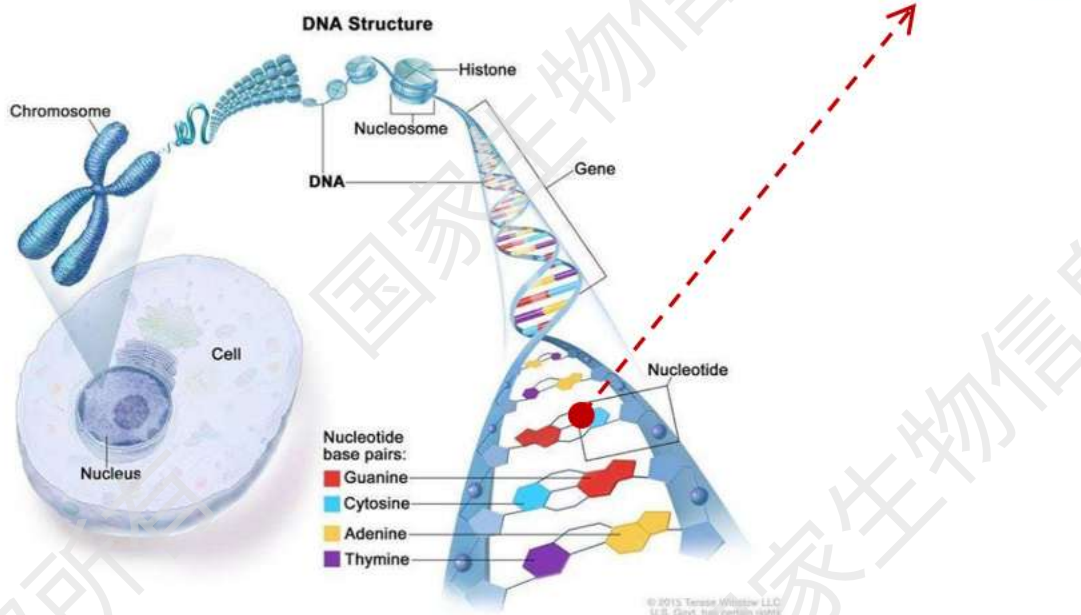


The human genome contains 3.2 billion BP

If you printed out the ~3 billion letters of the human genome in size 12 font, it would stretch from Houston to Boston!

Note: Only 20 letters of DNA sequence shown. To imagine what ~3 billion letters of the human genome sequence would look like, multiple this by 153 million!

How is our genome sequenced?

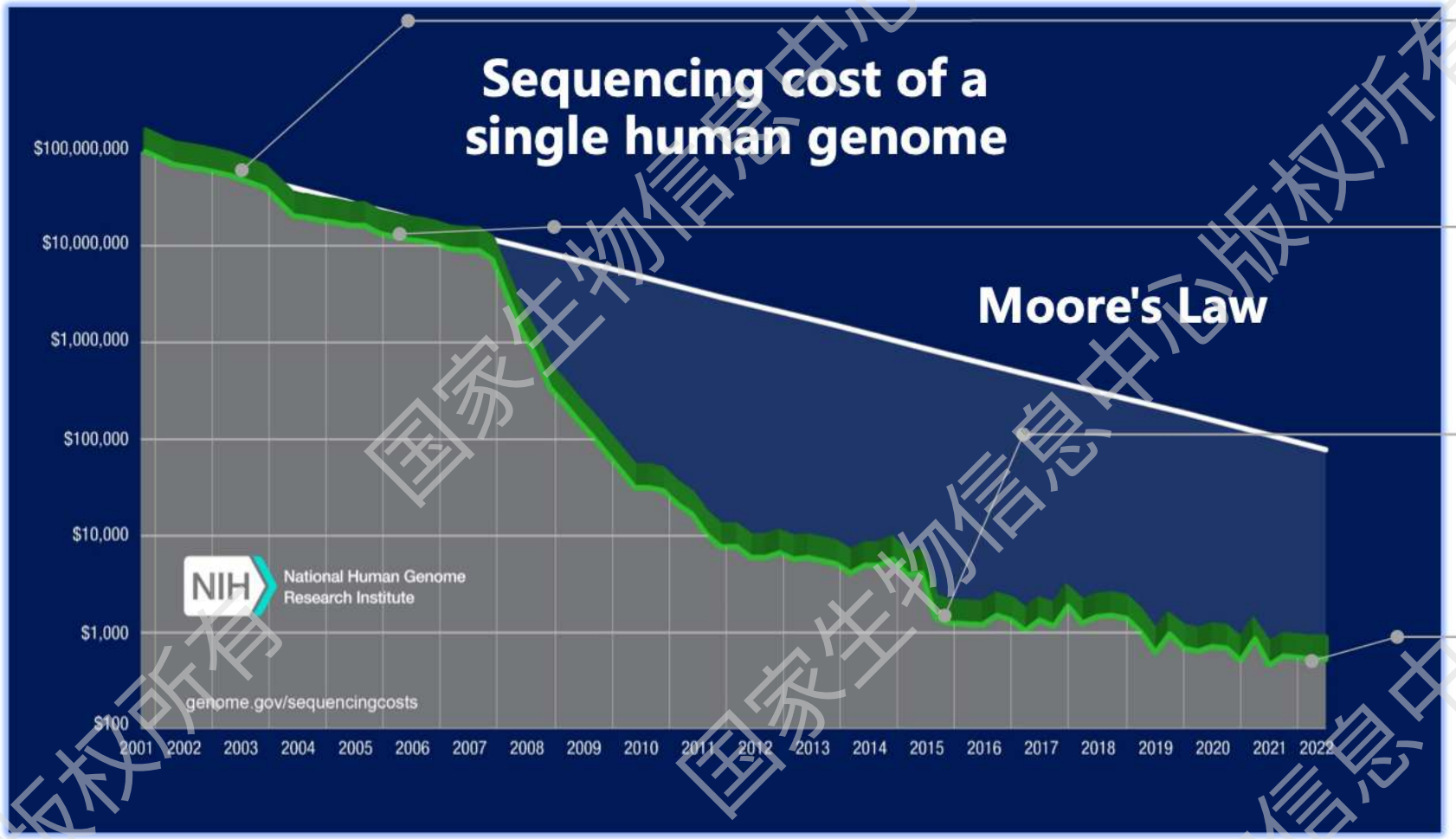


G → A

H1: GCTCATCGG--CTCGA
H2: GCTCATCGA--CTCTA

DNA Sequencer

Sequencing costs have decreased faster than Moore's Law



2003 — Human Genome Project-The first human genome- 13 years $3B

↓120-150 x

2006 — $20-25M per human genome

↓15,000 x

2015 — ~$15M per human genome

↓15 x

2022 — ~$200 per human genome

↓

2025 — < ¥1,000 per human genome

We are rapidly entering the era of "genomes for everyone"

Based on 30× sequencing depth

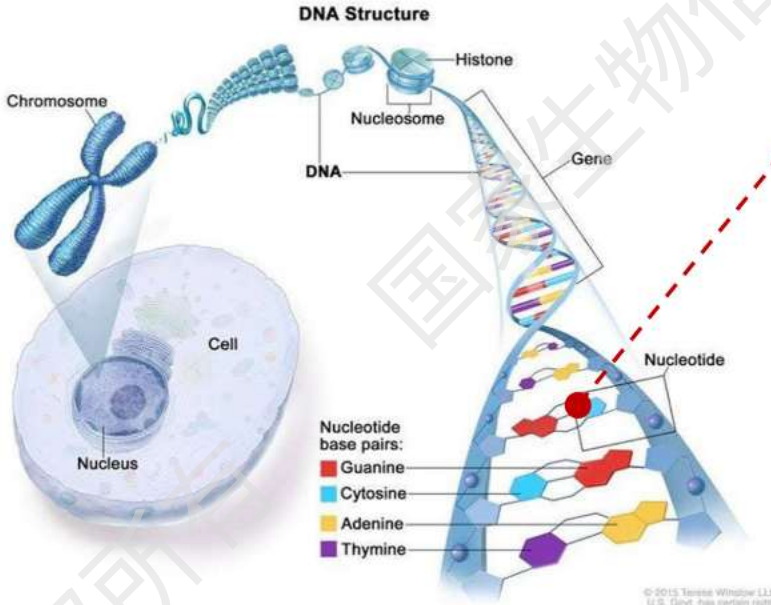
Credit: 谢晓亮院士

Genetic differences are the basis of disease and phenotypic variation

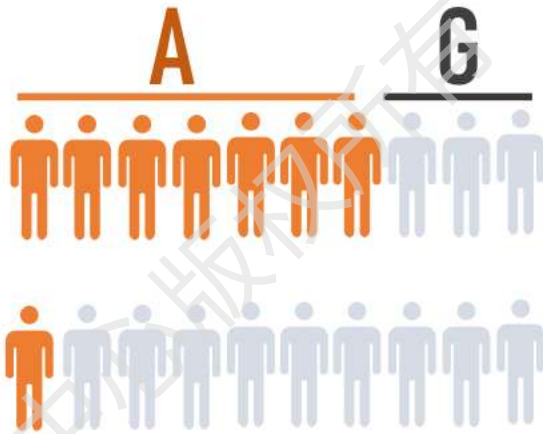
https://www.genome.gov/dna-day/15-ways/human-genomic-variation

The genome sequences of individuals differ by approximately 0.1%

Identifying gene–disease/phenotype associations



G → A



10000 samples
Case 70% A

10000 samples
Control 10% A

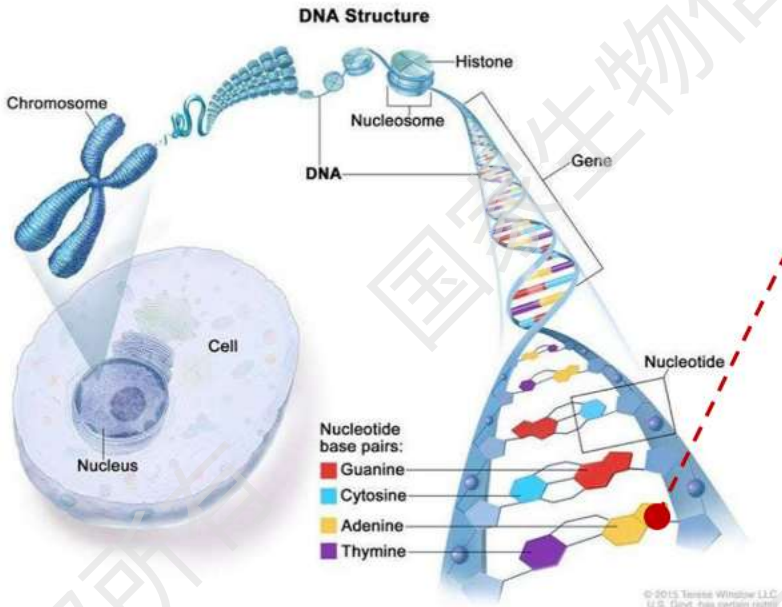
A

Increase disease risk

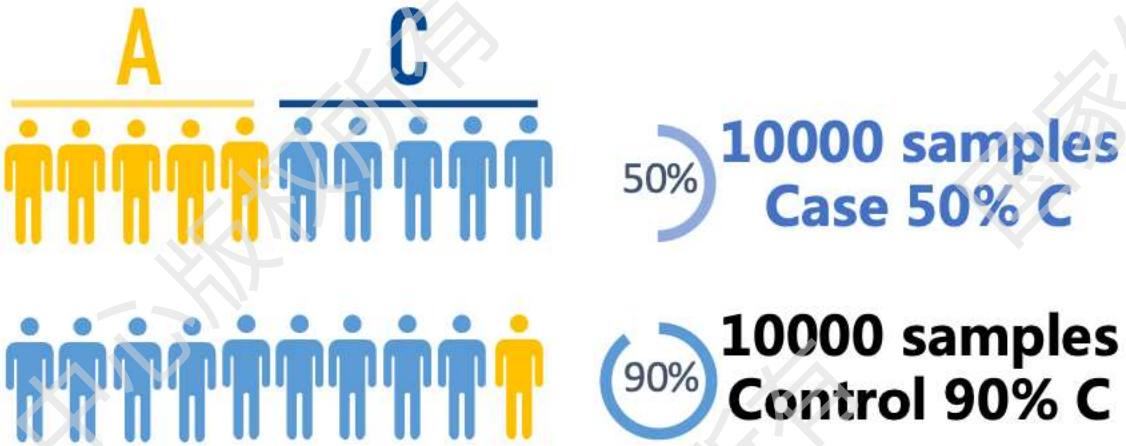
ALDH2,LDLR, APOB,LPL …

Khera, A. V. & Kathiresan, S.. *Nat. Rev. Genet* (2017)

Identifying gene–disease/phenotype associations



A → C



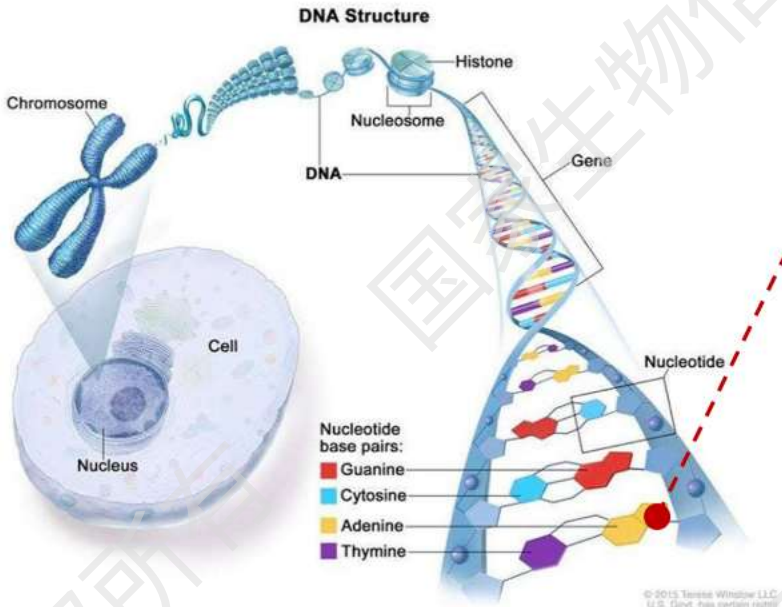
C

Reduce disease risk

PCSK9, ASGR1,NPC1L1 …

Khera, A. V. & Kathiresan, S.. *Nat. Rev. Genet* (2017)

Identifying gene–disease/phenotype associations



A → C

50% **10000 samples Case 50% C**

90% **10000 samples Control 90% C**

C

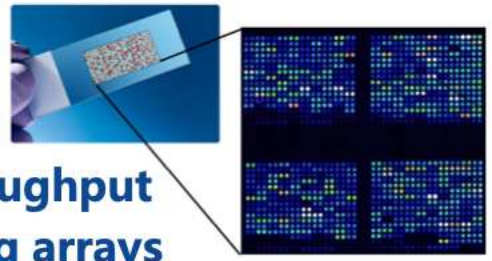
👉 **Not all mutations are harmful**

Reduce disease risk

PCSK9, ASGR1,NPC1L1 …

Khera, A. V. & Kathiresan, S.. *Nat. Rev. Genet* (2017)

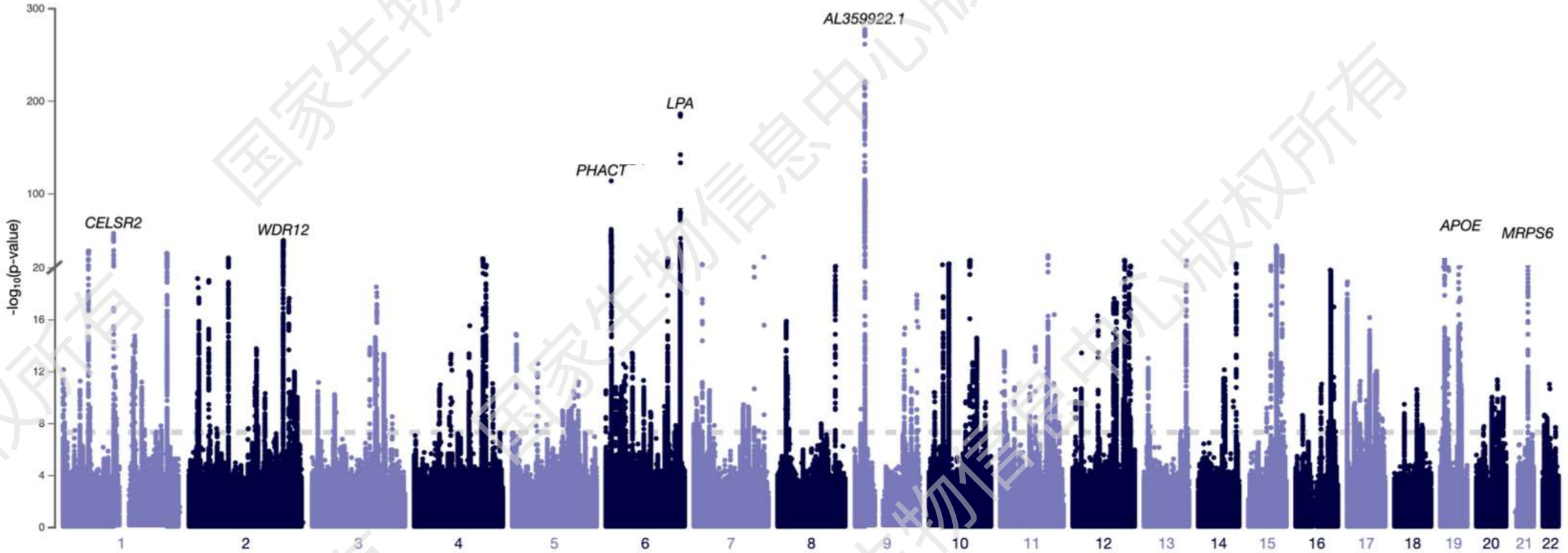
Coronary Artery Disease Genetic Map



High-throughput genotyping arrays

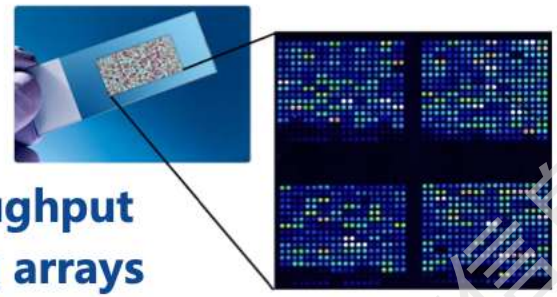
CARDIoGRAMplusC4D

180,000 CAD cases+ 1.17M controls GWAS ~ 300 genes



Aragam KG, et al., ***Nat Genet***. 2022;

CAD Genetic Map and Medications



High-throughput genotyping arrays

CARDIoGRAMplusC4D

180,000 CAD cases+ 1.17M controls GWAS

~ 300 genes

LDL-C

PCSK9 Inclisiran

Statin *HMGCR*

Aragam KG, et al., ***Nat Genet***. 2022;

Gene-targeted drugs are revolutionizing CAD treatment and prevention

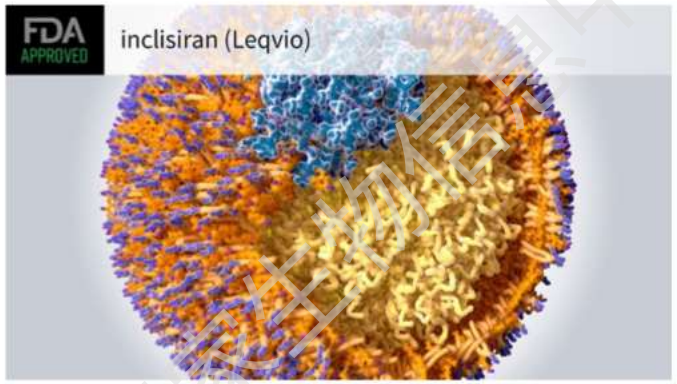
Every day
Statins

→

Every six months
PCSK9 siRNA

→

Once in a lifetime
Gene editing



Novaritis Leqvio(inclisiran)
Long-term lipid-lowering

NEJM. 2020
medpagetoday.com

verve THERAPEUTICS

The unmet medical needs

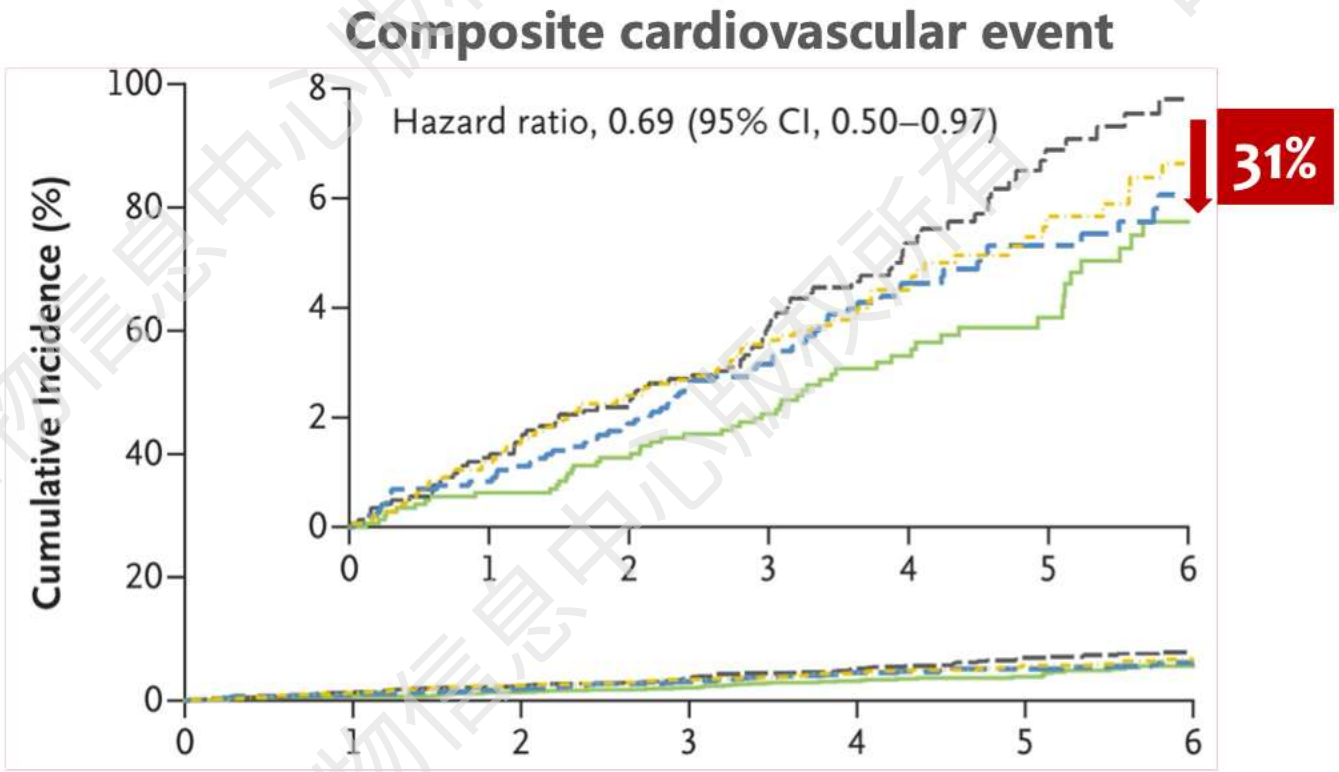
Can we
treat
every CAD
patient
effectively?

The unmet medical needs

Can we treat every patient effectively?

- Lipid-lowering
- Blood-pressure-lowering
- Antiplatelet

↓ 30~55%



5,713 individuals | 4.6 years of follow-up

Yusuf S, et al. N Engl J Med. 2021

MEDICAL NEEDS

Identify new genetic drug targets for CAD

Risk Factors for Cardiovascular Disease

Genetic


Monogenic


Polygenic


Family history


Somatic mutation


Telomere length

Environment


Smoking


Drinking


Diet


Aging


Exercise


Sleep

CAD


Biobank Japan

50% ~ 60%

Heritability

32k cases
146k controls

***ALDH2*, rs671 (p.Glu504Lys)**
Freq 24%, OR = 1.22, $P < 6.4^{-70}$

Koyama et al., ***Nat Genet***, 2020

AF

biobank uk
Improving the health of future generations

UK Biobank

60 ~ 62%

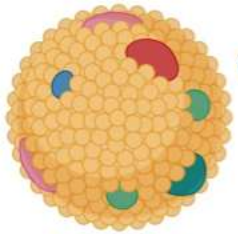
Heritability

30k cases
416k controls

***CDH2*, Rare LoF Variants**
OR = 10.23, $P = 5.3^{-3}$

Zöller B, et al., ***Circ Genomic Precis Med***, 2024

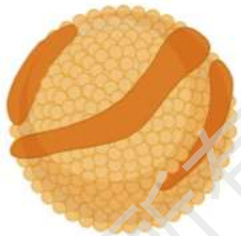
Challenges in Developing New Drugs for Cardiovascular Disease



HDL

Good cholesterol associated with reduced CAD risk

Lilly ~$800m	CSL ~$1B	MSD ~$500m	Pfizer ~$800m
>7 years	~6 years	~4 years	>6 years
Phase III	Phase III	Not marketed	Phase III
Abandoned	Failed		Failed
Evacetrapib	CSL-112	Anacetrapib	Torcetrapib



LDL

Bad cholesterol associated with increased CAD risk

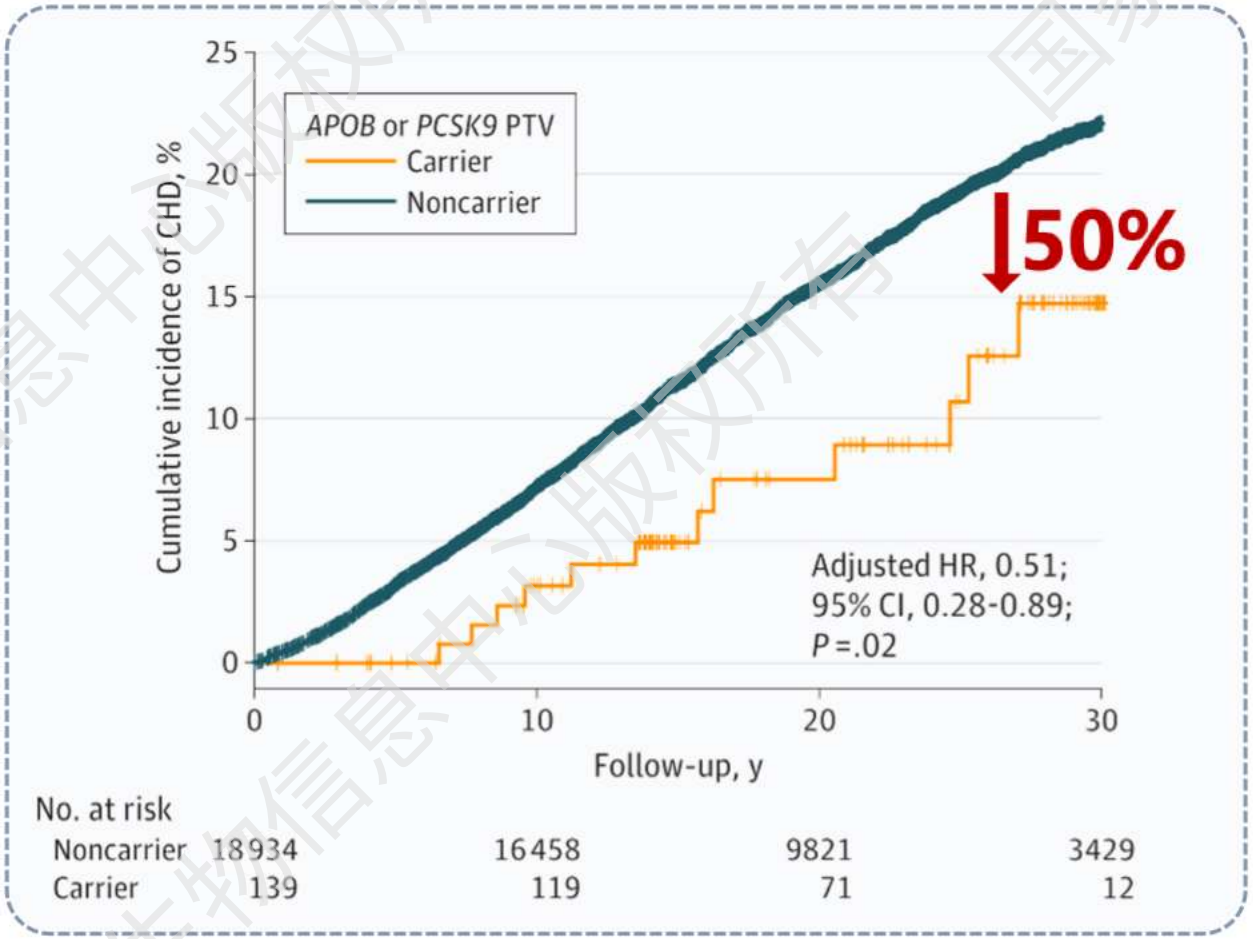
Pfizer ~$1B	MSD ~$600m	AMGEN >$1B	NOVARTIS ~$2B
~11 years	~6 years	~7 years	>7 years
Success	Success	Success	Success
Atorvastatin	Ezetimibe	Evolocumab	Inclisiran

How to increase clinical trial success rates, shorten development time, and reduce R&D costs?

Identifying gene–disease/phenotype associations

Gene Mutations Nature's Randomized Clinical Trials

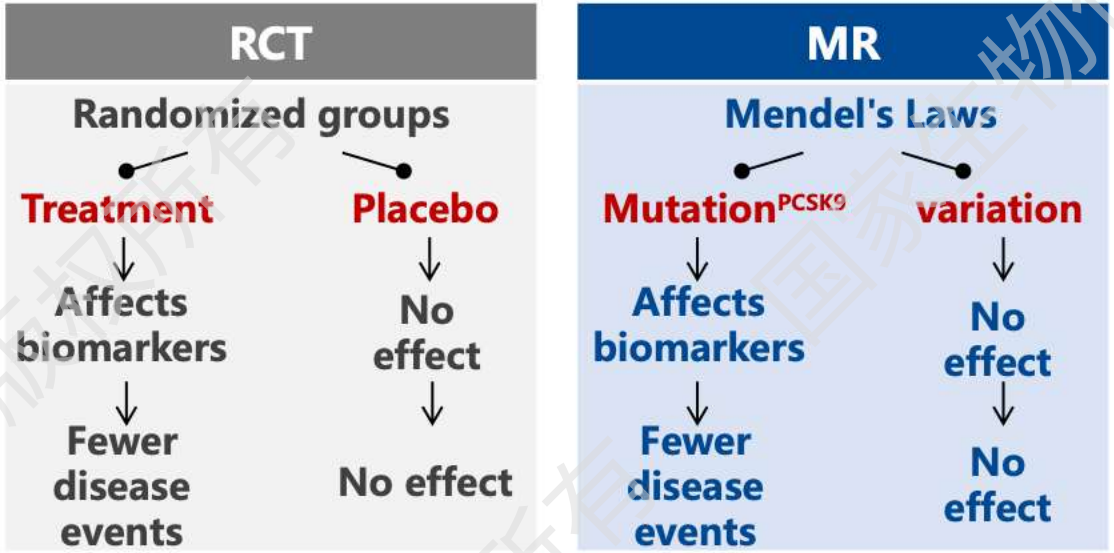
Loss-of-function mutations in PCSK9
Reduce the risk of CAD by 50%



Dron JS, et al. *JAMA Cardiol*. 2023.

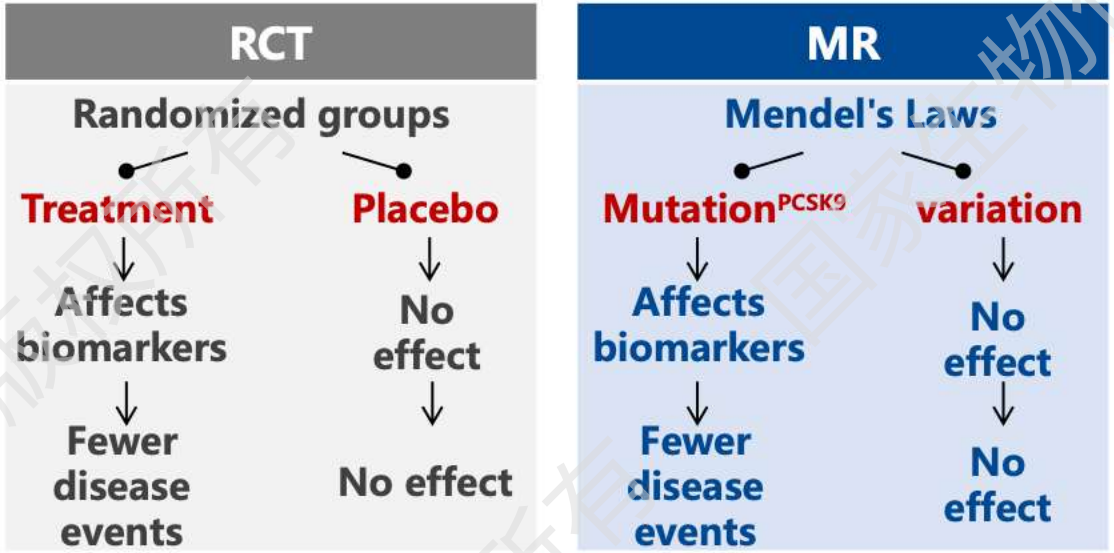
New Opportunities in Drug Development from Population-Scale Genomic Data

- Genetics influences the fundamental mechanisms of disease
- Represents lifelong exposure to risk factors
- Not affected by confounding factors
- Helps clarify causal relationships
- Helps assess safety issues



New Opportunities in Drug Development from Population-Scale Genomic Data

- Genetics influences the fundamental mechanisms of disease
- Represents lifelong exposure to risk factors
- Not affected by confounding factors
- Helps clarify causal relationships
- Helps assess safety issues



2021 FDA approvals: 50 drugs, ~2/3 supported by genetic evidence

Ochoa et al., Nat Rev Drug Discov. 2022;

Genetic evidence increases drug development success 2–3 fold

Nelson MR, Tipney H, Painter JL, *et al. Nat Genet.* 2015;
Minikel EV, Painter JL, Dong CC, Nelson MR. *Nature.* 2024

Top pharma companies invest in population cohorts

MERCK Genentech sanofi NOVARTIS Celgene

€80M

500k

FINNGEN

gsk GlaxoSmithKline

$300M

23andMe®

AMGEN AstraZeneca gsk GlaxoSmithKline Johnson&Johnson

Bristol Myers Squibb™ AstraZeneca REGENERON science to medicine® Biogen abbvie Pfizer Alnylam PHARMACEUTICALS Takeda

$100M

uk biobank

UK Biobank Allele Frequency Browser

500k samples WGS

$150M

500k WES

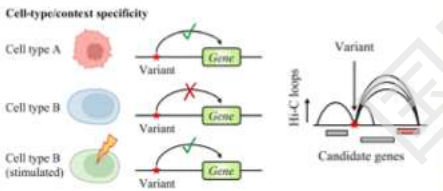
$50M

gsk GlaxoSmithKline

2016: 60% of new drug targets supported by human genetics

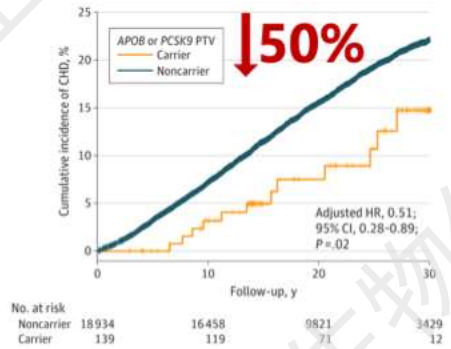
From genes to promising drug targets

Gene



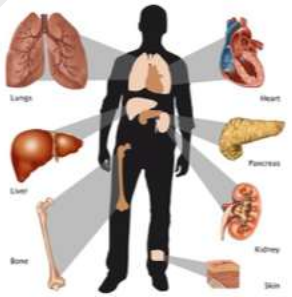
Identification of candidate genes

Effectiveness



Validation reducing disease risk

Safety



without nonspecific side effects

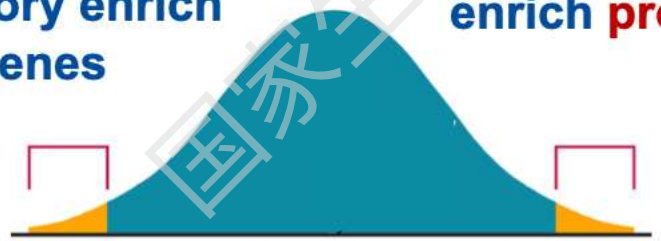
Drug targets

Promising drug target

Large families + Extreme phenotype + Large population cohort

Early onset and family history enrich risk genes

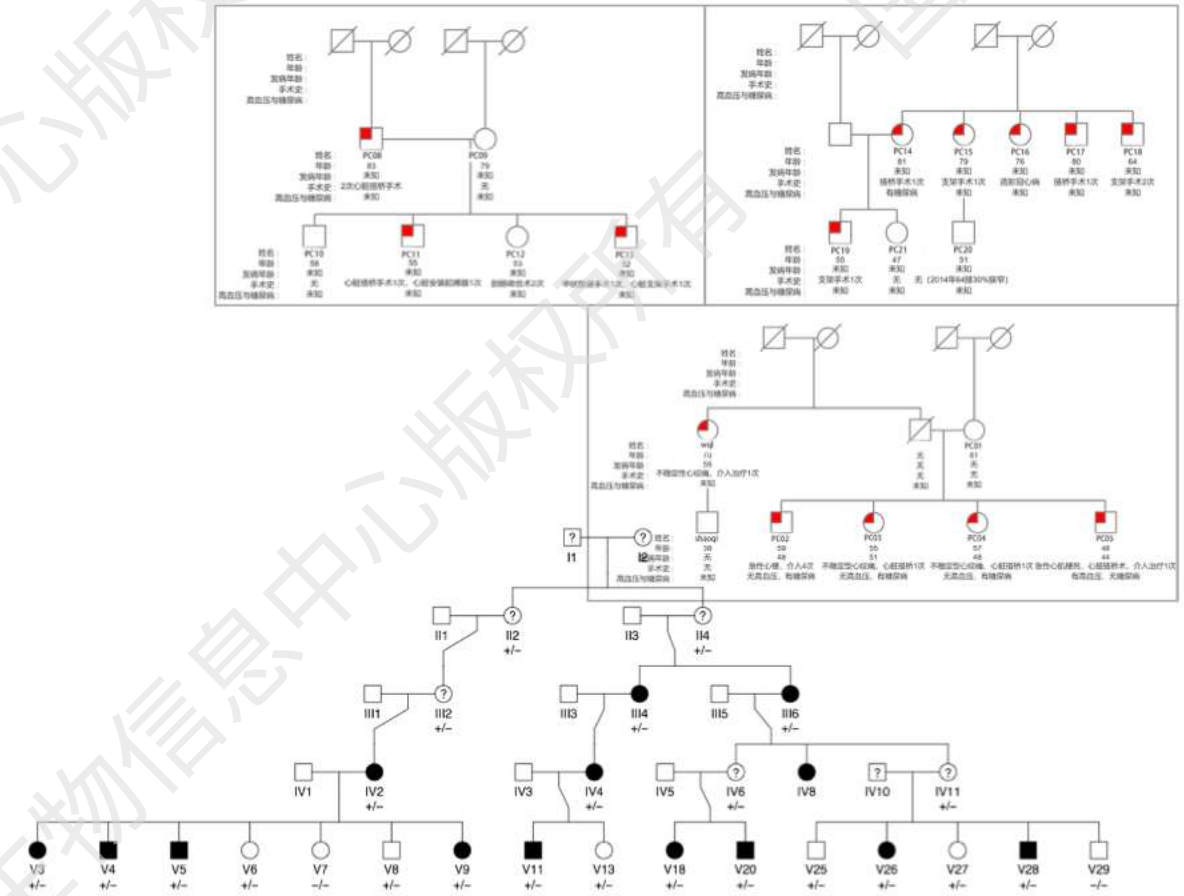
Old health cohort enrich protective genes

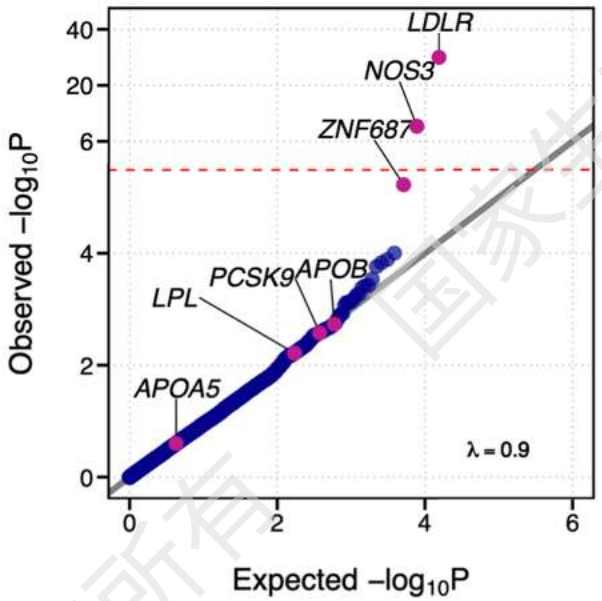


Extreme phenotype cohort

- **Early onset CAD （1k WGS）**
  - **Male ≤45 years, female ≤45 years, with a family history**
  - Low conventional risk, eg. blood pressure and blood lipids
  - No smoking, no drinking, and with a good lifestyle
- **Old health cohort （1k WGS）**
  - >75 years, good vascular status, no obvious signs of plaque

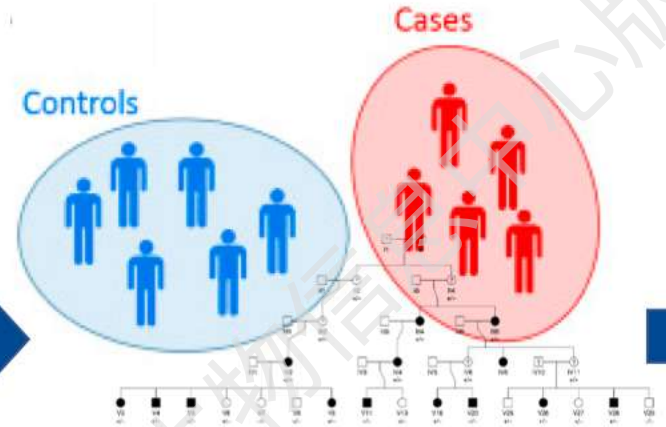
large families





Phase 1
41K CAD cases
217K controls

Circ Genomic Precis Med. 2022



Phase 2
50K CAD cases
510K controls

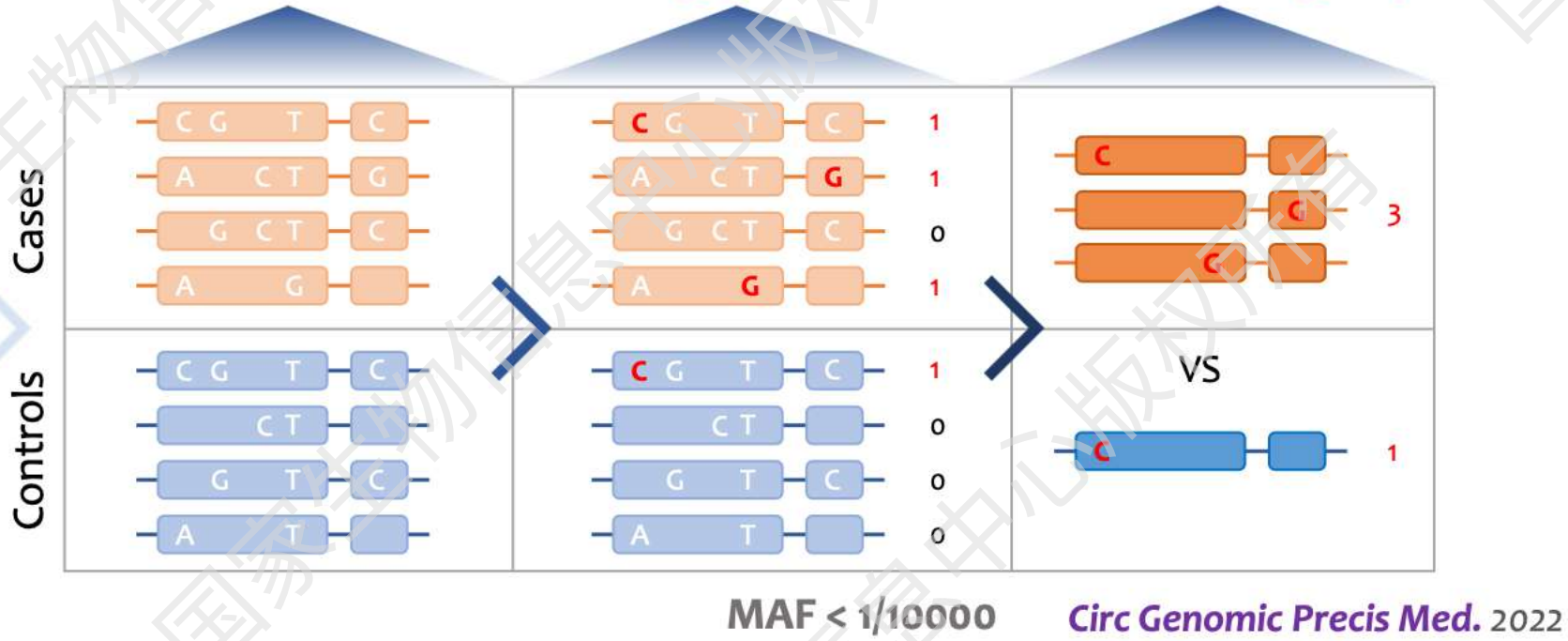
Phase 3
850K cohort

BROAD INSTITUTE
MASSACHUSETTS GENERAL HOSPITAL

Gene coding variants enrichment analysis

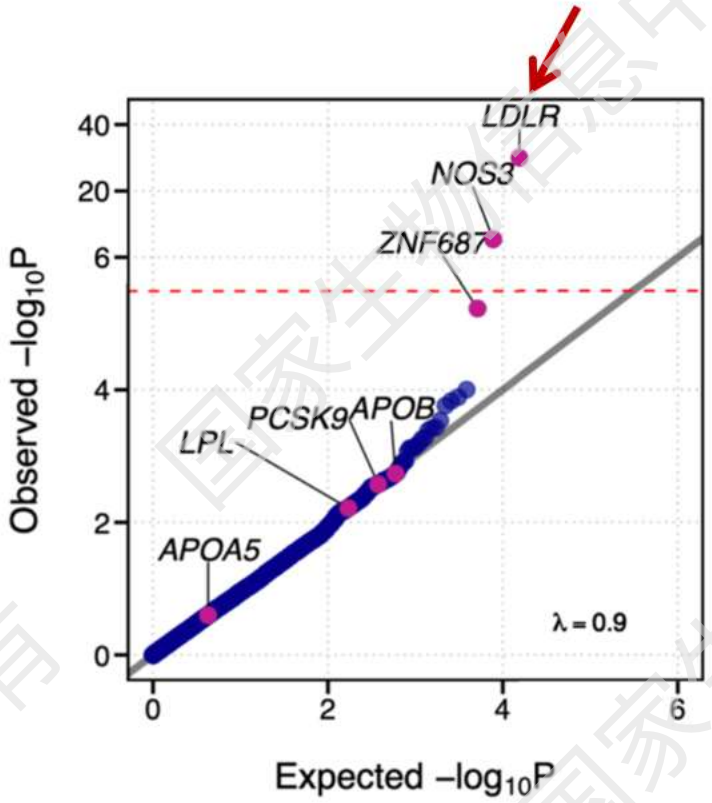
41K CAD cases

220K controls



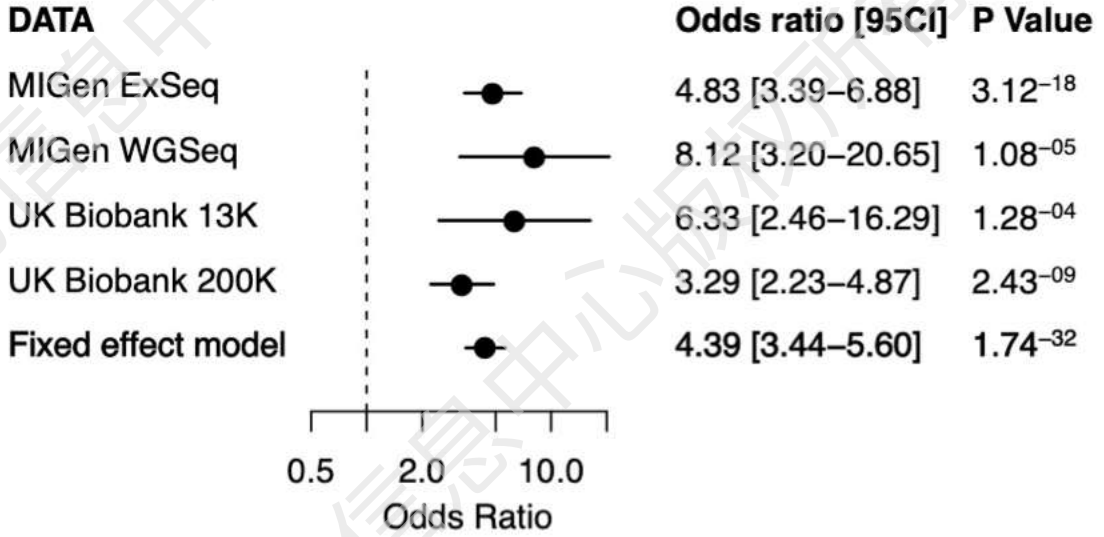
Circ Genomic Precis Med. 2022

Gene coding variants enrichment analysis



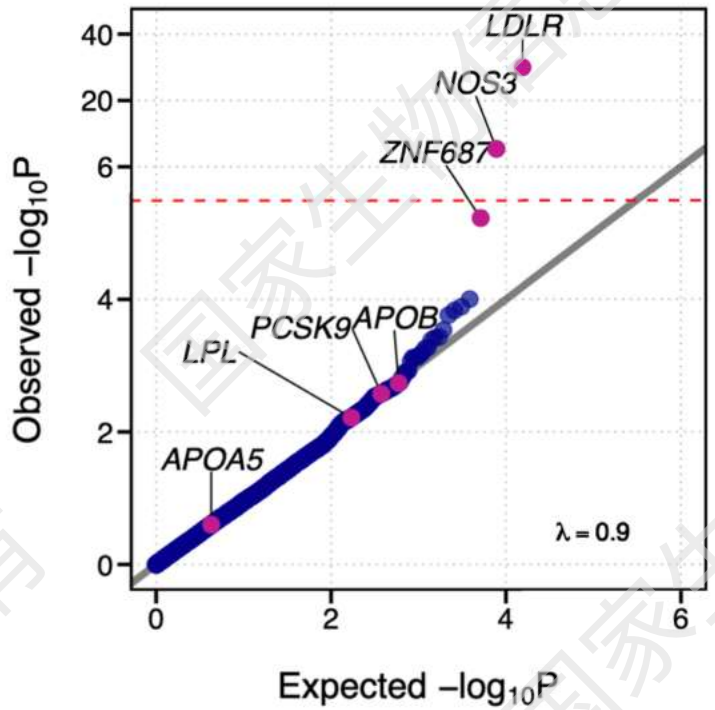
LDLR

- **Odds ratio 4.39**
- **P = 1.74 x 10^{-32}**
- **0.91% cases vs 0.34% controls**



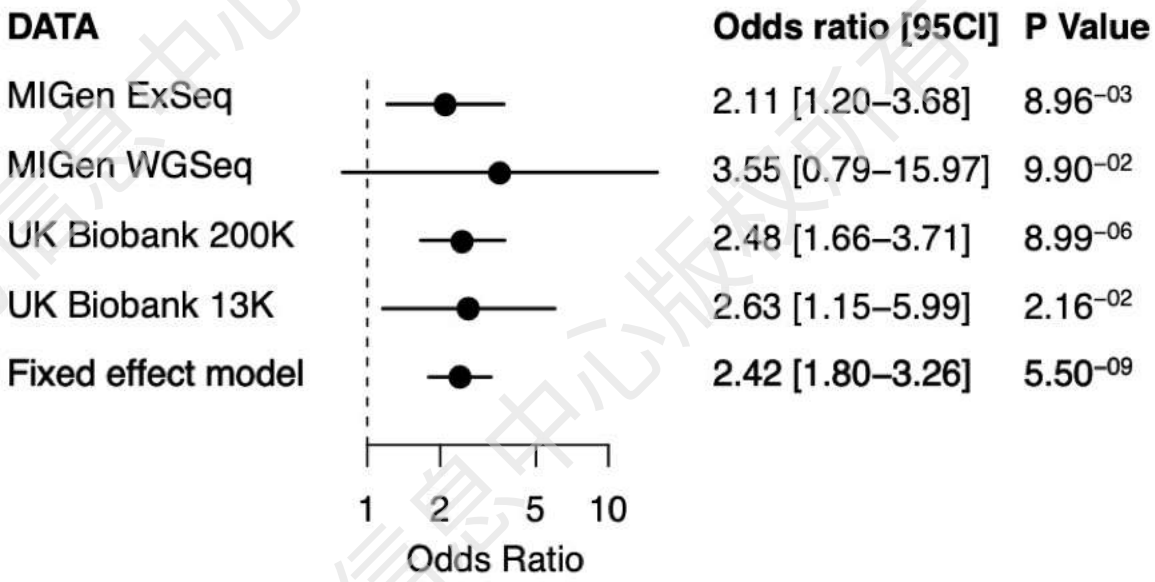
汪敏先*等, *Circ Genomic Precis Med.* 2022

Gene coding variants enrichment analysis



Nitric oxide synthase 3 (*NOS3*)

- **Odds ratio 2.42**
- **P = 5.5 x 10^{-9}**



20k protein-coding genes: 0.05/20000 = 2.5^{-6}

Circ Genomic Precis Med. 2022

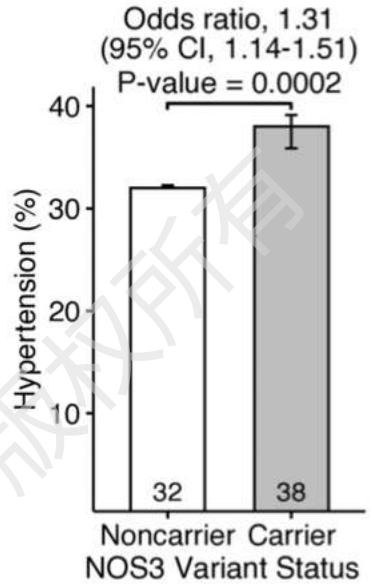
Molecular phenotypes dissect diease mechanisms

Association between NOS3 rare variants status with lipid and blood pressure biomarkers

TRAIT (unit)	Beta	*P-value*
LDL cholesterol* (mg/dL)	+ 0.8	0.47
HDL cholesterol* (mg/dL)	- 0.2	0.64
Total cholesterol* (mg/dL)	- 0.4	0.76
Triglycerides* (mg/dL)	- 0.2	0.94
Systolic blood pressure† (mmHg)	+ 3.3	5.0^{-6}
Diastolic blood pressure† (mmHg)	+ 1.3	7.0^{-4}

UKB lipid/blood pressure data of 200,000 individuals

Carriers of NOS3 rare variants 1.31-fold increased risk of hypertension



Pathway biomarkers indicate *NOS3* is lipid-independent

Genomic-based patients subtyping and treatment

Circ Genomic Precis Med. 2022

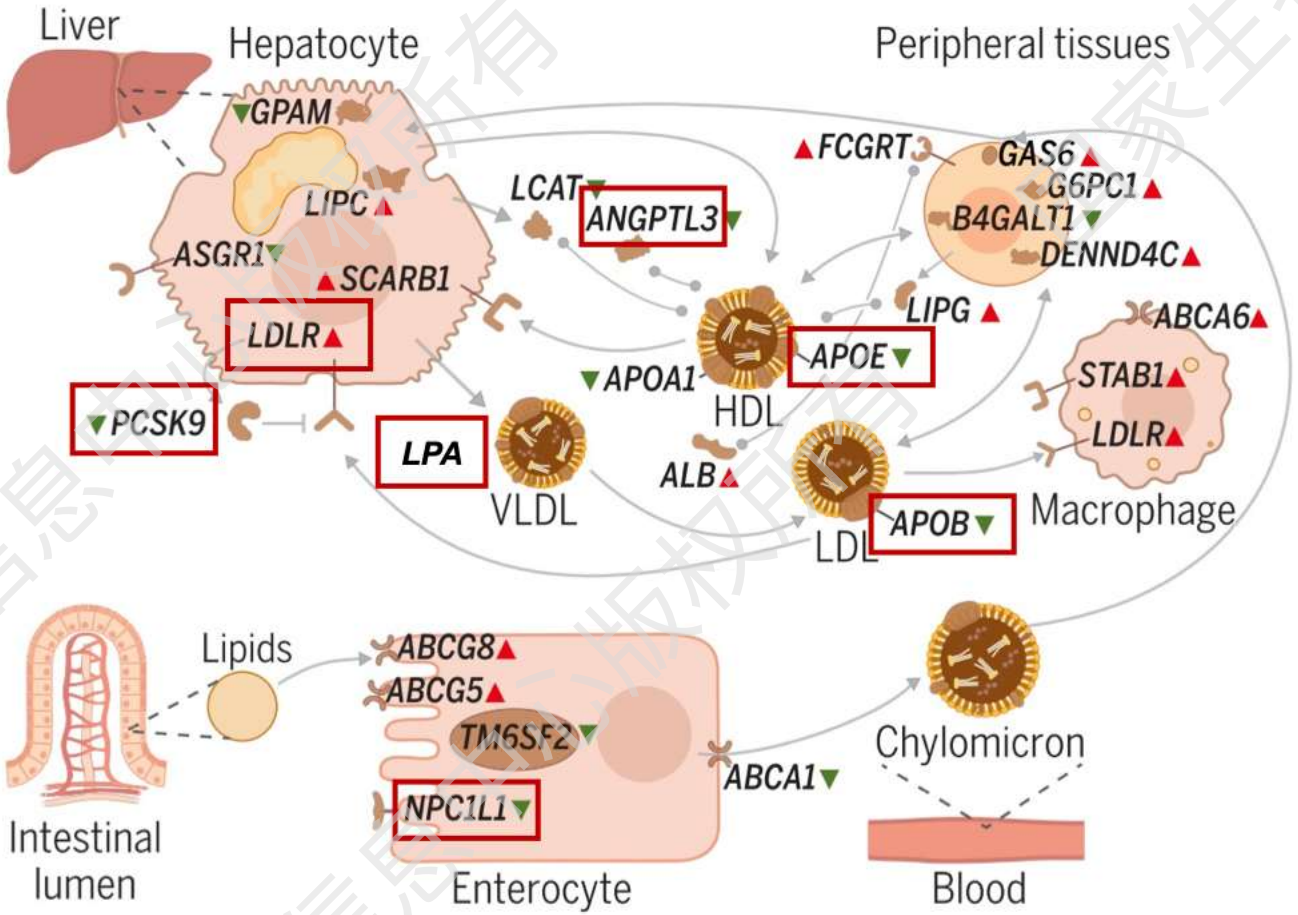
Preliminary results of the phase 2 study

Phase 2
50K CAD cases
510K controls

Rank	Gene	P Value
1	LPA	2.3^{-48}
2	LDLR	2.3^{-27}
3	APOE*	2.3^{-24}
4	PCSK9*	2.3^{-14}
6	LPL*	3.3^{-8}
13	ANGPTL4	6.5^{-7}
20	APOB*	3.3^{-6}
	NOS3	1.8^{-6}
	NPC1L1*	3.3^{-3}

* Genetic variants reduce disease risk

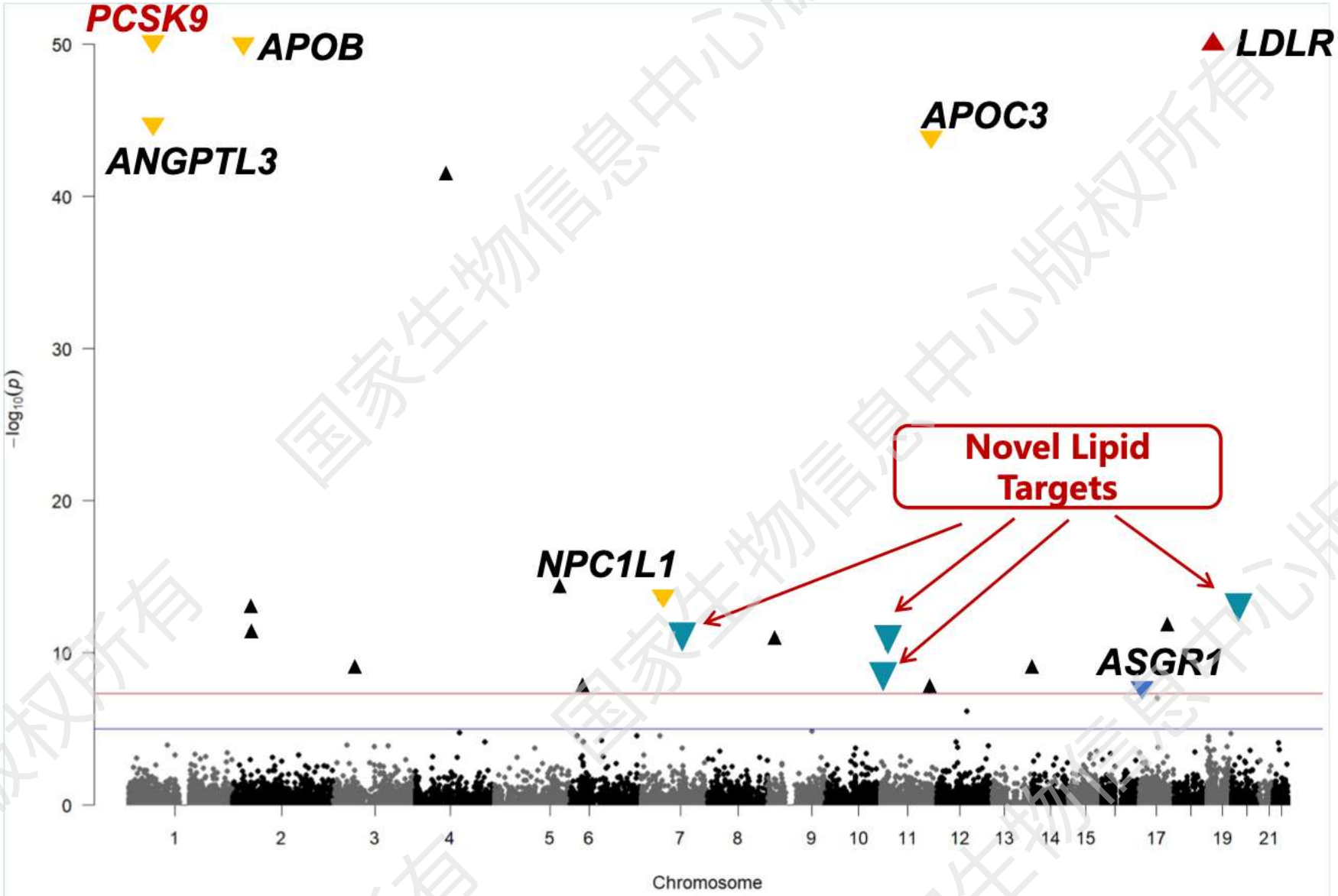
Unpublished



Human lipid pathway

Fiziev P, McRae J, Dron J, et al. ***Science***, 2023.

Genetic Map of Low-Density Lipoprotein Cholesterol in 500,000 Individuals



biobank uk

UK Biobank
500k LDL Burden Test
mask1 : lofHC

- beta > 0, Elevated lipid levels
- beta < 0, Reduced lipid levels
- Well-established genes
- FDA approved
- Clinical trials
- Novel lipid targets

Unpublished

廖双巧

[Rare] protective genetic variants guide the successful development of new drugs

Population cohort

Reduced risk of CAD

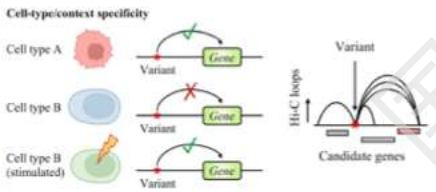
Drug targets

Gene	*PCSK9*	*APOC3*	*NPC1L1*	*ANGPTL3*
Mutation frequency	1/50	1/150	1/650	1/300
Performance	80% ↓risk	40% ↓risk	53% ↓risk	34% ↓risk
FDA approved Drugs	McAb: Alirocumab, Evolocumab ASO: Inclisiran	ARO-APOC3 (RNAi)	Ezetimibe	EvinaCumab

NEJM. 2015, 2017, 2018, 2021

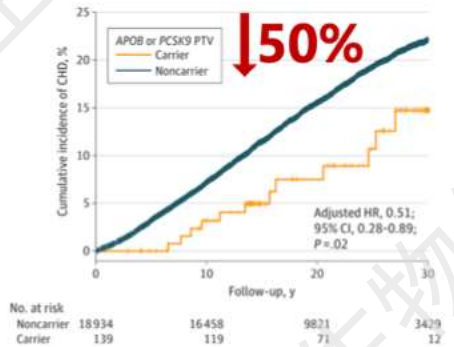
From candidate genes to promising drug targets

Gene



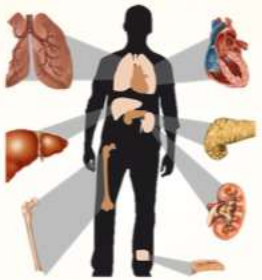
Identification of candidate genes

Effectiveness



Validation reducing disease risk

Safety

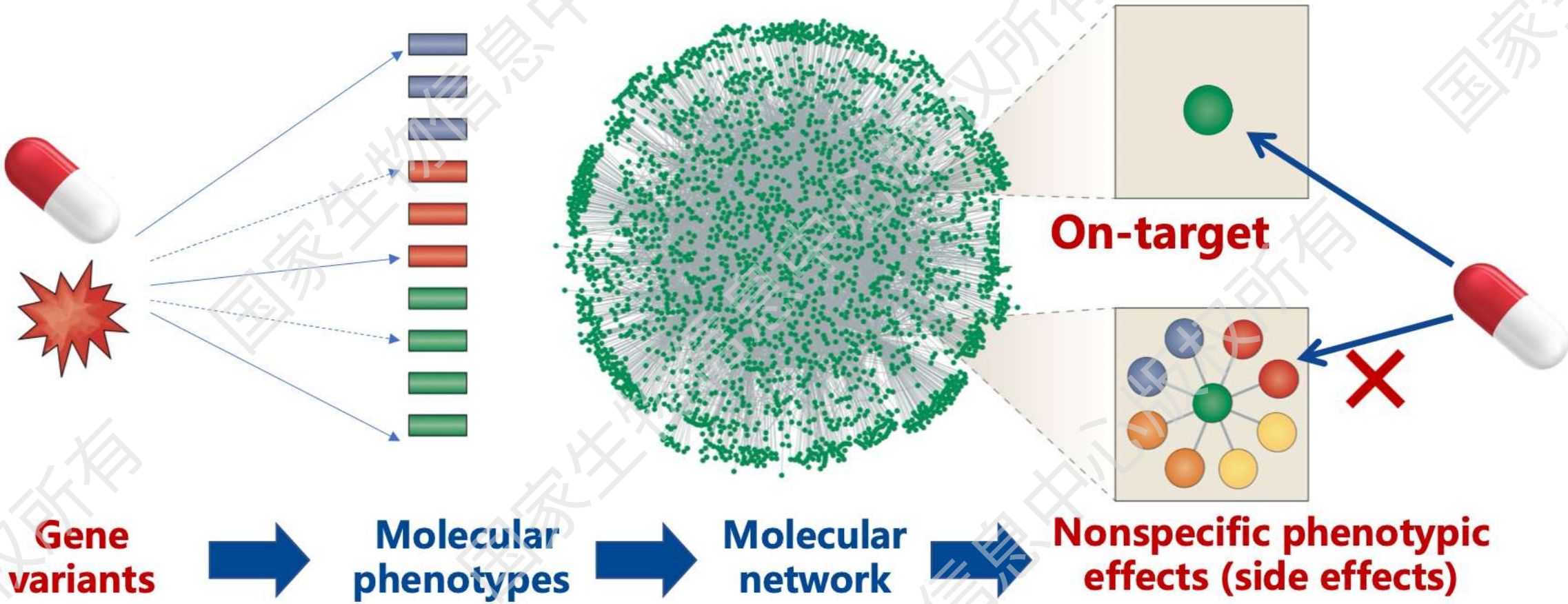


without nonspecific side effects

Drug targets

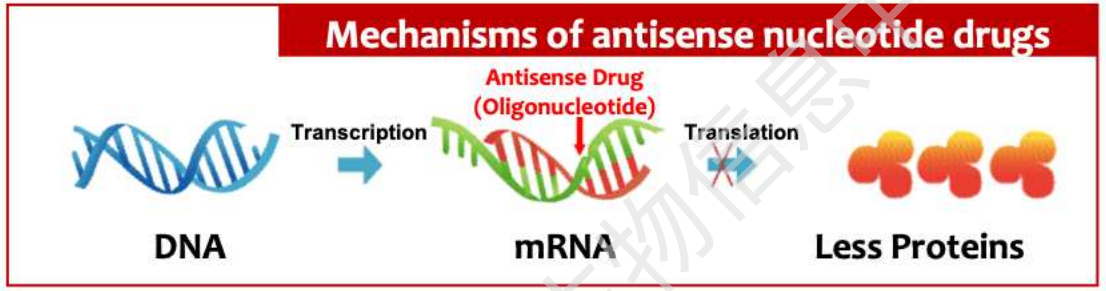
Promising drug target

Drug safety assessment by molecular phenotypes



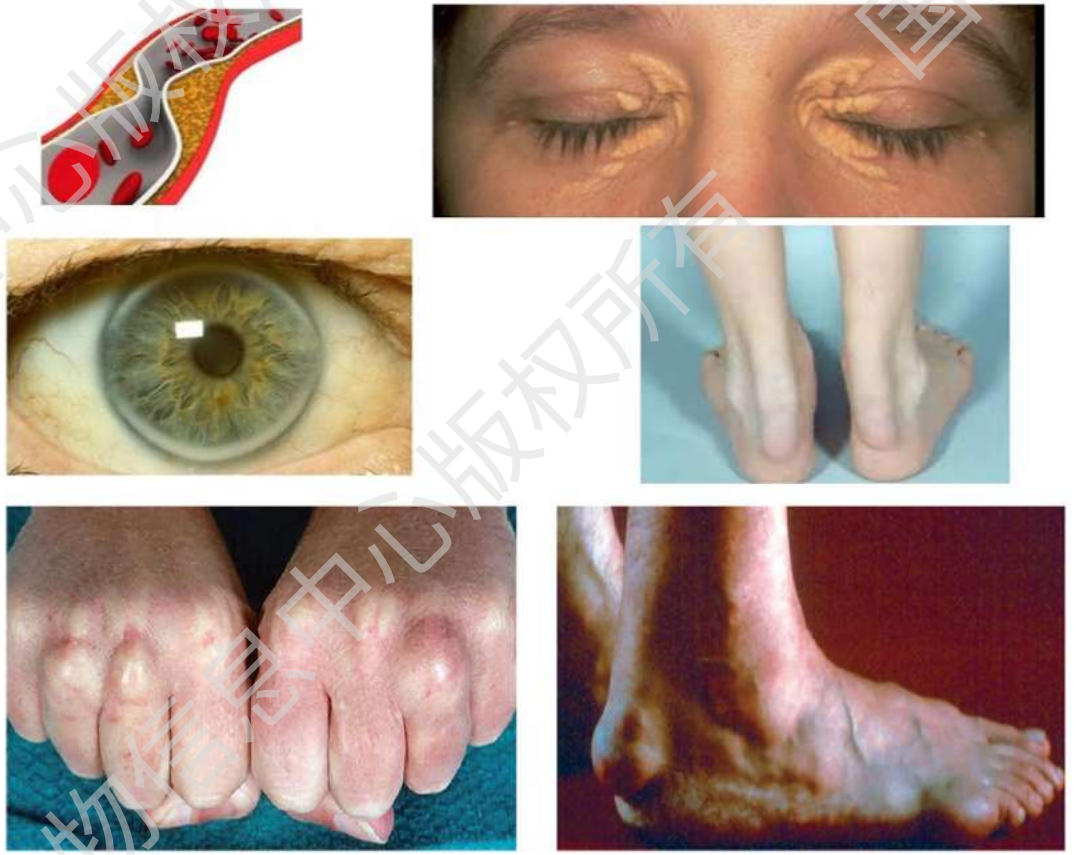
Gene variants mimic drug effects

Drug safety assessment by molecular phenotypes



- **Volanesorsen: FDA-approved treatment for familial chylomicronemia syndrome**
  - **Antisense oligonucleotide drugs**
  - **The target is APOC3 gene, which reduces triglyceride**
  - **Phase 3 study:**
    45% with acute thrombocytopenia

Familial Chylomicronemia Syndrome
triglyceride abnormality



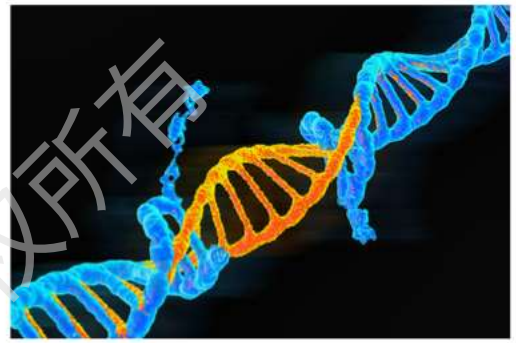
Drug safety assessment by molecular phenotypes

UK Biobank whole exome sequencing of 42K individuals

Gene mutation mimics drug intervention

➢ **Gene mutation:**

clinical trial by nature

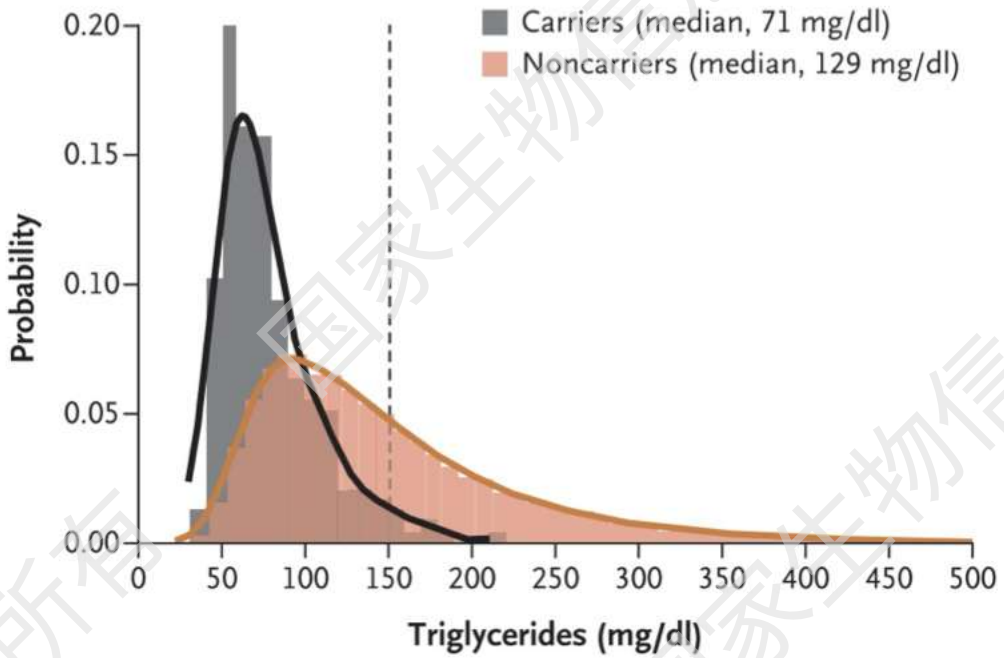


UK Biobank
10 years follow-up

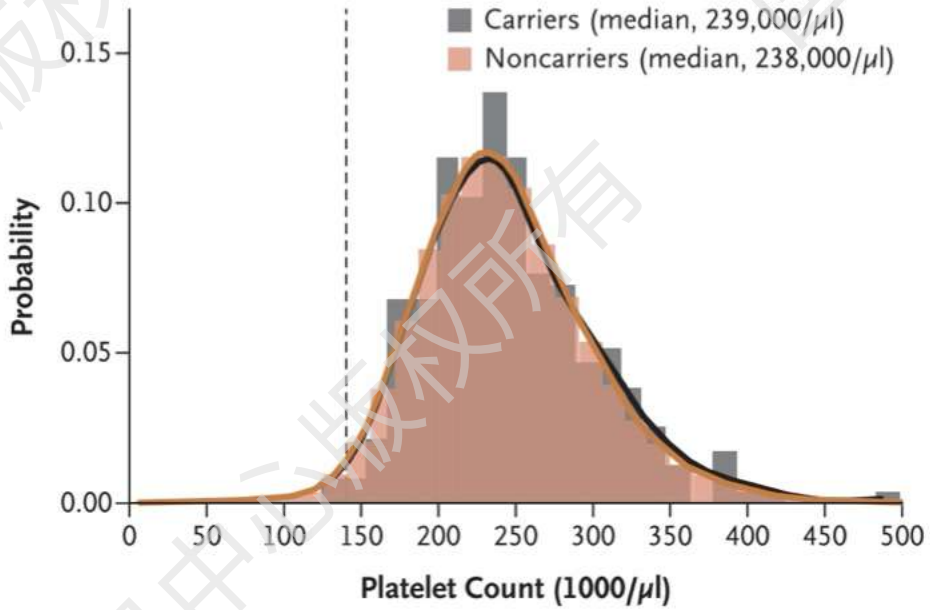
New England Journal of Medicine, 2019

Drug safety assessment by molecular phenotypes

APOC3 genetic mutations lower triglyceride levels



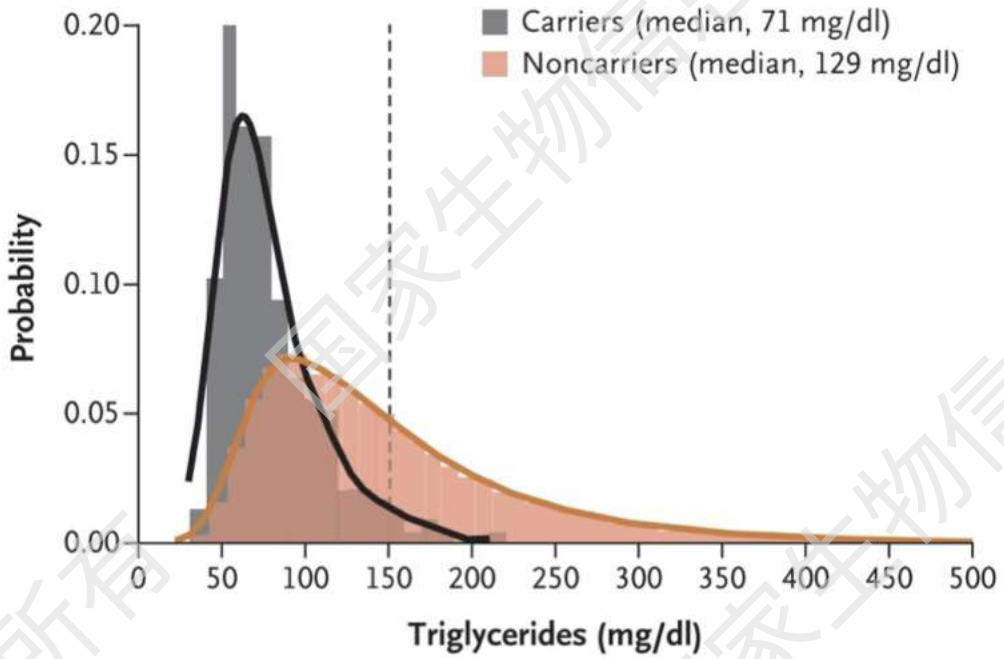
APOC3 genetic mutations and platelet count levels



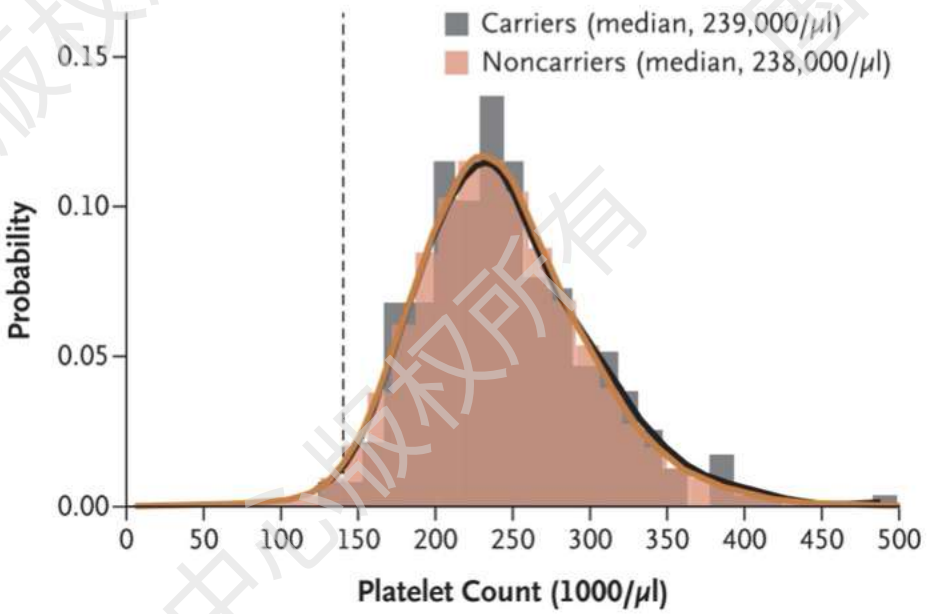
New England Journal of Medicine, 2019

Drug safety assessment by molecular phenotypes

APOC3 genetic mutations lower triglyceride levels



APOC3 genetic mutations and platelet count levels



The side effects of Volanesorsen is from the drug delivery system rather than the target gene

Target gene is safe!

Drug safety assessment by molecular phenotypes

The NEW ENGLAND JOURNAL of MEDICINE

1. Witztum JL, Gaudet D, Freedman SD, et al. Volanesorsen and triglyceride levels in familial chylomicronemia syndrome. N Engl J Med 2019;381:531-42.
2. The TG and HDL Working Group of the Exome Sequencing Project, National Heart, Lung, and Blood Institute. Loss-of-function mutations in *APOC3*, triglycerides, and coronary disease. N Engl J Med 2014;371:22-31.

DOI: 10.1056/NEJMc1912350

effect and early intervention to mitigate and maintain patients on treatment. We are currently investigating a galactosamine-conjugated *APOC3* oligonucleotide that, in a phase 1–2a study conducted over the course of 3 months, achieved similar levels of triglyceride reduction at lower doses, with no reported changes in platelet counts.[3]

Reply from drug developers:
Phase 1-2a, modified delivery system with the same target had no side effects

The same approach is applicable to other genetic drug development

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Olezarsen for Hypertriglyceridemia in Patients at High Cardiovascular Risk

Phase 2b :
Olezarsen -- Targeting APOC3 messenger RNA (ASO), found no safety concerns

Prohaska et al, NEJM 2024.

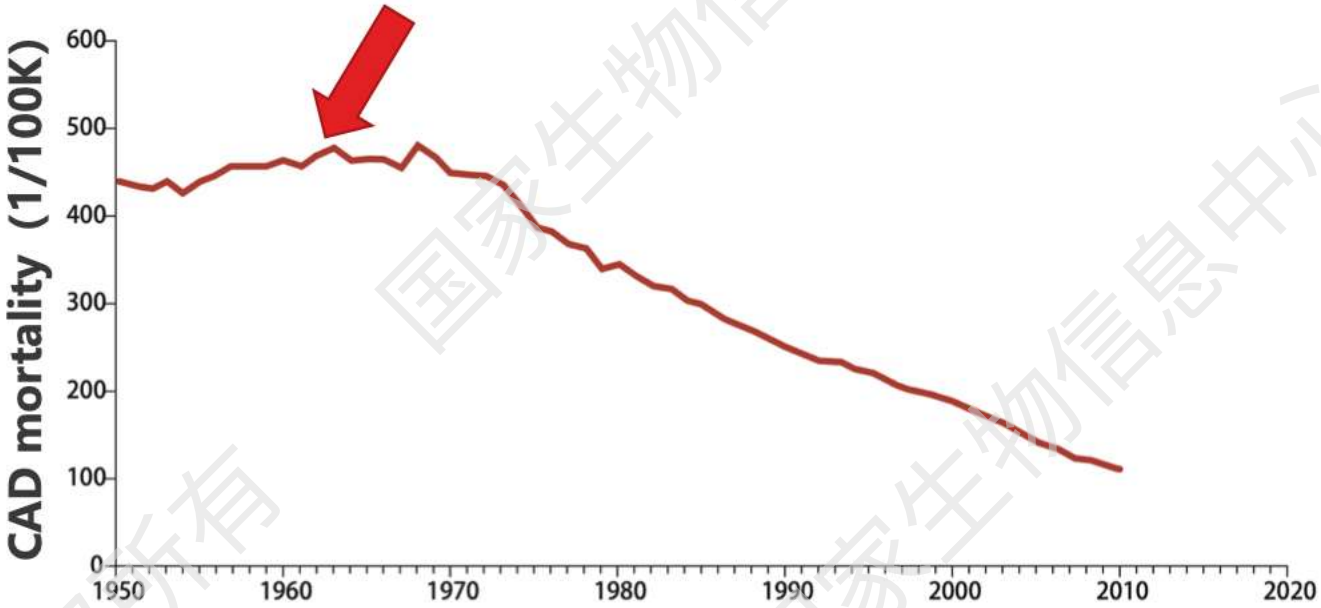
For a longer, healthier life

American Heart Association®
Healthy for Good™

Life's Essential 8™

80% of Cardiovascular diseases are preventable

Discovery of CAD risk factors in 1960s



CAD mortality in the United States

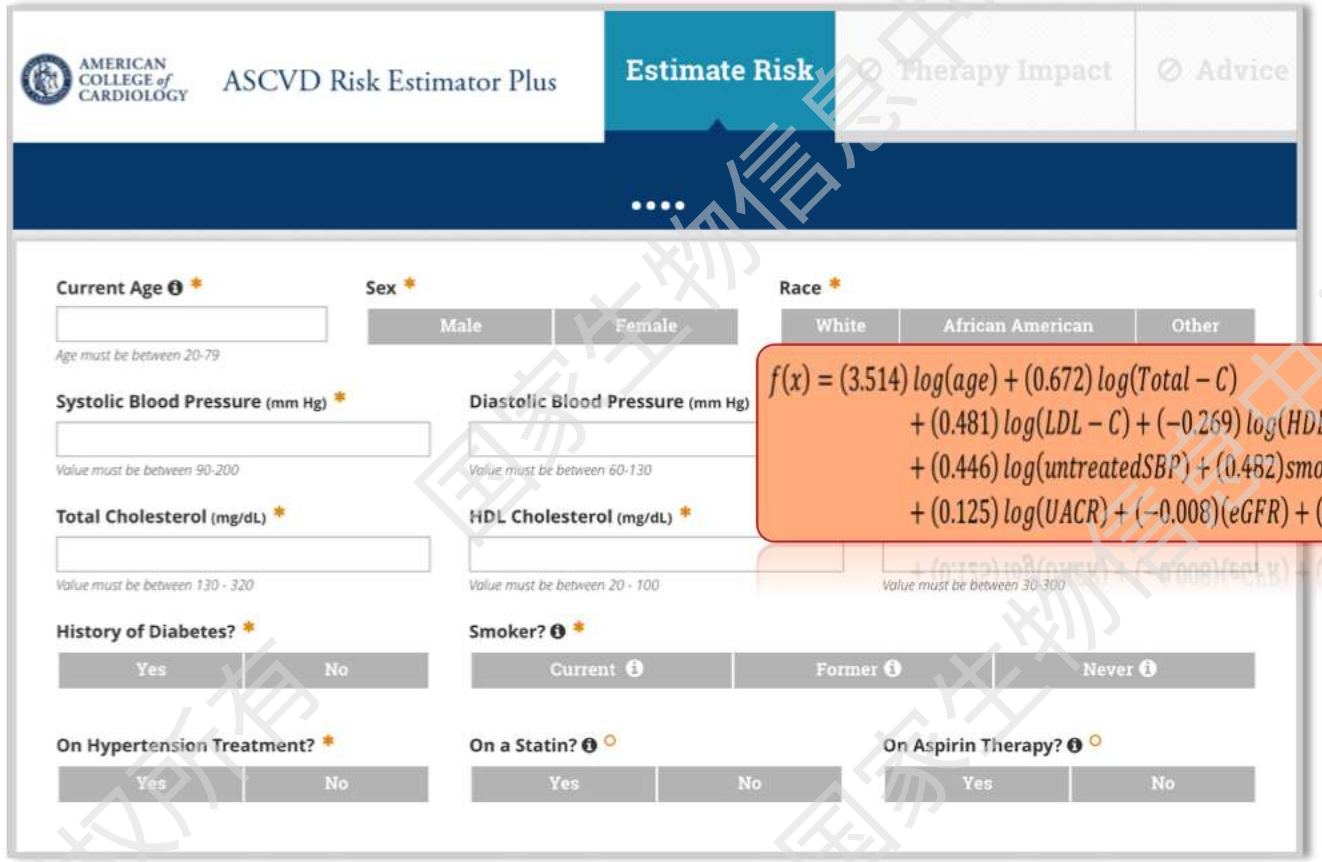
Nabel, E. G. & Braunwald, E. ***N. Engl. J. Med.*** *(2012).*
Mensah, G. A. et al. ***Circulation Research*** *(2017).*

The NEW ENGLAND JOURNAL of MEDICINE

200 NEJM ANNIVERSARY ARTICLE

The discovery of CAD risk factors and implementation of preventive medicine have led to remarkable achievements in disease prevention!

Preventive Medicine Based on Probabilistic Models



$$f(x) = (3.514)\log(age) + (0.672)\log(Total - C) + (0.481)\log(LDL - C) + (-0.269)\log(HDL - C) + (0.478)\log(treatedSBP) + (0.446)\log(untreatedSBP) + (0.482)smoking + 9.046(DM) + (0.125)\log(UACR) + (-0.008)(eGFR) + (-2.118)(\log(age) * DM)$$

https://tools.acc.org/ascvd-risk-estimator-plus/#!/calculate/estimate/

ACC/AHA Clinical Guidelines : PCE 10-Year ASCVD risk > 7.5% Recommended Drug Interventions

10-year risk for ASCVD
Low-risk (<5%)
Borderline risk (5% to 7.4%)
* Intermediate risk (7.5% to 19.9%)
* High risk (≥20%)

Preiss, D. & Kristensen, S. L. Can. J. Cardiol. (2015). 2013 ACC/AHA Guideline on the Assessment of Cardiovascular Risk.

Huangdi Neijing: "Superior doctors prevent illness before it occurs."

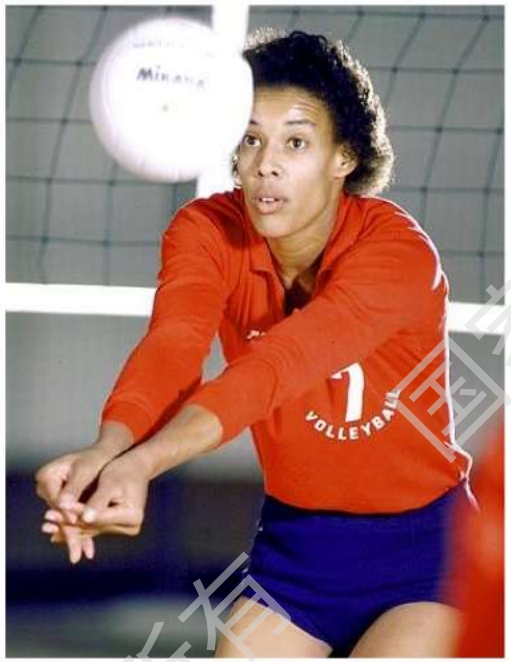
ACC/AHA PCE: American College of Cardiology/American Heart Association Pooled Cohort Equations

New concept of disease precision prevention

Early identification of individuals at high risk of disease, **especially before the onset of disease**, through targeted lifestyle interventions and cholesterol-lowering medications can reduce lifetime disease risk!

- **How can we effectively identify high-risk individuals?**

Monogenic Genetic Diagnosis and Disease Risk Stratification



Olympic Volleyball Star
Flo Hyman
Marfan syndrome
FBN1 mutation + Sudden Cardiac Death

2013 Angelina Jolie
BRCA1 + Family history
Mother died from ovarian cancer at age 56

汪敏先等. Rare Genetic Variants Associated With Sudden Cardiac Death in Adults.
Journal of the American College of Cardiology, 2019.

Monogenic Genetic Diagnosis and Disease Risk Stratification

Genetics in Medicine (2022) 24, 1407–1414

ELSEVIER

Genetics in Medicine

www.journals.elsevier.com/genetics-in-medicine

ACMG STATEMENT

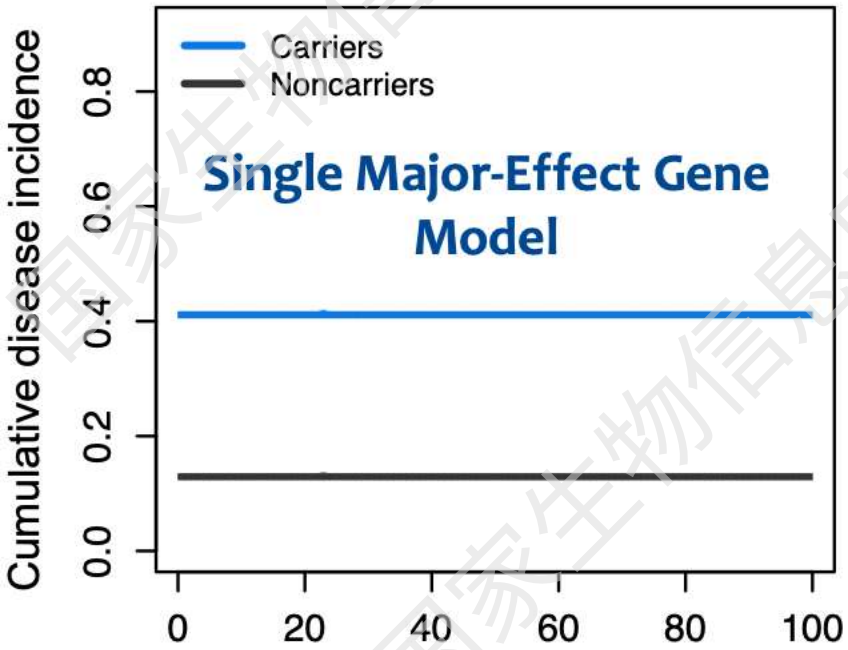
ACMG SF v3.1 list for reporting of secondary findings in clinical exome and genome sequencing: A policy statement of the American College of Medical Genetics and Genomics (ACMG)

Miller DT, et al. *Genet Med*. 2023;25(8)

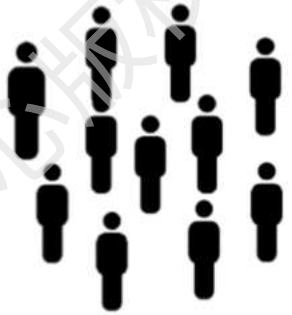
➢ ***81 clinically actionable gene list***

Phenotype	Gene
…	…
Familial hypercholesterolemia	*LDLR*
	APOB
	PCSK9
Marfan syndrome, Loeys-Dietz syndromes, and familial thoracic aortic aneurysms and dissections	*FBNI1*
	TGFBR1
	TGFBR2
	SMAD3
	ACTA2
	MYH11
…	…

Monogenic Genetic Diagnosis and Disease Risk Stratification

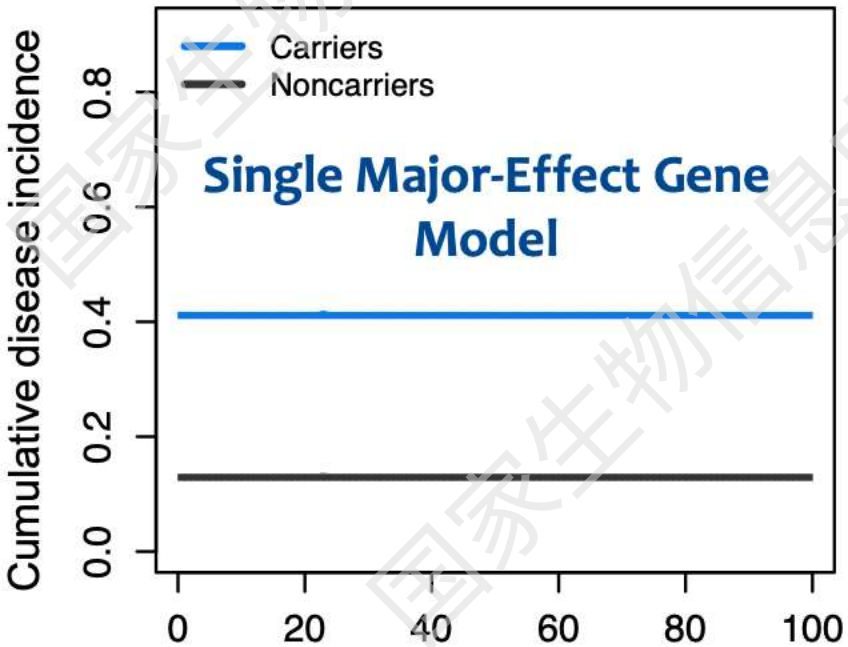


High risk Carriers



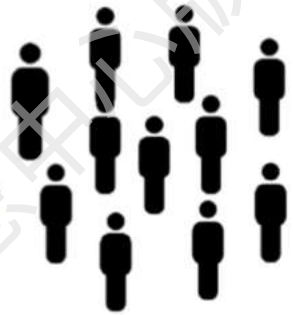
Average risk Noncarriers

Monogenic Genetic Diagnosis and Disease Risk Stratification



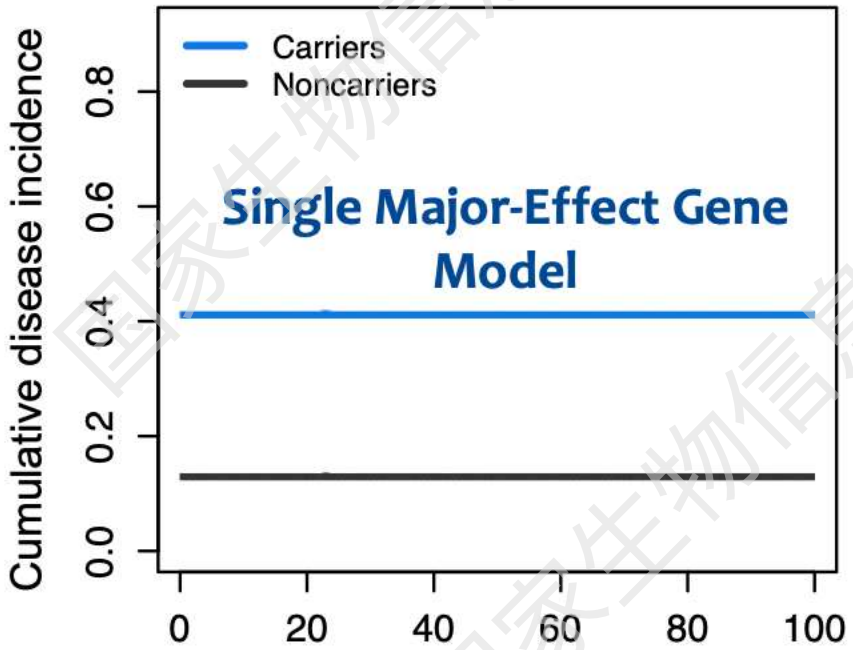
High risk Carriers

- Population Explained
- Incomplete Penetrance



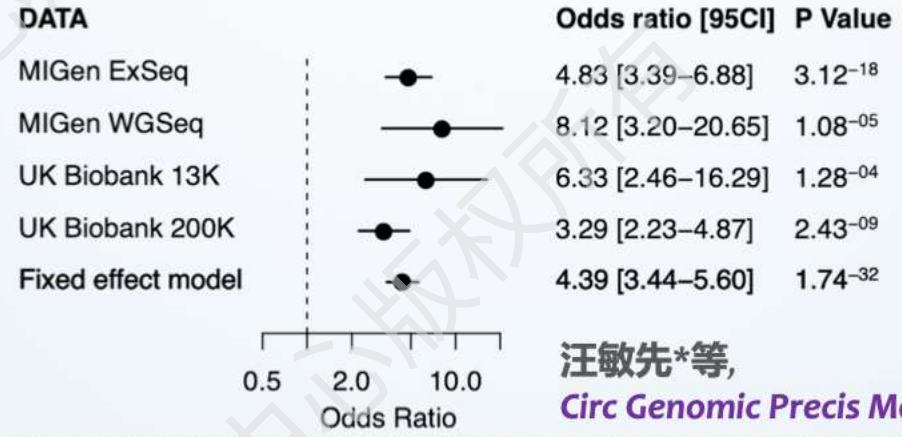
Average risk Noncarriers

Monogenic diagnosis has limited utility in complex diseases—— CAD

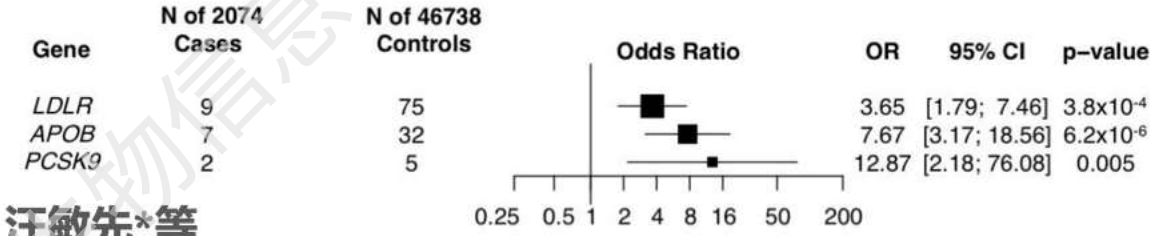


LDLR | 41k CAD cases and 220k controls

- Odds ratio 4.39
- P = 1.74 x 10^{-32}
- 0.91% cases vs 0.34% controls



汇敏先*等,
Circ Genomic Precis Med. 2022



汇敏先*等
Nature Communications. 2020

Monogenic diagnosis has limited utility in complex diseases —— AF

Gene	Case Carrier	Control Carrier	Percentage	Odds Ratio	95%CI	P-value	Total
TTN	571	4149	**1.1344%**	**1.66**	1.52-1.81	1.09E-26	4720
RPL3L	343	2722	**0.7366%**	**1.57**	1.39-1.77	1.36E-12	3065
PKP2	75	485	**0.1346%**	**1.76**	1.45-2.14	2.17E-07	560
C10orf71	45	305	**0.0841%**	**1.96**	1.41-2.74	1.95E-04	350
CTNNA3	41	193	**0.0562%**	**2.51**	1.74-3.62	4.71E-06	234
KDM5B	34	196	**0.0553%**	**2.58**	1.75-3.80	1.08E-05	230
MYBPC3	30	175	**0.0493%**	**2.75**	1.81-4.17	1.40E-05	205
RBM20	25	147	**0.0413%**	**2.18**	1.39-3.44	1.66E-03	172
PPA2	15	115	**0.0312%**	**2**	1.03-3.88	3.96E-02	130
PDLIM3	15	95	**0.0264%**	**2.22**	1.09-4.52	2.80E-02	110
DES	16	85	**0.0243%**	**2.99**	1.70-5.26	5.55E-04	101
FLNC	14	79	**0.0224%**	**2.71**	1.25-5.87	1.13E-02	93
PRKAG2	11	63	**0.0178%**	**2.72**	1.13-6.54	2.51E-02	74
ACTN2	12	62	**0.0178%**	**2.67**	1.37-5.19	7.48E-03	74
GATA5	9	40	**0.0118%**	**2.32**	1.29-4.17	1.00E-02	49
ACTC1	6	33	**0.0094%**	**3.6**	1.04-12.49	4.36E-02	39
GATA6	6	18	**0.0058%**	**4.8**	1.81-12.74	4.40E-03	24
TGFB3	5	11	**0.0038%**	**7.13**	2.13-23.87	3.32E-03	16
WWTR1	3	11	**0.0034%**	**3.5**	1.19-10.27	9.65E-03	14
NKX2-5	4	8	**0.0029%**	**7.5**	1.91-29.44	6.01E-03	12
CDH2	3	6	**0.0022%**	**10.23**	2.26-46.22	5.31E-03	9

UK Biobank, Total N = 41 6085, AF = 30 797

Rufan Zhang … Minxian Wang#, Akl Fahed#, ***JAMA cardiology.*** *2025 (IF = 15)*

Polygenic Risk Score (PRS)

10 BREAKTHROUGH TECHNOLOGIES 2018

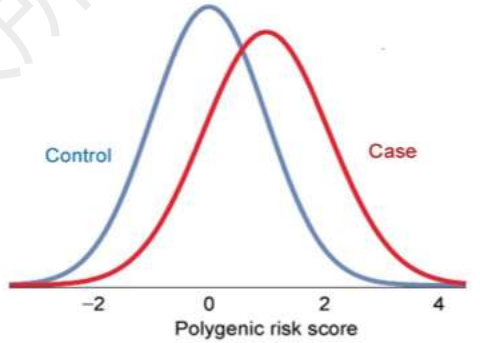
MIT Technology Review

Genetic Fortune-Telling

"polygenic risk score" (PRS)

Next-gen individual health prediction

Method innovations	25k	100k	1M	5M
PRIMED consortium	emerge network: Electronic Medical Records and Genomics (eMERGE) Network; NIH National Human Genome Research Institute	Wisdom: Women Informed to Screen Depending on Measures of Risk (Wisdom Study) (WISDOM)	All of Us RESEARCH PROGRAM	UK Research and Innovation: Accelerating detection of disease (ADD) challenge program
 NIH awards $38 million to improve utility of polygenic risk scores in diverse populations	• $75million for 5 Years, • **RRS for 10 conditions,** • return results to 25,000 participants.	• **ClinicalTrials.gov Identifier: NCT02620852,**	• improve risk prediction, early detection and early intervention, • **By 2023 to 2024, return PRS for 5 million participants.**	
• $38million for 5 Years, • Consortium Members				



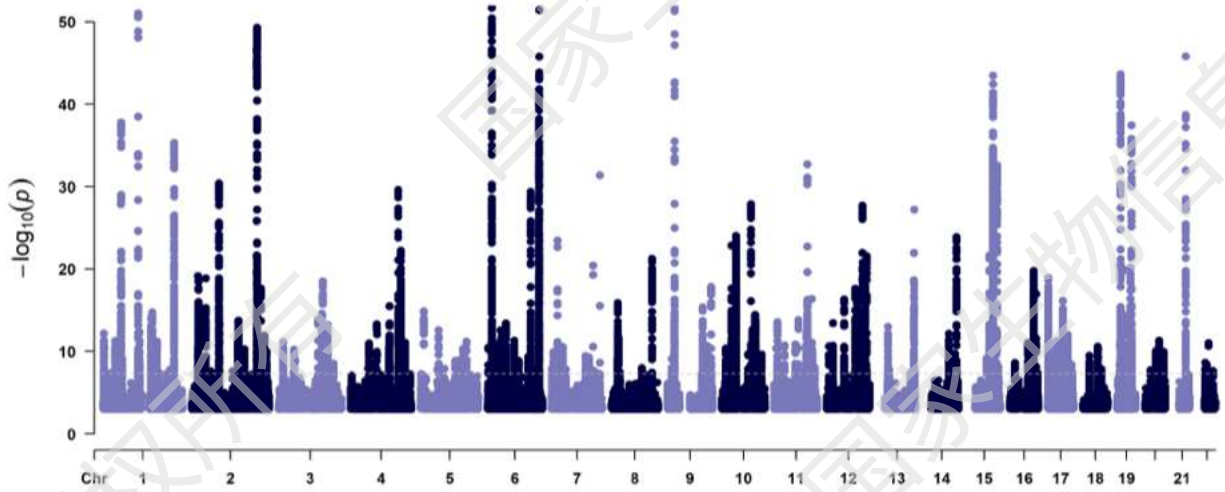
$$PRS_i = \sum_j^M \hat{\beta}_j \times dosage_{ij}$$

Early warning across common complex diseases , "Measure once, benefit for life"

Genetic Maps for CAD and AF

CARDIoGRAMplusC4D

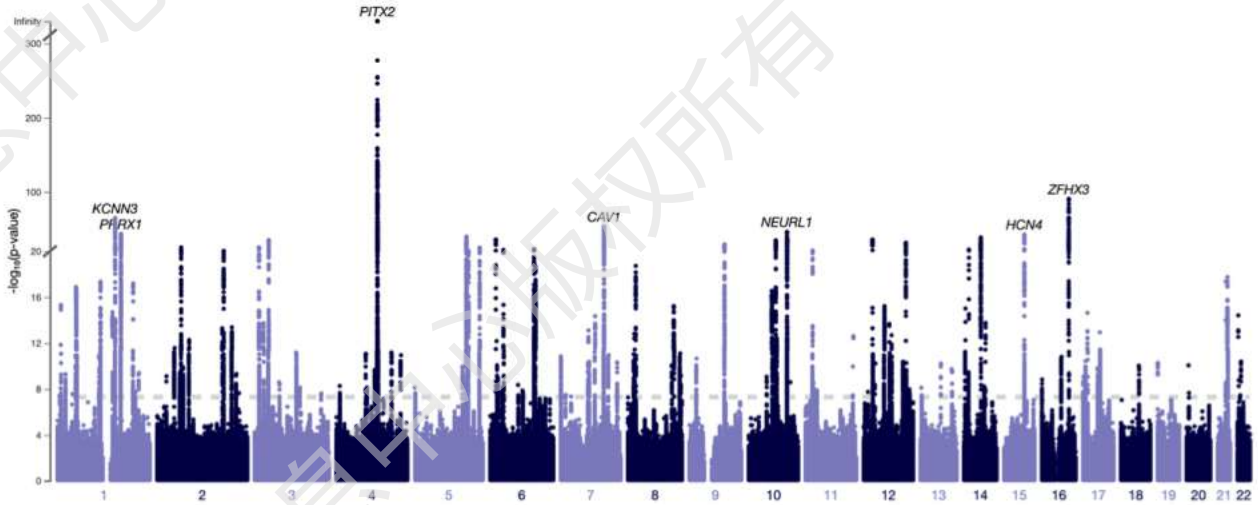
180k CAD cases + 1,170k controls
300 independent genetic loci identified by GWAS



Aragam KG, et al., ***Nat Genet***. 2022

MVP

85k AF cases + 540k controls
>300 independent genetic loci identified by GWAS



Anurag V., et al., ***Science***. 2024

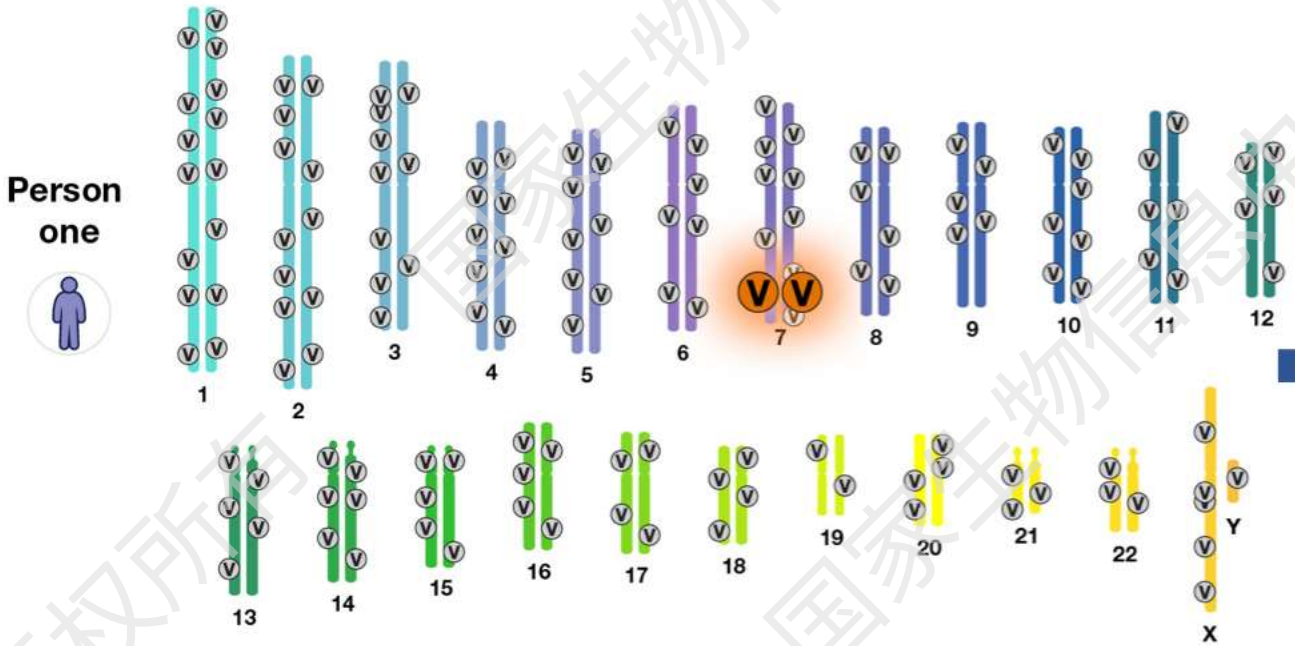
Polygenic Risk Score (PRS)

CARDIoGRAMplusC4D

CAD GWAS

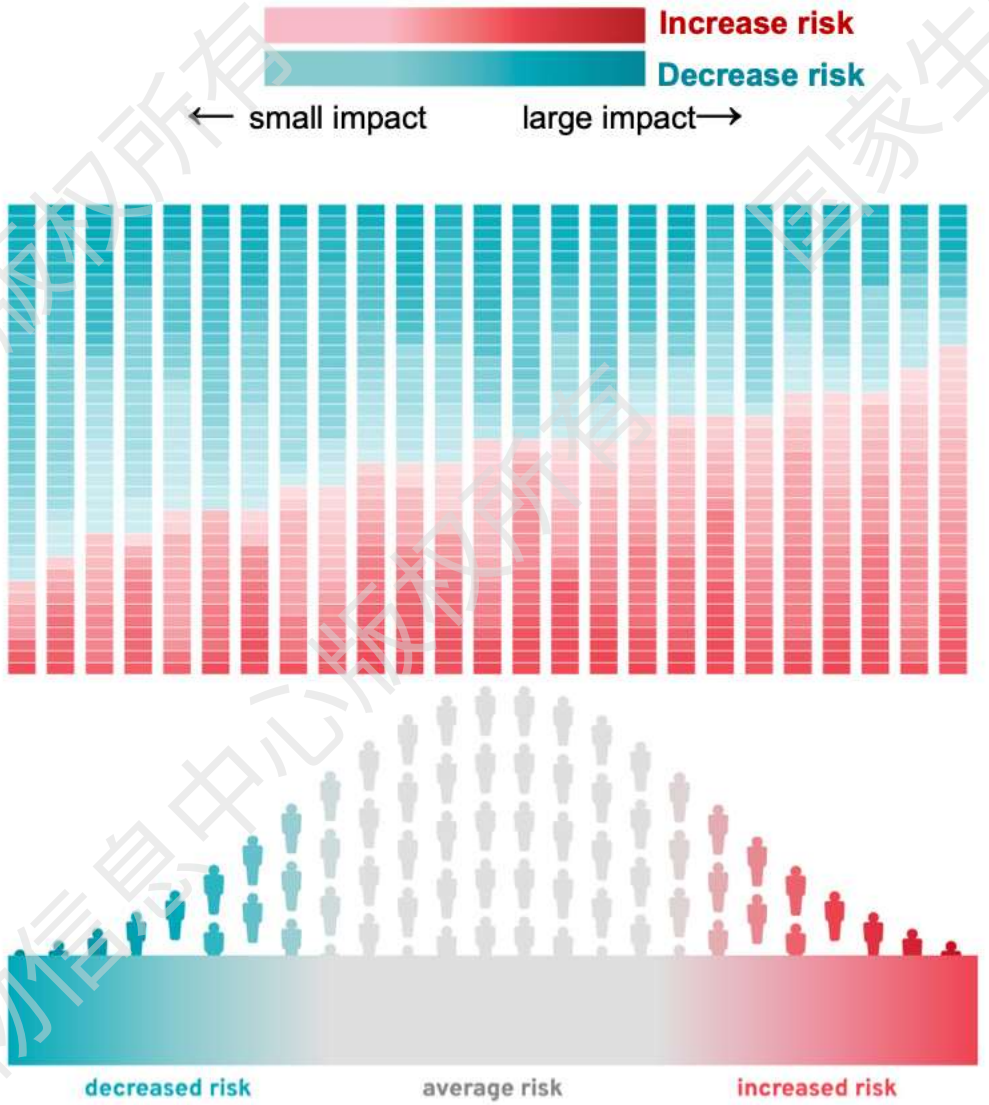
180K CAD cases + 1.2M controls

identified 300 associated variants



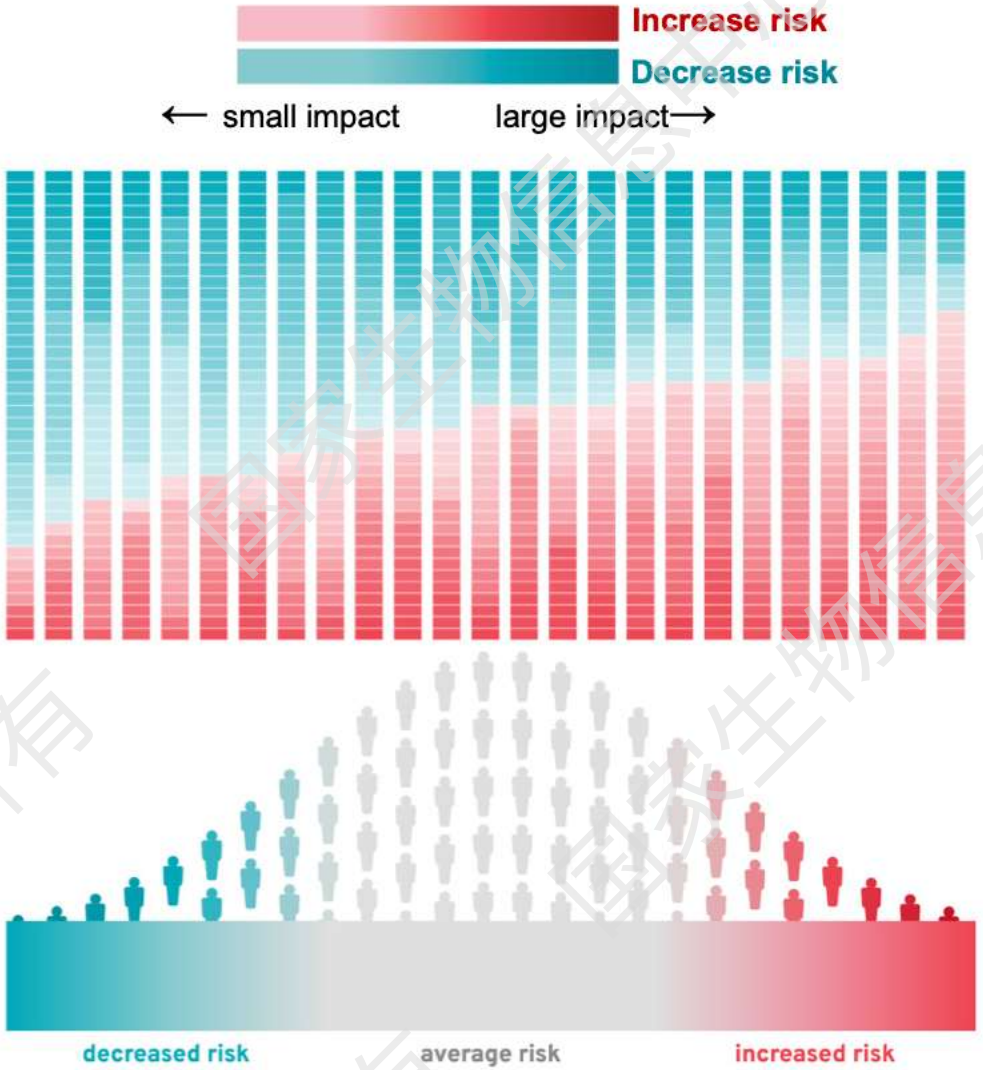
Letter 'V' for disease-associated mutations.

Rare large-effect mutations and common variants with small effects.

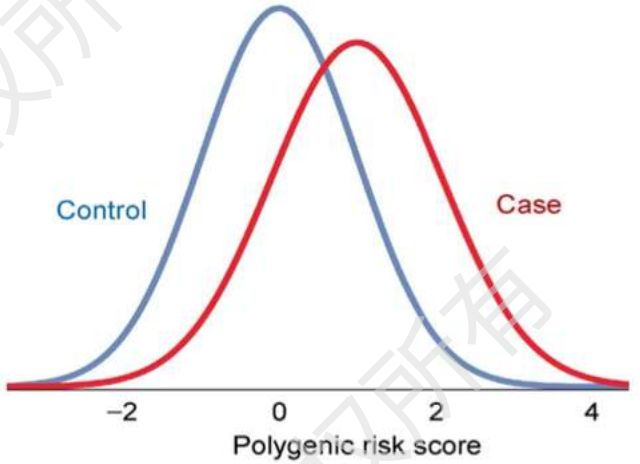


http://polygenicscores.org/

Polygenic Risk Score (PRS)



http://polygenicscores.org/

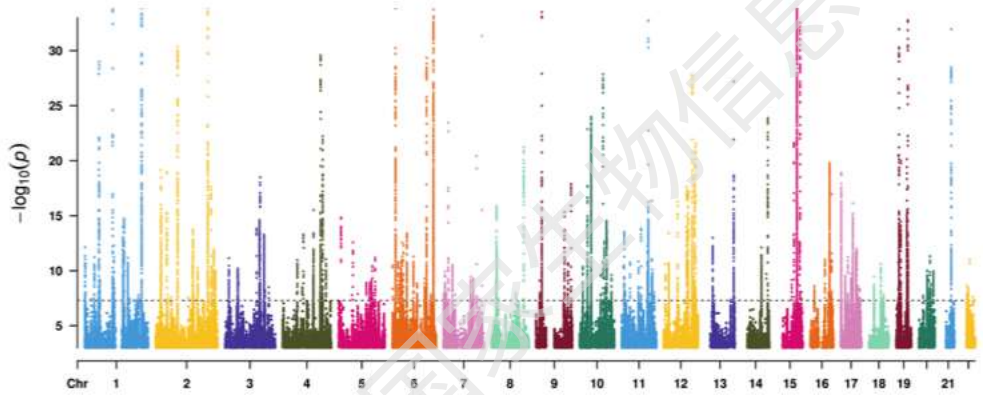


$$PRS_i = \sum_j^M \hat{\beta}_j \times dosage_{ij}$$

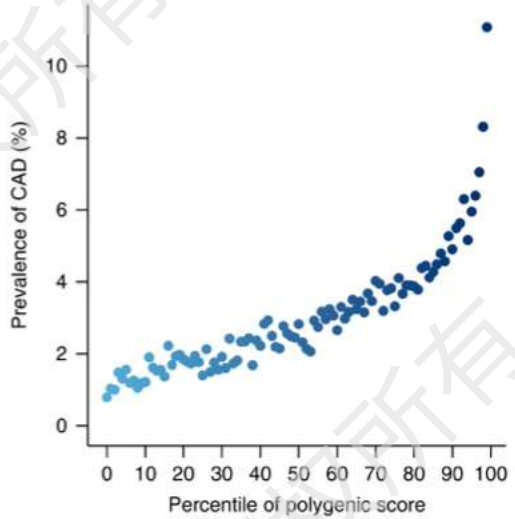
Effect size **#Effective alleles**

Polygenic Risk Score (PRS)

CAD GWAS



CAD risk



Khera A V., et al. *Nat Genet.* 2018

$$PRS_i = \sum_j^M \hat{\beta}_j \times dosage_{ij}$$

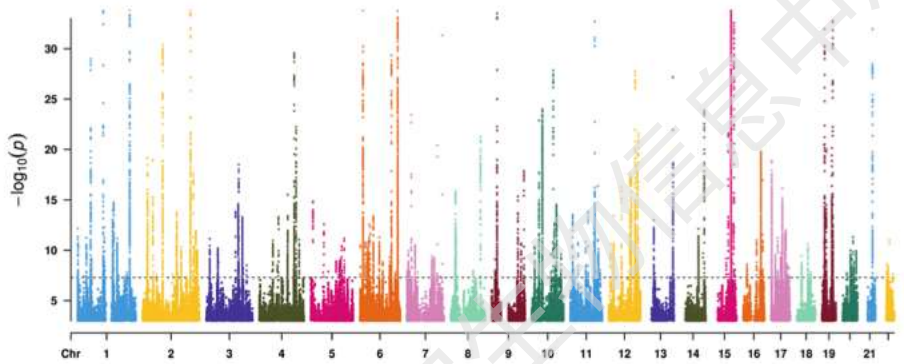
Effect size

No. of Effective alleles

Traditional methods
Predict CAD risk of by CAD GWAS weight

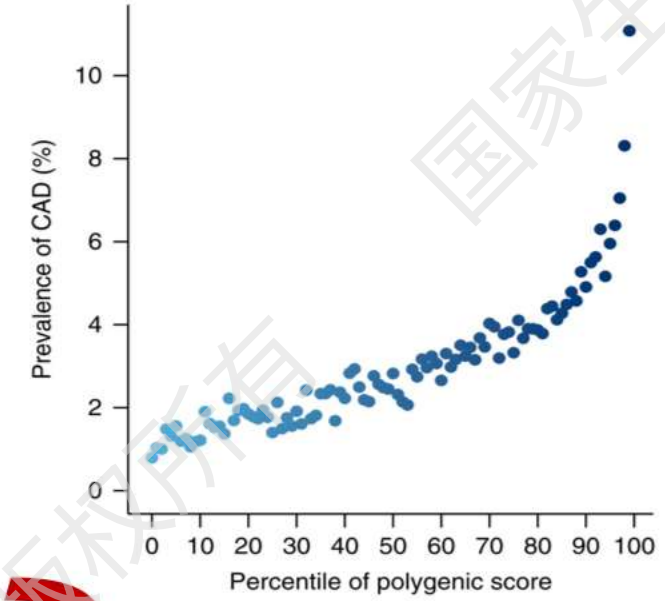
Polygenic risk model for CAD -- *GPS 2.0*

CAD GWAS



Risk factors and comorbidities

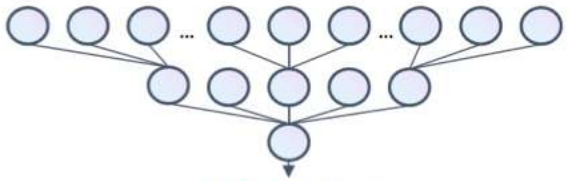
CAD risk



?

Polygenic risk model for CAD -- *GPS 2.0*

Two-layer neural network



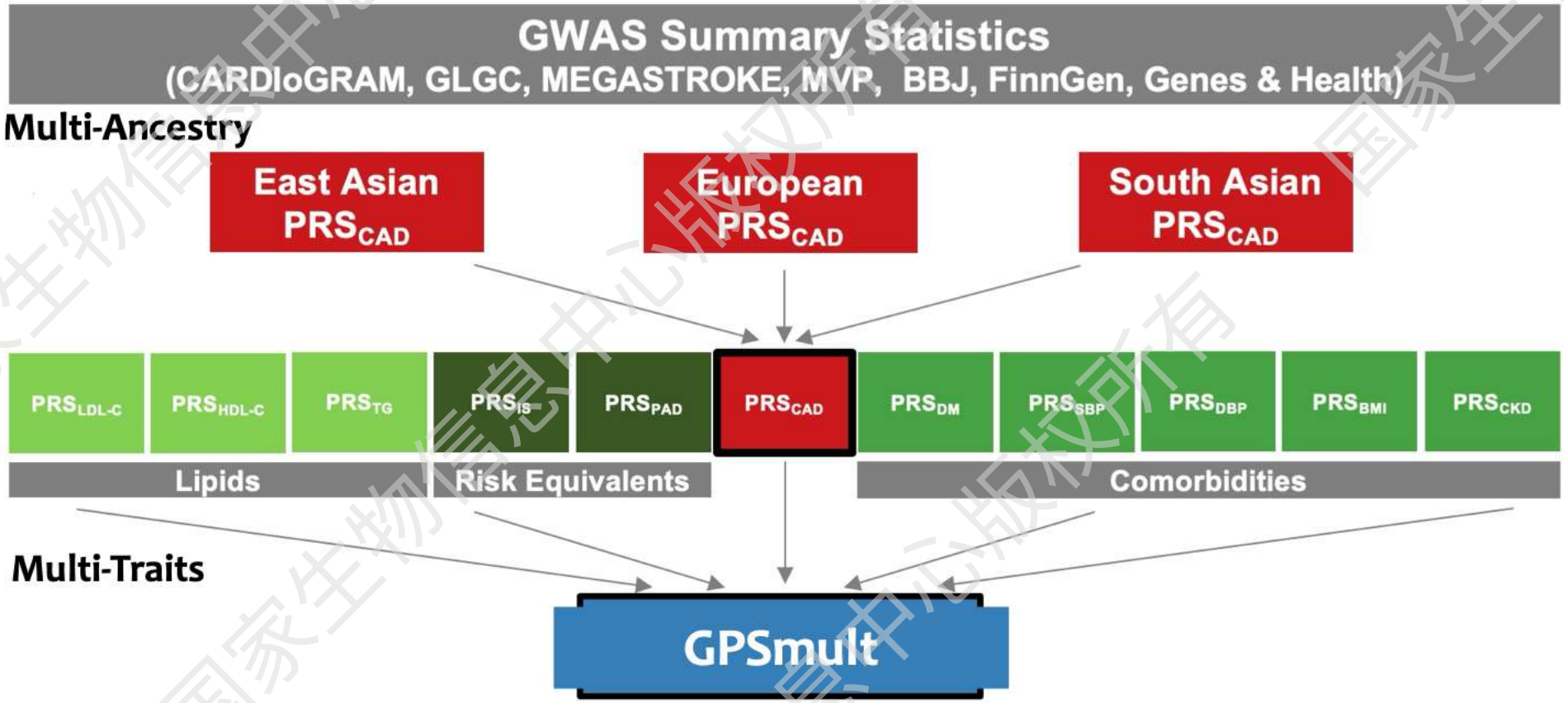
CAD status

BROAD INSTITUTE

MASSACHUSETTS GENERAL HOSPITAL

- **CardiogramPlusC4D**
  - **>260k CAD cases**
  - **>1170k controld**

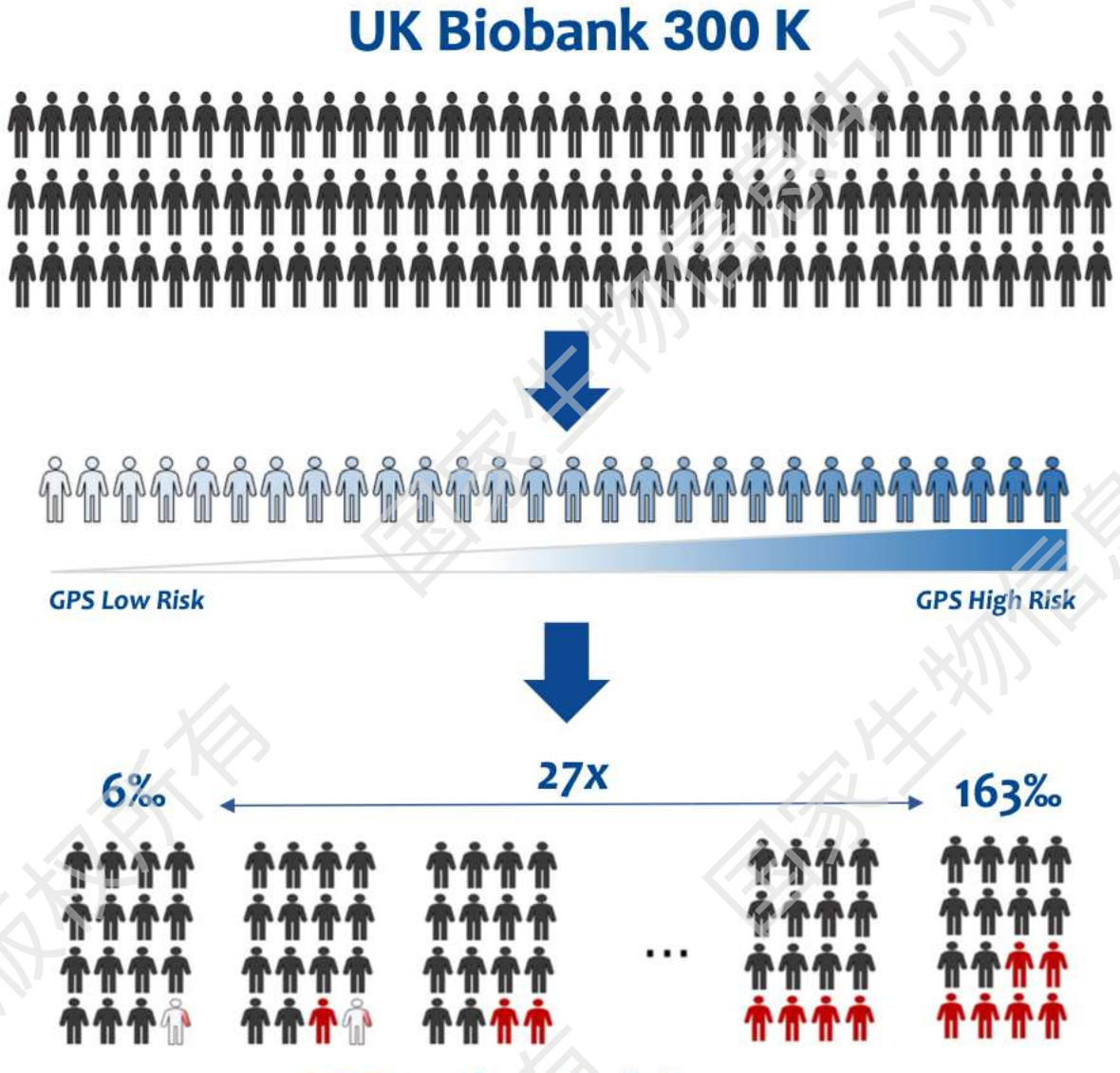
- **CAD comorbidities and risk factors**
  - **> 1.3M individuals**

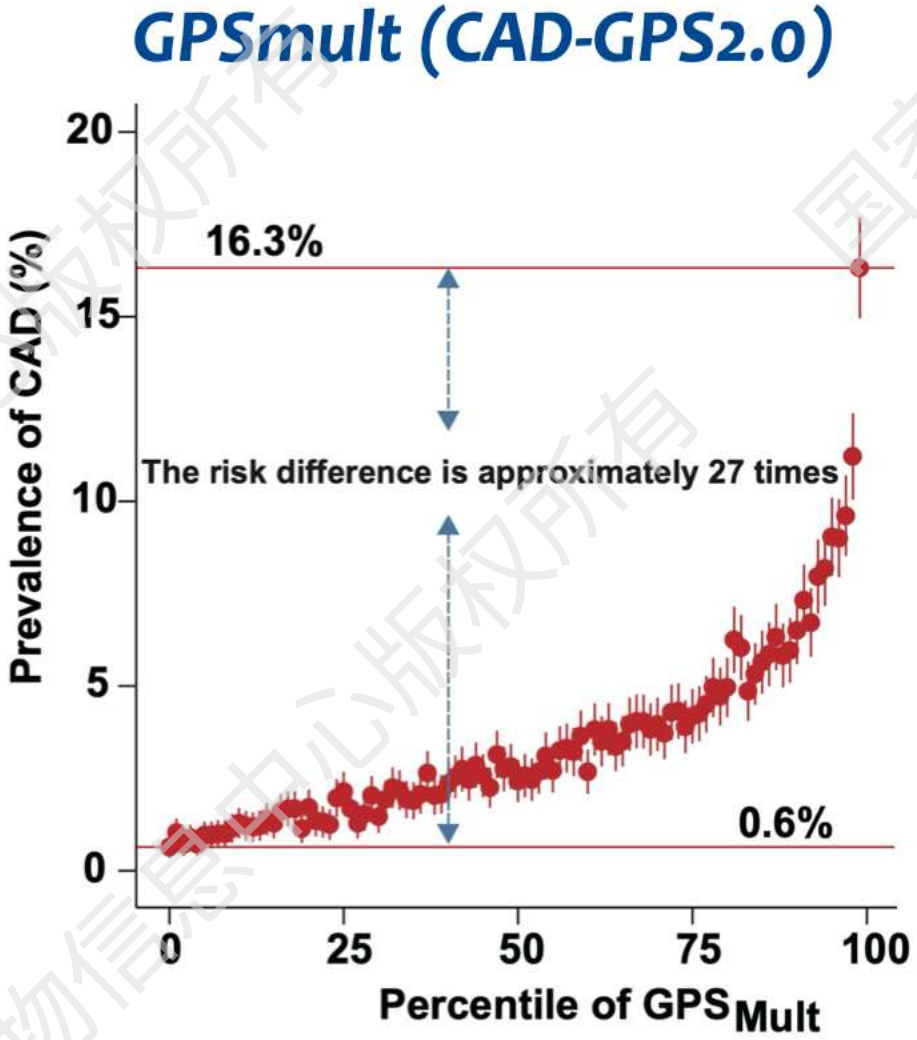


Integrate genetic risk from multi-ethnic and comorbidity GWASs

Nature Medicine. 2023

Polygenic risk model for CAD -- *GPS 2.0*



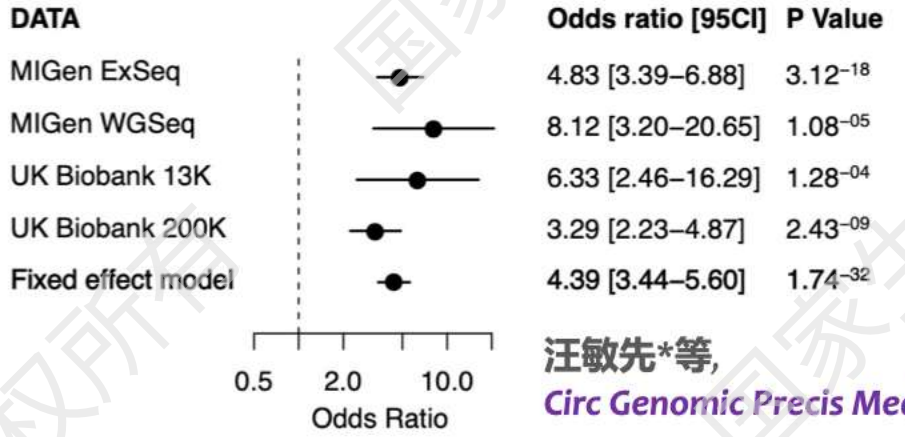


Nature Medicine. 2023

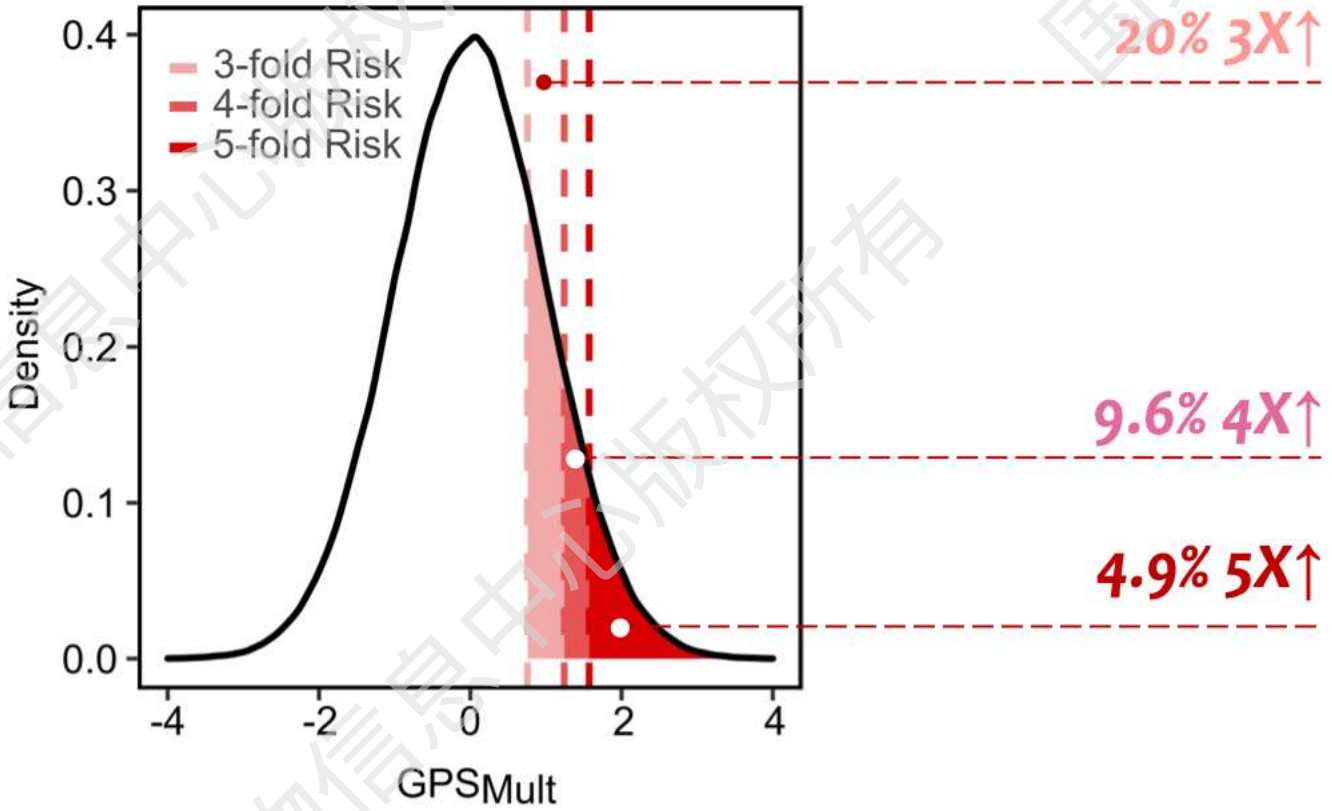
Polygenic Risk Scores and Disease Risk Stratification

LDLR | 41k CAD cases and 220k controls

- Odds ratio 4.39
- P = 1.74 x 10^{-32}
- 0.91% cases vs 0.34% controls

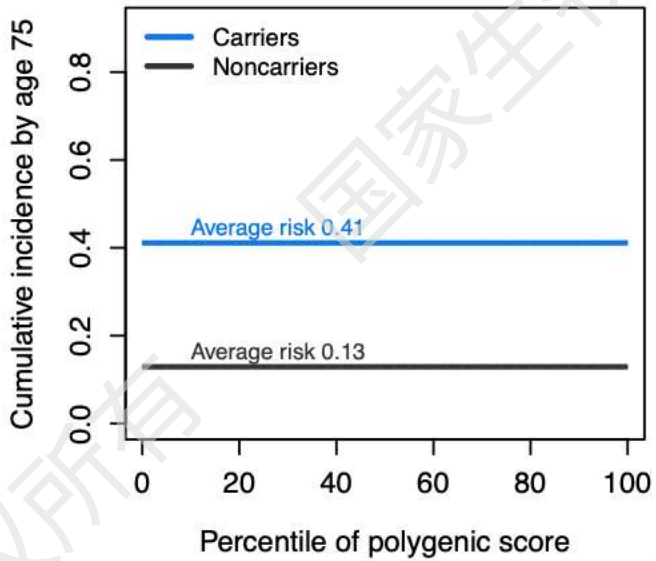


汪敏先*等,
Circ Genomic Precis Med. 2022



汪敏先*#等,
Nature Medicine. 2023

Personalized Genetic Risk Assessment V2.0

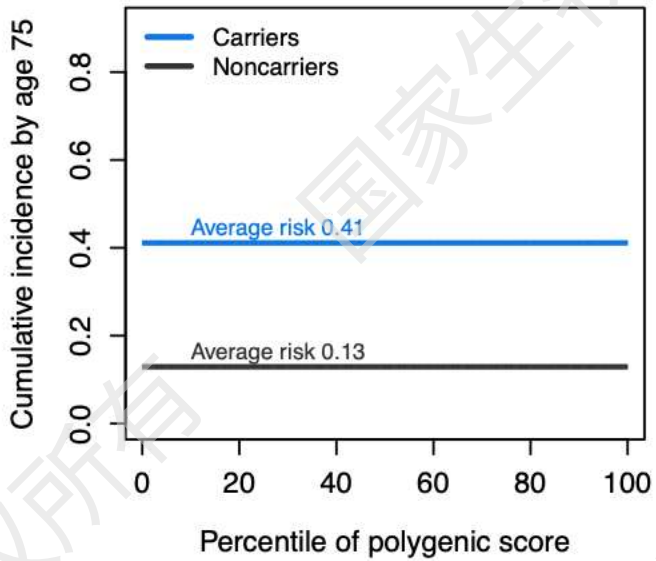


?
Monogenic mutation

High risk carriers

Average risk non-carriers

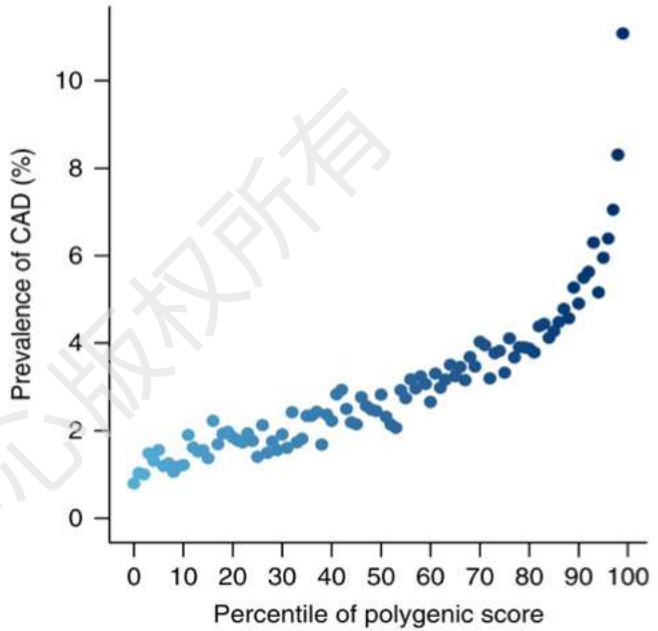
Personalized Genetic Risk Assessment V2.0



?

Monogenic mutation

Σ



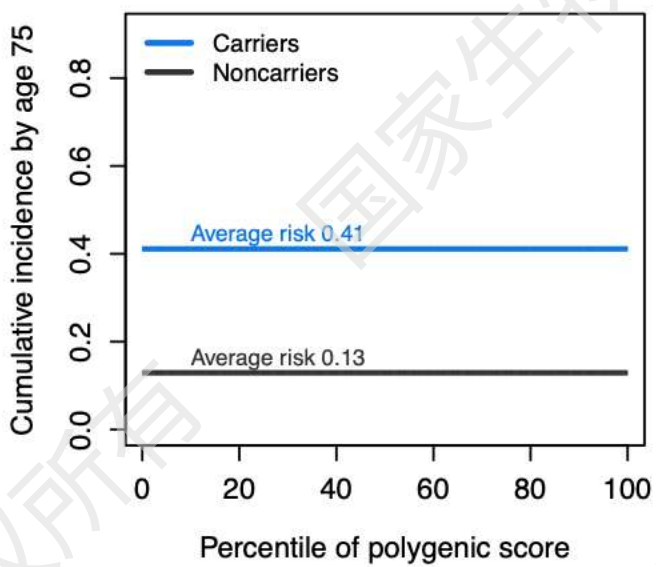
High risk carrier

Average risk non-carriers

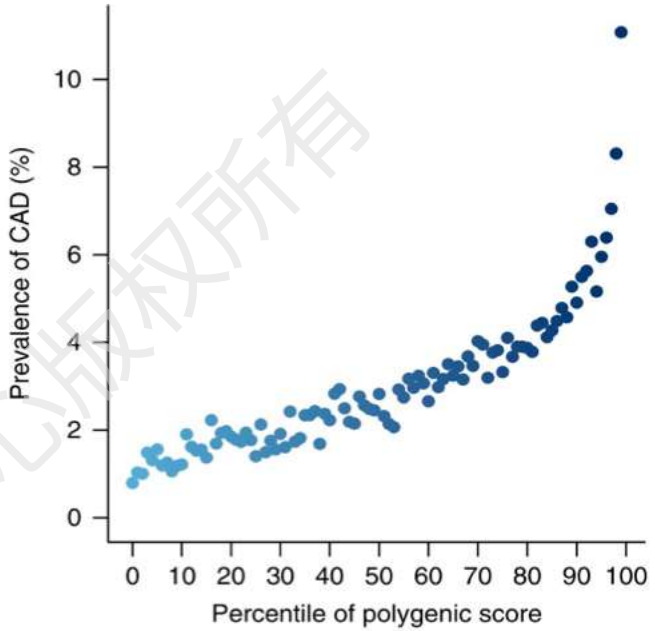
Polygenic risk summarize small effects

Personalized Genetic Risk Assessment V2.0

? Interplay of multiple effects?



? Monogenic mutation

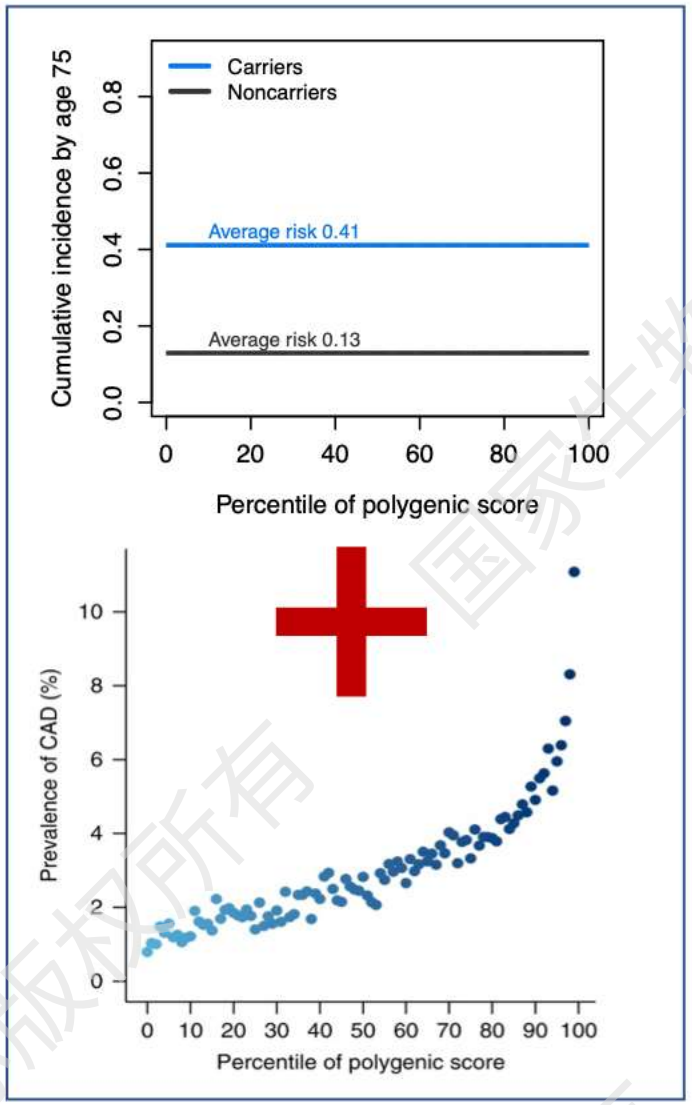


High risk carrier

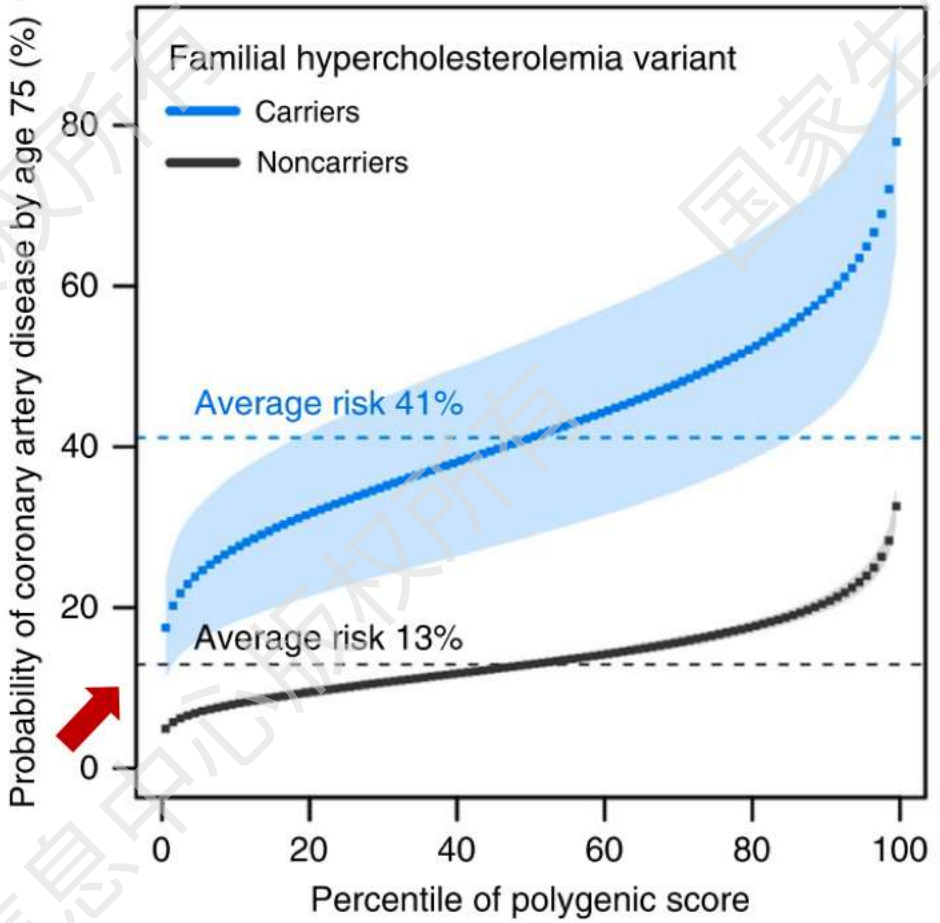
Average risk non-carriers

Polygenic risk small effects size accumulation

Personalized Genetic Risk Assessment For CAD



Joint modeling of monogenic variants and polygenic risk

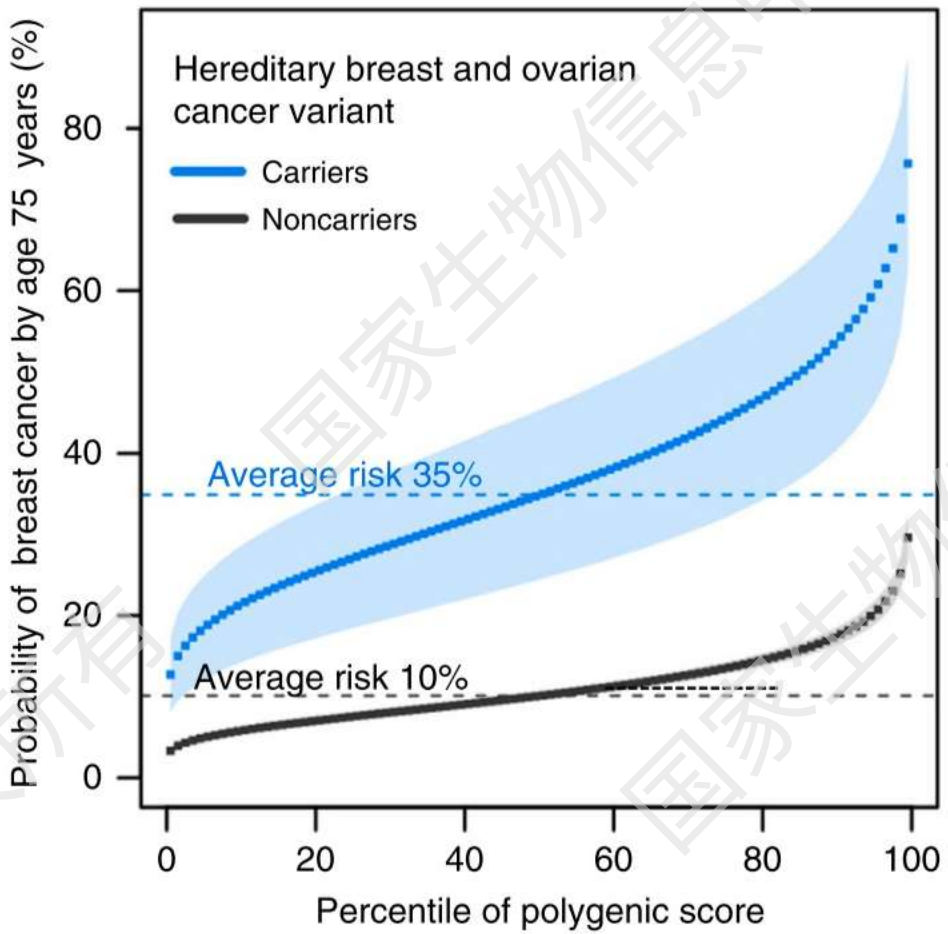


ACMG Variation	Avg	Low	High
+	41%	17%	78%
-	13%	4.9%	33%

Nature Communications. 2020

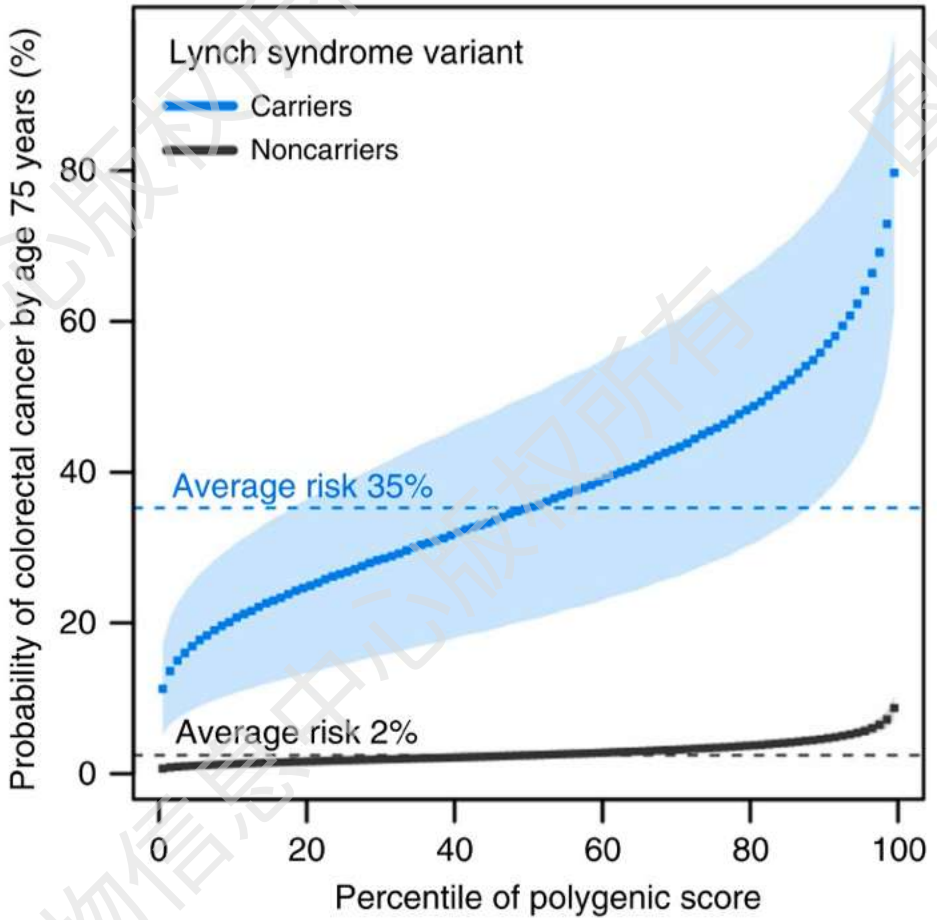
V2.0: Monogenic + Genome-Wide Background Susceptibility

Breast cancer absolute disease risk



564 Cases + 25,260 Controls, 8 Years follow up.

Colorectal cancer



361 Cases + 48,812 Controls, 8 Years follow up.

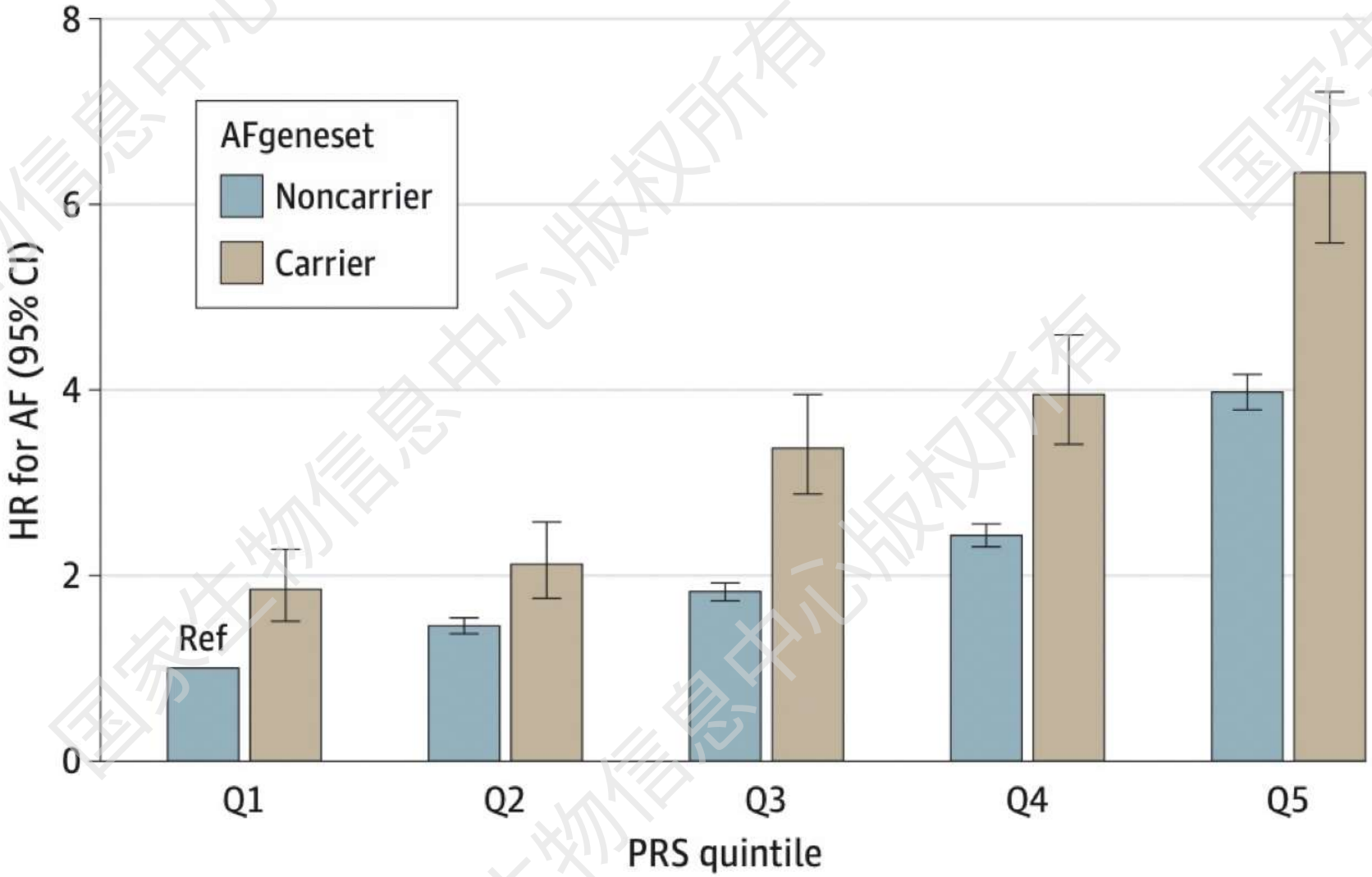
汪敏先*等, *Nature Communications.* 2020

Monogenic + Genome-Wide Background Susceptibility (AF)

biobank uk

AF = 30 797

N = 41 6085



Rufan Zhang ... Minxian Wang#, Akl Fahed#, *JAMA cardiology.* 2025

Personalized Genetic Risk Assessment V2.0

- **Highlighted by Science**
- **Top 10 clinical genetics advances of 2020 (U.S. NHGRI Advisory Council)**

YEAR IN REVIEW

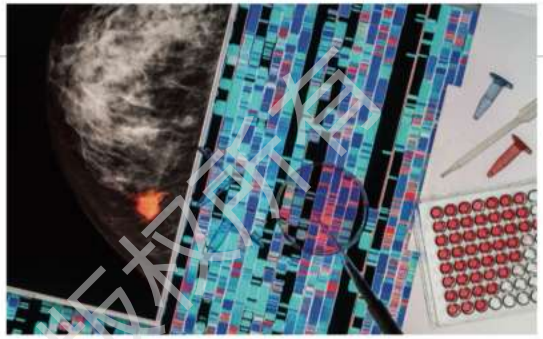
Genomic Medicine Year in Review: 2020

Teri A. Manolio,[1,*] Carol J. Bult,[2] Rex L. Chisholm,[3] Patricia A. Deverka,[4] Geoffrey S. Ginsburg,[5] Madison Goldrich,[1] Gail P. Jarvik,[6] George A. Mensah,[7] Mary V. Relling,[8] Dan M. Roden,[9] Robb Rowley,[1] Cecelia Tamburro,[1] Marc S. Williams,[10] and Eric D. Green[1]

Introduction

In December 2019, *The American Journal of Human Genetics* introduced a new annual feature[1] designed to chronicle ten key advances in applying genomic information to clinical care in the previous twelve months of published literature. The Genomic Medicine Working Group of the National Advisory Council for Human Genome Research of the National Human Genome Research Institute (NHGRI) and understandable trend toward fewer citations of those published later in the year.

This year the Working Group has again selected its ten most significant advances among the 45 recognized accomplishments and resources published during the 12 months ending August 31, 2020. These papers clustered into 13 broad categories, two of which—"Genomic Medicine Resource" and "Pilot Implementation"—were added for the 2020 review to genotype-guided treatment has not been studied. Two randomized trials compared outcomes of rapid genotype-guided versus standard antiplatelet therapy in patients undergoing percutaneous coronary intervention (PCI). Both trended in favor of genotyping but came to different stated conclusions based on prespecified statistical analysis plans. Neither trial was powered to study outcomes in carriers of two LOF alleles, the group unable to bioactivate clopidogrel at all.



A woman's genetic background can powerfully modify her cancer risk from a *BRCA1* mutation.

BIOMEDICINE

'Polygenic' analyses may sharpen disease risk predictions

Genetic background modifies impact of single disease genes

By **Jocelyn Kaiser**

Women who learn that they have a mutation in the breast cancer gene *BRCA1* face a wrenching decision. Their doctor or genetic counselor will likely tell them that women with such mutations have, on average, a 72% lifetime risk of breast cancer and a 44% risk of ovarian cancer. Given that, up to half decide to have prophylactic mastectomies, and many have ovaries removed, too.

But recent studies show a woman could receive a more individualized, accurate cancer risk estimate by factoring in other gene variants. A preprint posted last month finds that a person's "polygenic" background influences not only the disease risk conferred by a *BRCA1* defect, but also risks from single gene mutations linked to colorectal cancer and heart disease. Some individuals were very likely to develop cancer or heart disease by age 75, the analysis showed, whereas in others the risk was not much greater than in a person without the high-risk mutation.

"It's pretty striking," says cardiologist and geneticist Amit Khera of Massachusetts General Hospital (MGH) in Boston, leader of the study, which is on the medRxiv preprint server. "It's become clear that there are both monogenic and polygenic [disease] drivers. The future is to assess both."

"The message is a very important one for patients and clinicians," says Teri Manolio of the National Human Genome Research Institute in Bethesda, Maryland. "Carriers of *BRCA1* mutations or other pathogenic variants don't invariably develop disease, and genomics can be used to help parse carriers who are at lower risk." Others caution, however, that risk scores summing how dozens to thousands of other genetic variants interact with a single major disease gene aren't yet accurate enough to use in the clinic. The new paper "is teasing at the possibility, but there's a lot of work to be done," says Harvard University epidemiologist Peter Kraft.

MGH cardiology fellow Akl Fahed and others in Khera's group explored polygenic influences on the three important single-gene disorders in the United States: familial hypercholesterolemia, which leads to sky-high cholesterol levels and dramatically elevates risk of heart disease; Lynch syndrome, a flaw in DNA repair that brings a lifetime risk of colorectal cancer of about 60%; and inherited breast cancer, caused by variants in *BRCA1* or *BRCA2*. They took advantage of databases that combine medical and genomic information from thousands of people, enabling researchers to tally how the many genetic variants with subtle effects modify disease risks and complex traits such as height.

Drawing on some 50,000 participants in the UK Biobank and 19,000 women tested for *BRCA* genes by the company Color Genomics, the team found that polygenic background strongly modified the risk of carrying a mutation in the key genes for the three disorders. For a small proportion of major disease gene carriers, other genetic variants boosted their overall risk of cancer or heart disease to about 80%, well above the average of 30% to 40% that Khera's group estimated for its study populations based on just the single disease gene mutations. (The team's monogenic disease risk predictions are lower than many other estimates for several possible reasons, Khera notes, including that the UK Biobank participants are healthier than the general population.) At the other extreme, the polygenic analysis suggested that a few people with those mutations have much lower risks than predicted by their single mutation alone, as low as 11% for colon cancer, 13% for breast cancer, and 17% for heart disease—not much higher than other people in general.

Khera's group says adding polygenic data to single-gene tests could help people decide whether to take aggressive steps to head off disease—mastectomy or removal of the ovaries for women carrying *BRCA* mutations or frequent colonoscopies for people with Lynch syndrome. But the new study does not include enough data for clinical decisions, says genetic epidemiologist Antonis Antoniou of the University of Cambridge in the United Kingdom. Only 116 women in the UK Biobank sample had *BRCA* mutations, which he notes "is an extremely small number to make inferences about risks."

Two years ago, Antoniou led a study that reported on how polygenic scores influence risks in 25,000 carriers of *BRCA* mutations and found nearly as wide a range of overall cancer risks. His team has incorporated those data into a breast cancer risk estimator along with factors such as family history.

The MGH study is "an important and exciting paper" that complements other work, says David Ledbetter, chief scientific officer for the Geisinger Health System in Danville, Pennsylvania. His team recently looked at 92,000 participants in an ongoing genomic medicine study called MyCode, focusing on those who carried mutations predisposing them to 11 rare disorders that affect traits such as height, weight, and cholesterol levels. Incorporating polygenic scores helped predict those traits, the group reported on 25 October in *Nature Communications*.

It may be a while before physicians are comfortable telling patients how genetic backgrounds modify the risk posed by a major disease gene mutation. But some companies already offer polygenic scores for cancer and other diseases, and tests that combine both kinds of information are imminent. Before insurance companies agree to pay for such tests, Ledbetter cautions, "They're going to want to see much more clinical validation"—including for minorities, because current polygenic analyses draw on data primarily from people of European ancestry. ■

PHOTO: MOSAIC/SCIENCE SOURCE

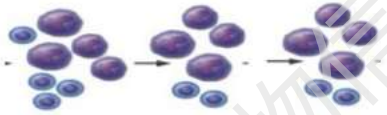
Published by AAAS

Genetic Risk Factors for Cardiovascular Disease

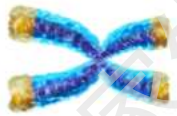
Germline Mutations

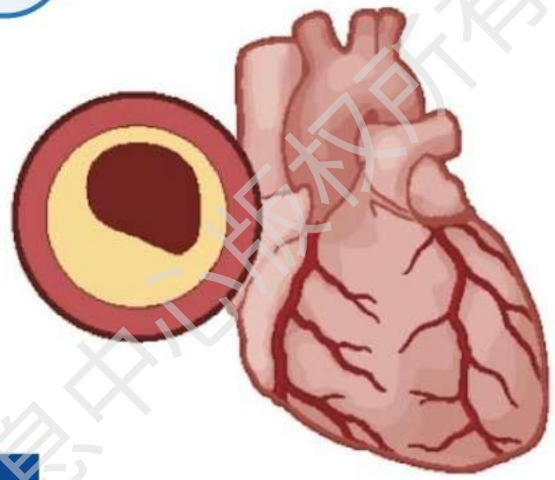
- Monogenic → *LDLR, PCKS9 …*
- Polygenic →→→→→→→→→→ *Polygenic risk*

Somatic Mutations
CHIP



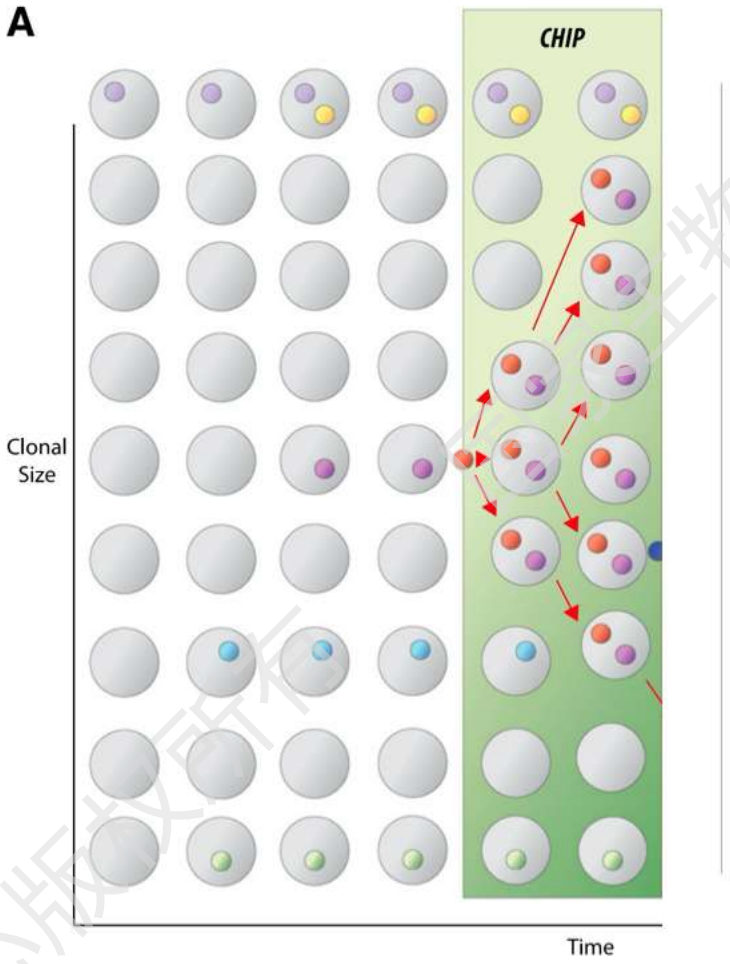
Telomere Length
LTL





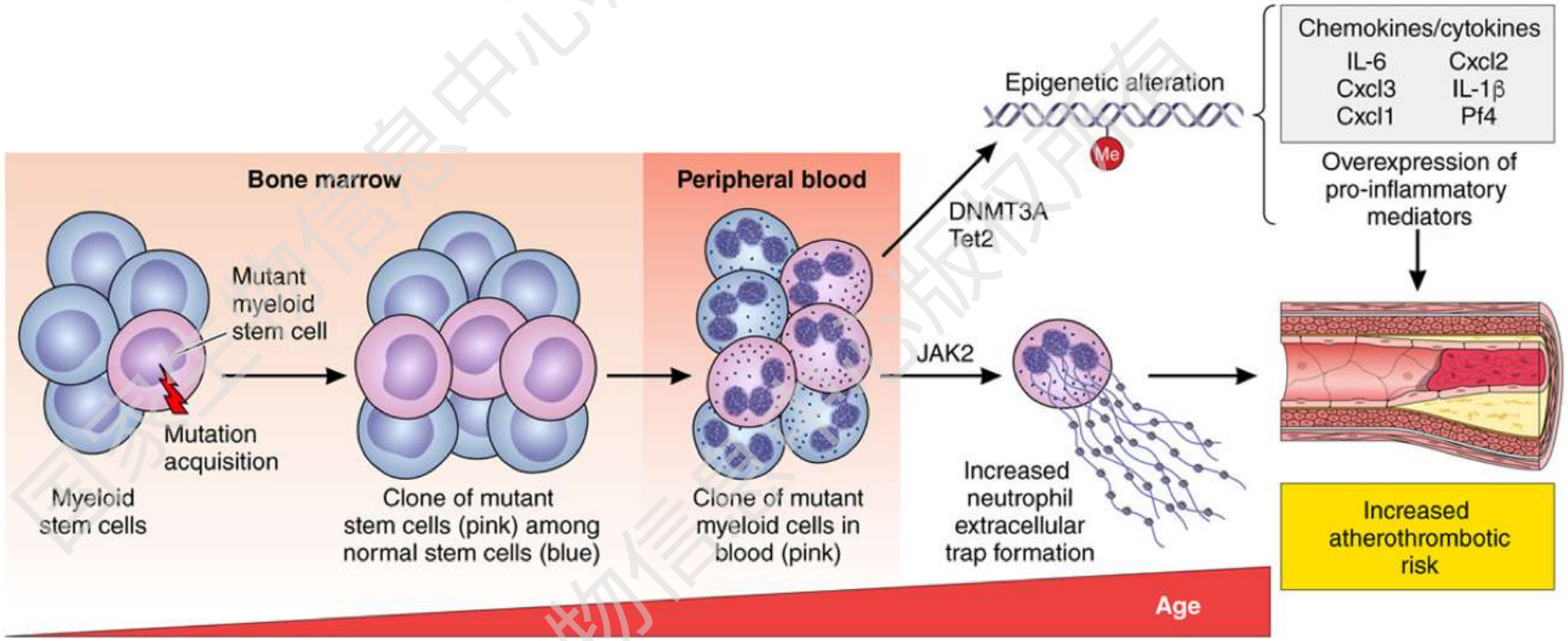
Genetic Factors

Clonal Hematopoietic Somatic Mutations and CAD risk



- **Clonal hematopoiesis of indeterminate potential （CHIP）**
  - > Age>65 ~ 10%-15%
  - > Age>80 ~ 30%

Clonal hematopoiesis of indeterminate potential (CHIP), also known as age-related clonal hematopoiesis (ARCH)

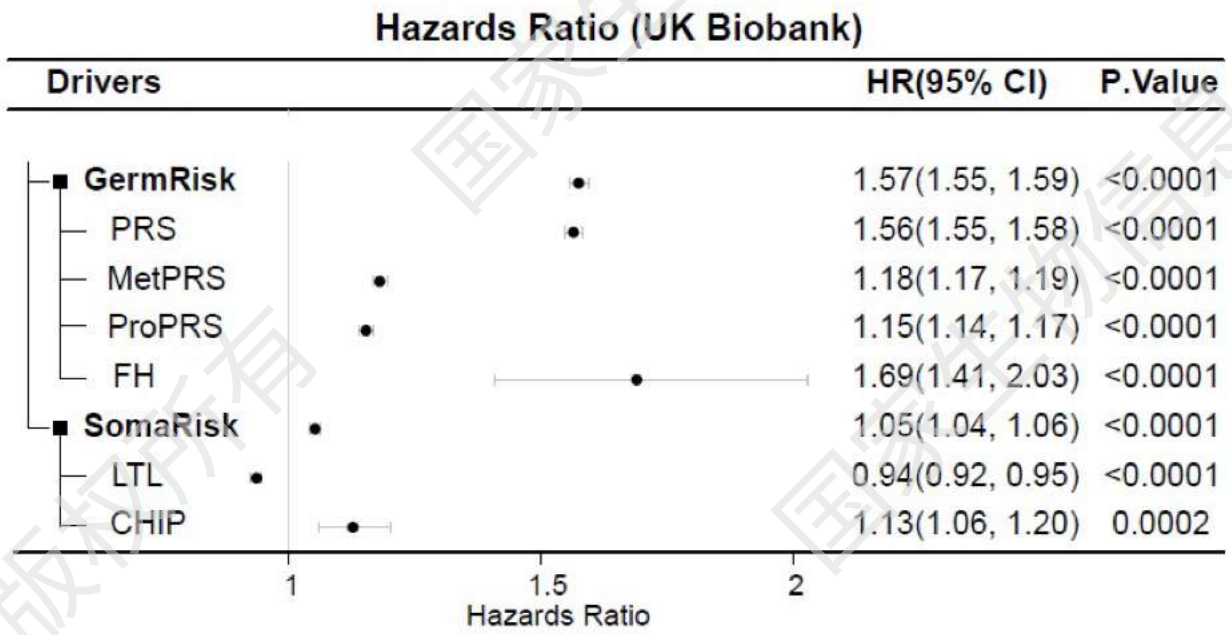


Libby P, Ebert BL. *Circulation*. 2018;
Steensma DP, Bejar R, Jaiswal S, et al. *Blood*. 2015

Contribution of Somatic Mutations and Telomere Length to Cardiovascular Disease Risk

biobank uk

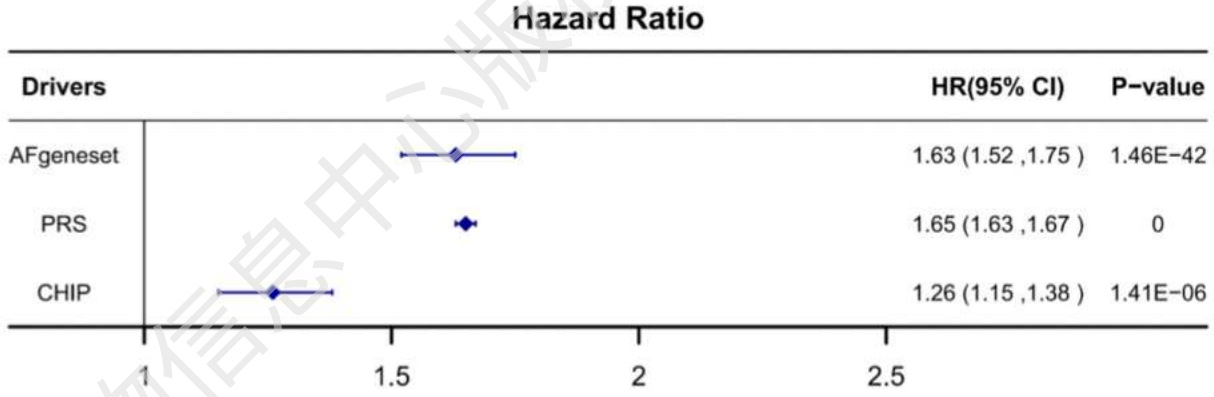
CAD = 42 602
N = 391,536



Xiong Yang … Minxian Wang#, Akl Fahed#
Nat. commun. *(in press)*

biobank uk

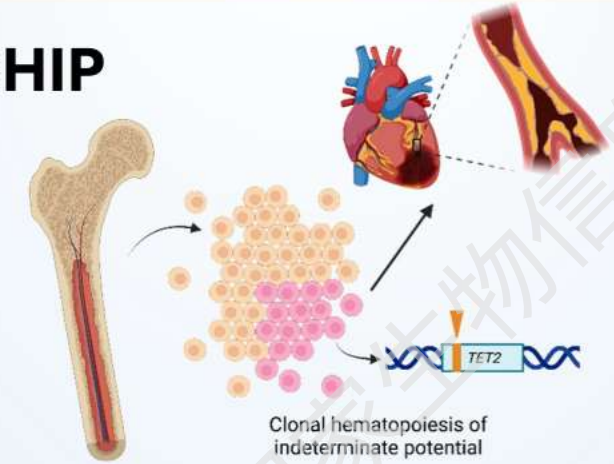
AF = 30 797
N = 41 6085



Rufan Zhang … Minxian Wang#, Akl Fahed#
JAMA cardiology. *2025*

V3.0: Monogenic + PRS + Somatic Mutations

CHIP

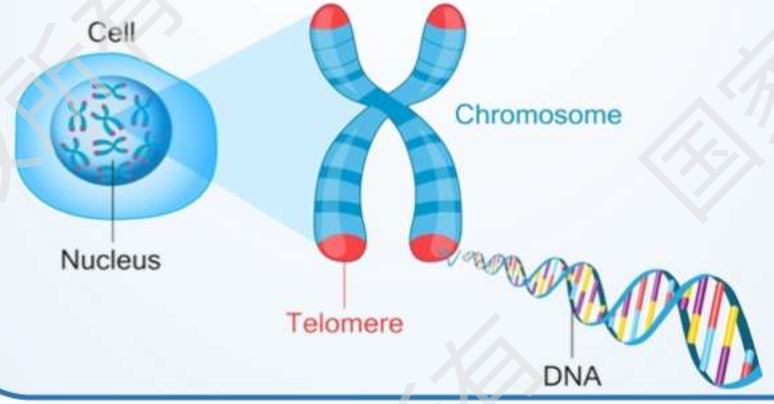


Natarajan P. *ATVB*. 2023;43(1):3-14.

Germline Mutations

- Monogenic → *LDLR, PCKS9 ...*
- Polygenic →→→→→→→→→→ *Polygenic risk*

Telomere length



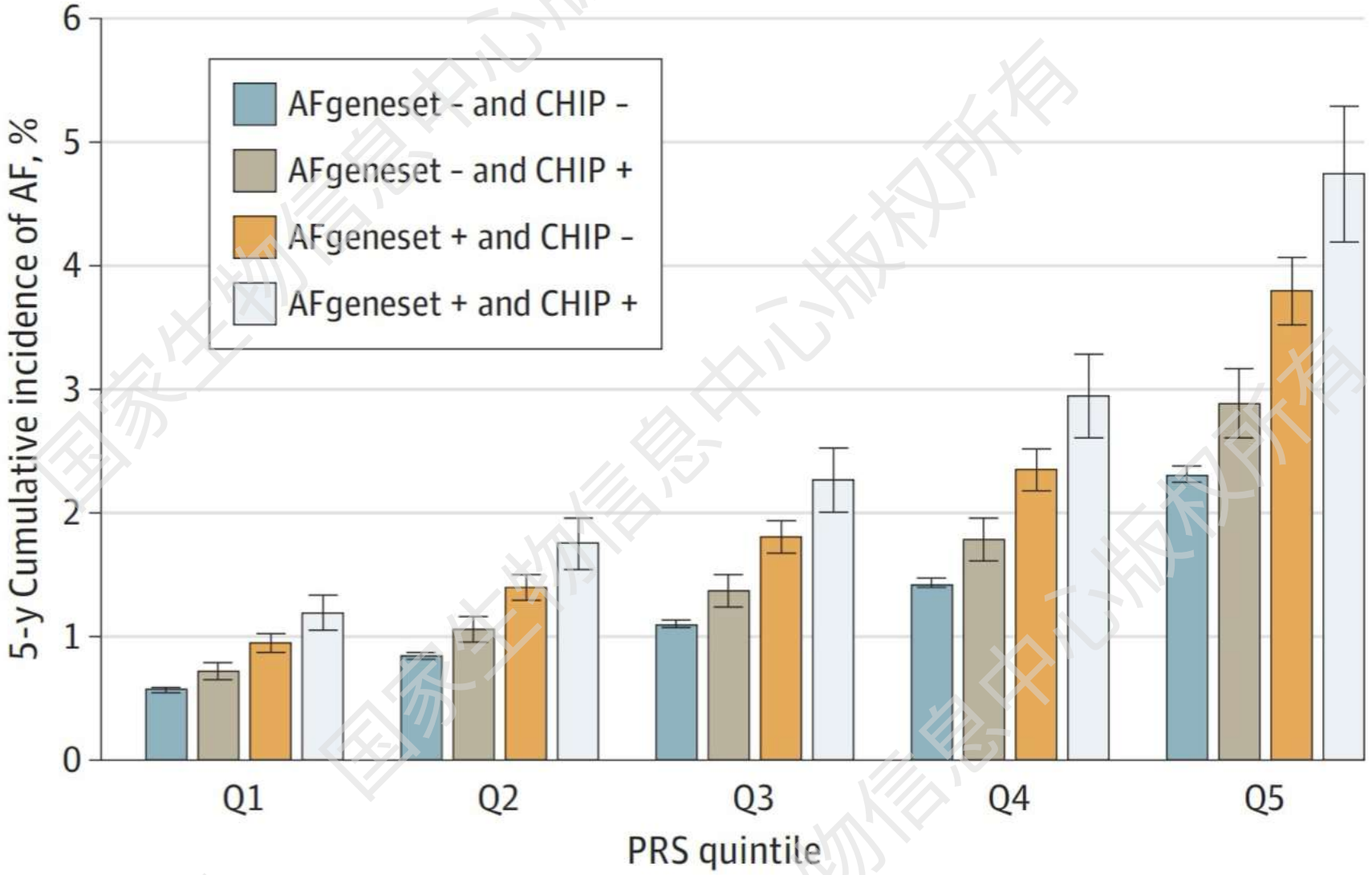
Integrated Genomic Model (IGM)

coxph(formula = Surv(SurvYear, IsIncident)

~ Monogenic + PRS + CHIP + LTL

+ covariates, data = idf)

V3.0: Monogenic + PRS + Somatic Mutations（AF）



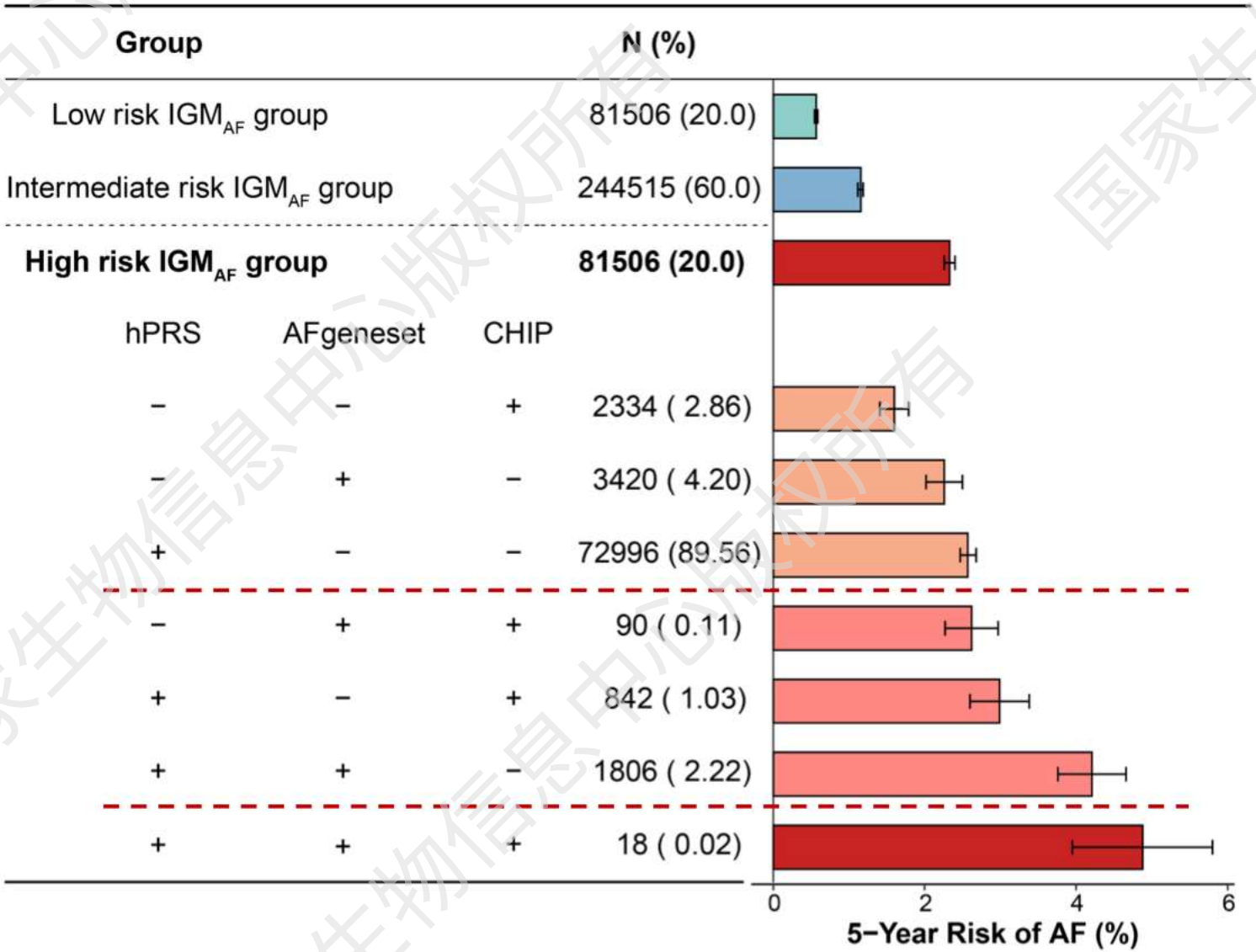
Rufan Zhang … Minxian Wang#, Akl Fahed#. *JAMA cardiology.* 2025 (IF = 15)

V3.0: Monogenic + PRS + Somatic Mutations (AF)

biobank uk

AF = 30 797

N = 41 6085



Rufan Zhang … Minxian Wang#, Akl Fahed#

JAMA cardiology. 2025 (in press, IF = 15)

V3.0: Monogenic + PRS + Somatic Mutations（CAD）

biobank uk

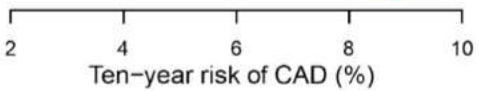
CAD = 42 602
N = 391,536

NHLBI TOPMed

CAD = 4 326
N = 34 177

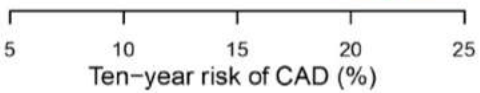
UK Biobank

Group			N(%)
Low risk IGM group			75454 (20%)
Intermediate risk IGM group			226368 (60%)
High risk IGM group			**75454 (20%)**
High germline risk	High somatic risk		
–	–		624 (0.8%)
–	+		4527 (6.0%)
+	–		54785 (72.6%)
+	+		15520 (20.6%)



TOPMed

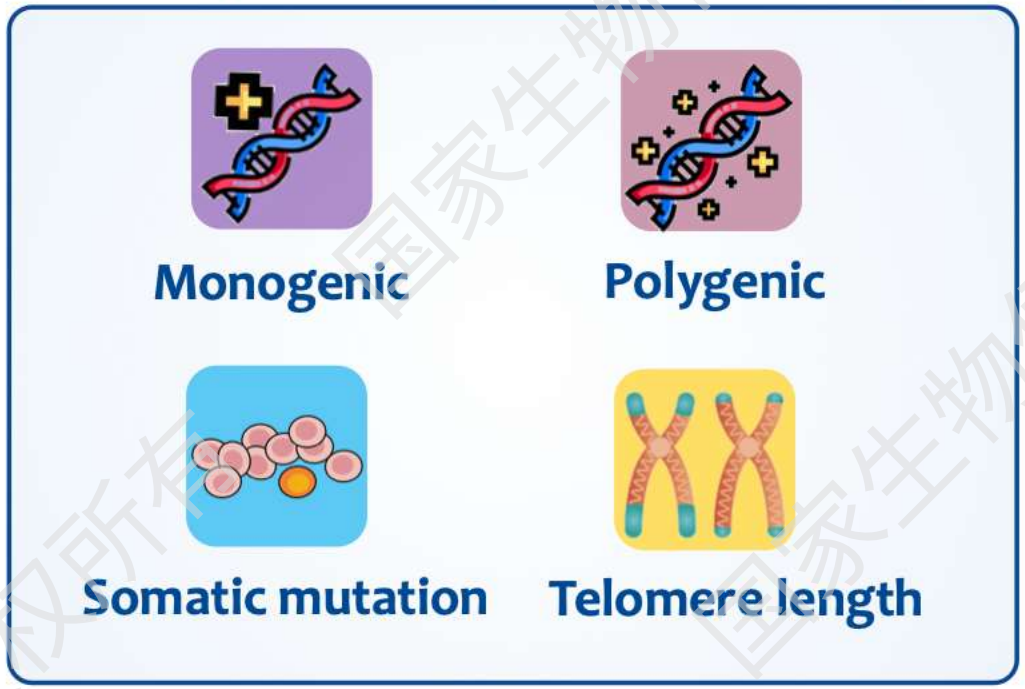
Group			N(%)
Low risk IGM group			5504 (20%)
Intermediate risk IGM group			16512 (60%)
High risk IGM group			**5504 (20%)**
High germline risk	High somatic risk		
–	–		54 (1.0%)
–	+		313 (5.7%)
+	–		4016 (73.0%)
+	+		1121 (20.4%)



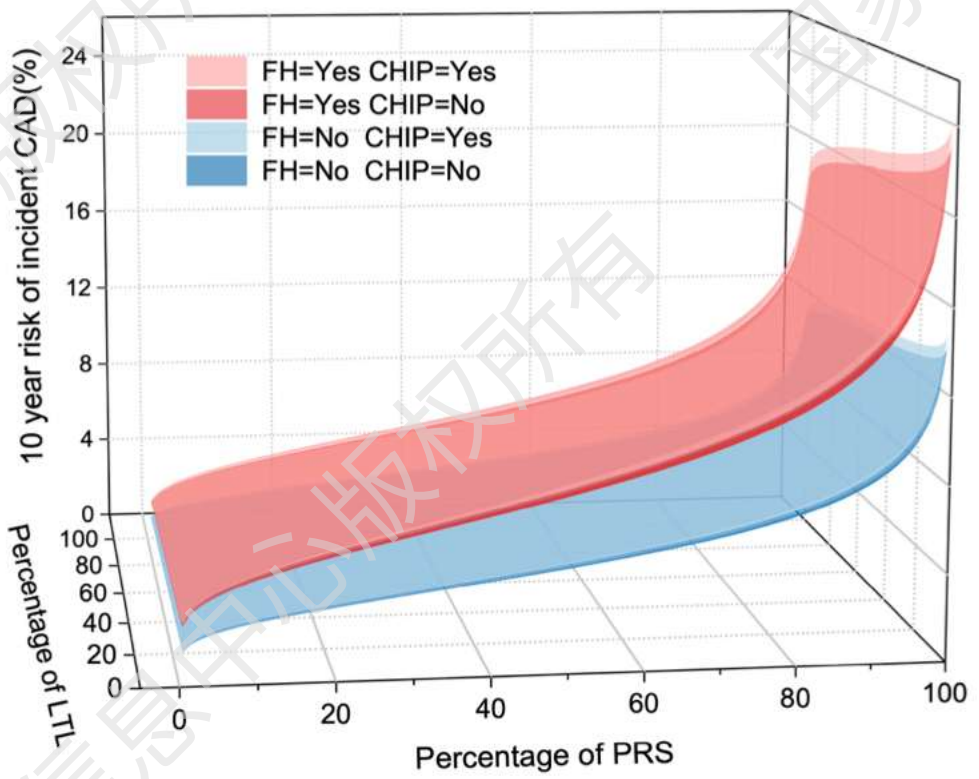
Xiong Yang … Minxian Wang#, Akl Fahed#
Nat commu. (round review)

Personalized risk prediction V3.0

multiple genetic risk components joint modeling

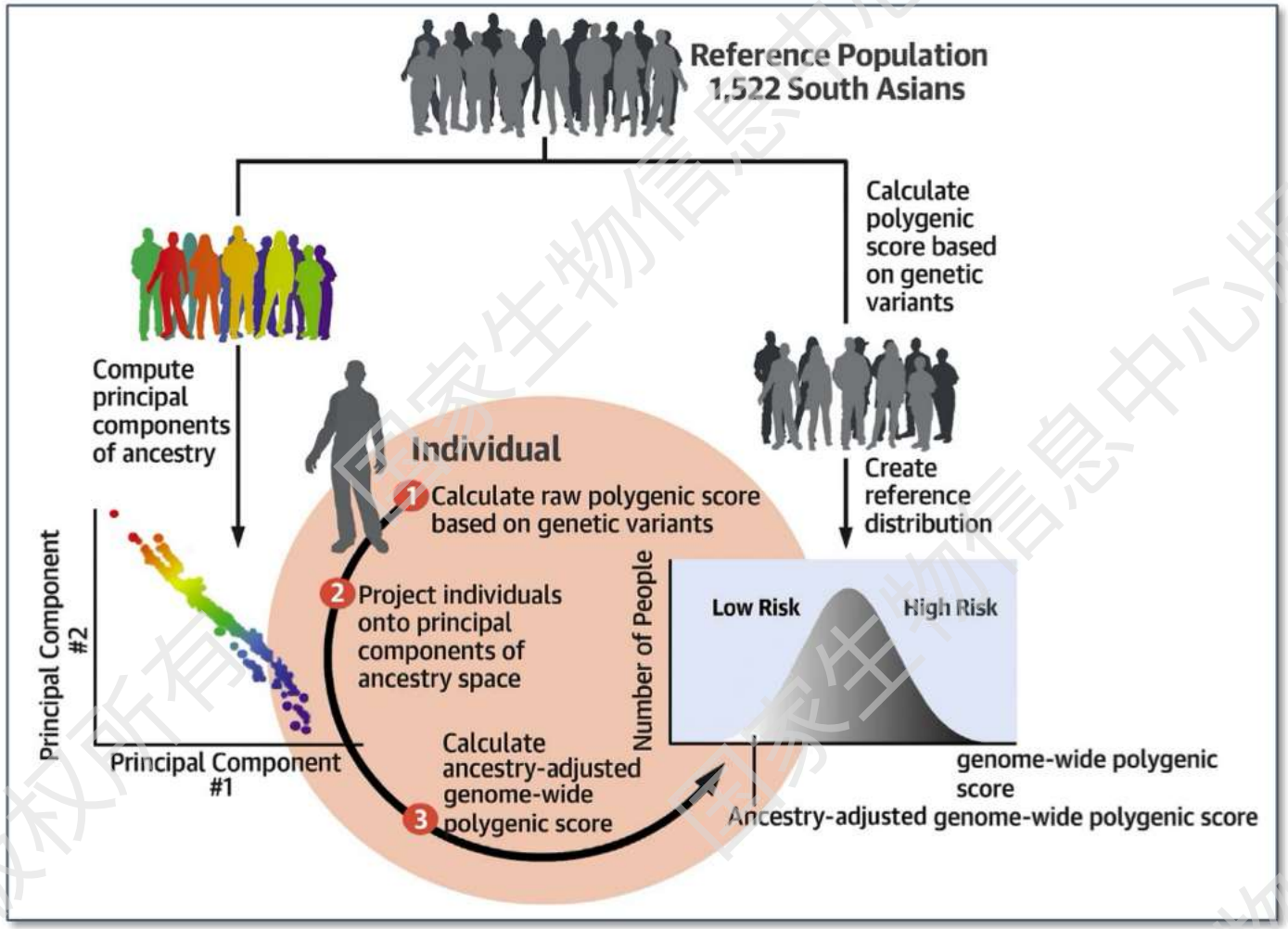


UK Biobank 400K individuals



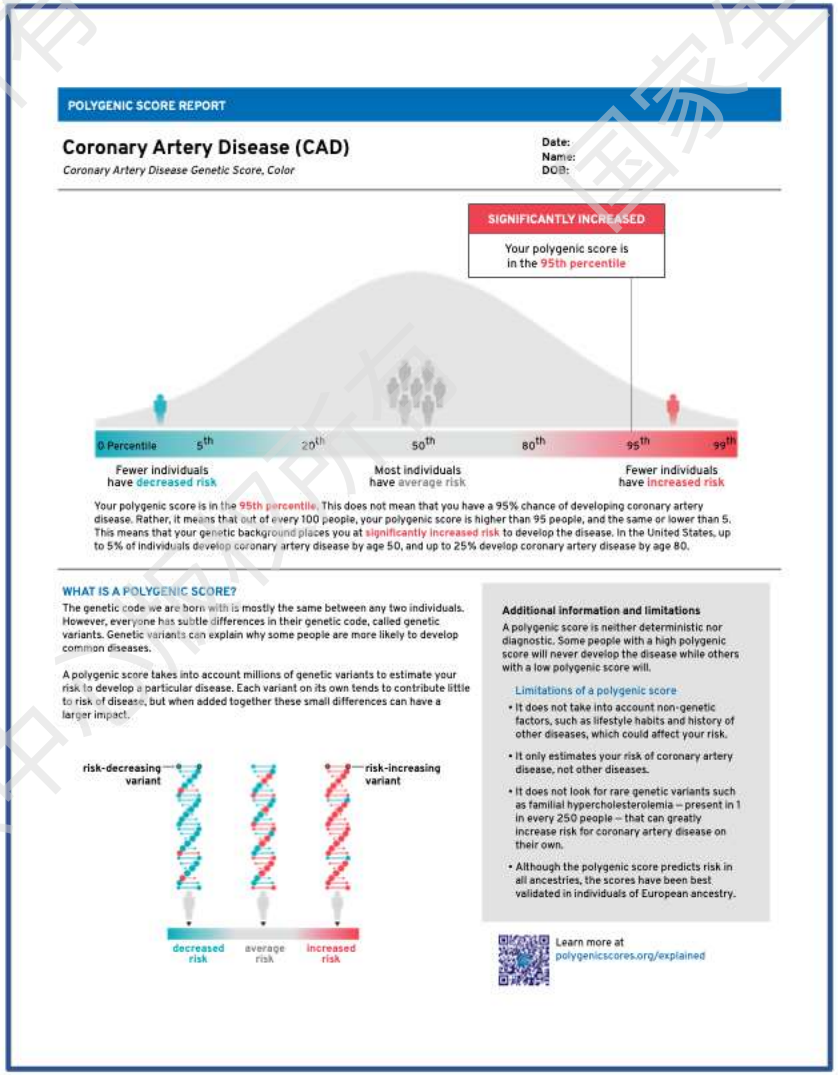
Unpublished

Personalized Genetic Risk Assessment and Clinical Applications



汪敏先等 *J. Am. Coll. Cardiol.* 2020 (IF 24)

Risk Report



POLYGENIC SCORE REPORT

Coronary Artery Disease (CAD)

Coronary Artery Disease Genetic Score, Color

Date:
Name:
DOB:

SIGNIFICANTLY INCREASED

Your polygenic score is in the 95th percentile

0 Percentile | 5th | 20th | 50th | 80th | 95th | 99th

Fewer individuals have decreased risk

Most individuals have average risk

Fewer individuals have increased risk

Your polygenic score is in the 95th percentile. This does not mean that you have a 95% chance of developing coronary artery disease. Rather, it means that out of every 100 people, your polygenic score is higher than 95 people, and the same or lower than 5. This means that your genetic background places you at significantly increased risk to develop the disease. In the United States, up to 5% of individuals develop coronary artery disease by age 50, and up to 25% develop coronary artery disease by age 80.

WHAT IS A POLYGENIC SCORE?

The genetic code we are born with is mostly the same between any two individuals. However, everyone has subtle differences in their genetic code, called genetic variants. Genetic variants can explain why some people are more likely to develop common diseases.

A polygenic score takes into account millions of genetic variants to estimate your risk to develop a particular disease. Each variant on its own tends to contribute little to risk of disease, but when added together these small differences can have a larger impact.

risk-decreasing variant

risk-increasing variant

decreased risk | average risk | increased risk

Additional information and limitations

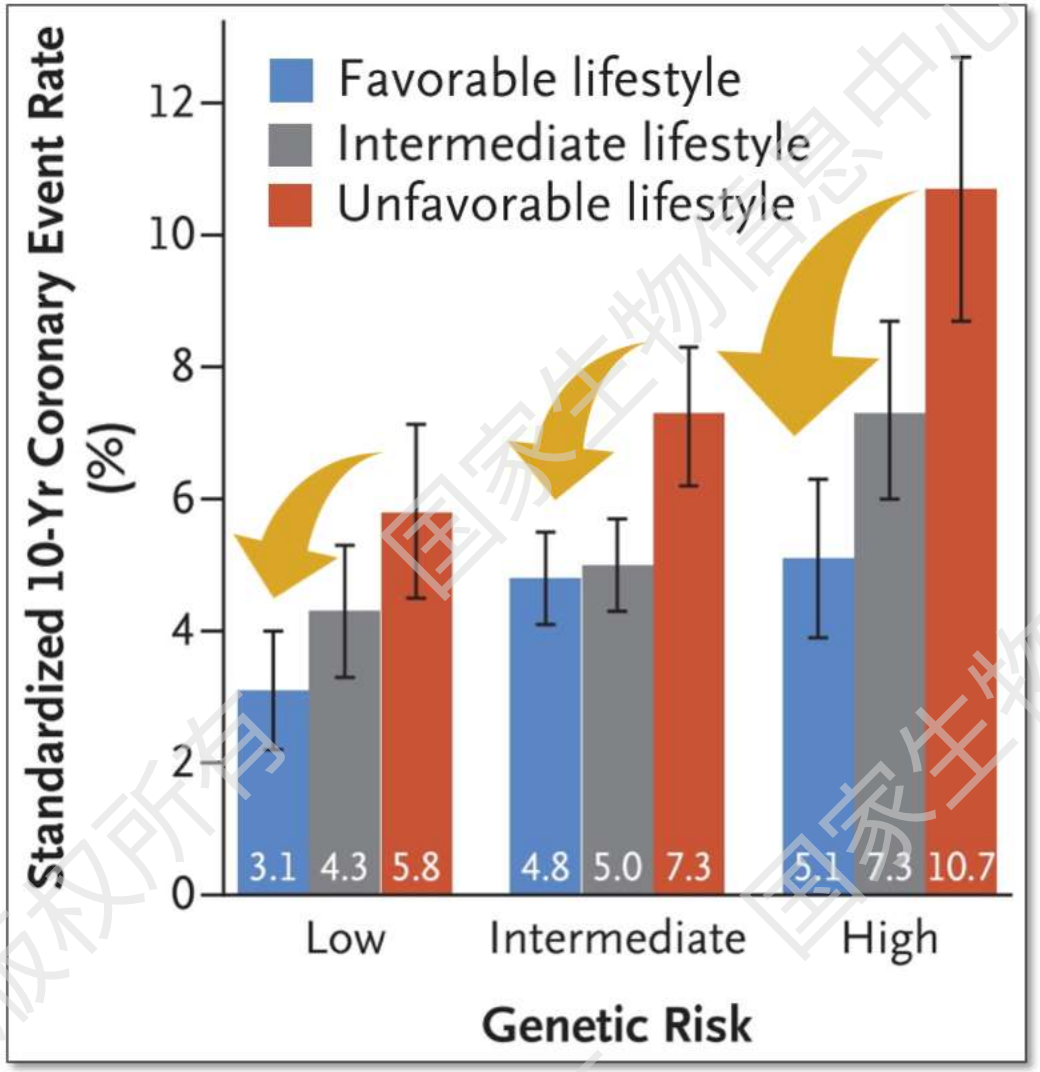
A polygenic score is neither deterministic nor diagnostic. Some people with a high polygenic score will never develop the disease while others with a low polygenic score will.

Limitations of a polygenic score

- It does not take into account non-genetic factors, such as lifestyle habits and history of other diseases, which could affect your risk.
- It only estimates your risk of coronary artery disease, not other diseases.
- It does not look for rare genetic variants such as familial hypercholesterolemia – present in 1 in every 250 people – that can greatly increase risk for coronary artery disease on their own.
- Although the polygenic score predicts risk in all ancestries, the scores have been best validated in individuals of European ancestry.

Learn more at polygenicscores.org/explained

DNA is not destiny !

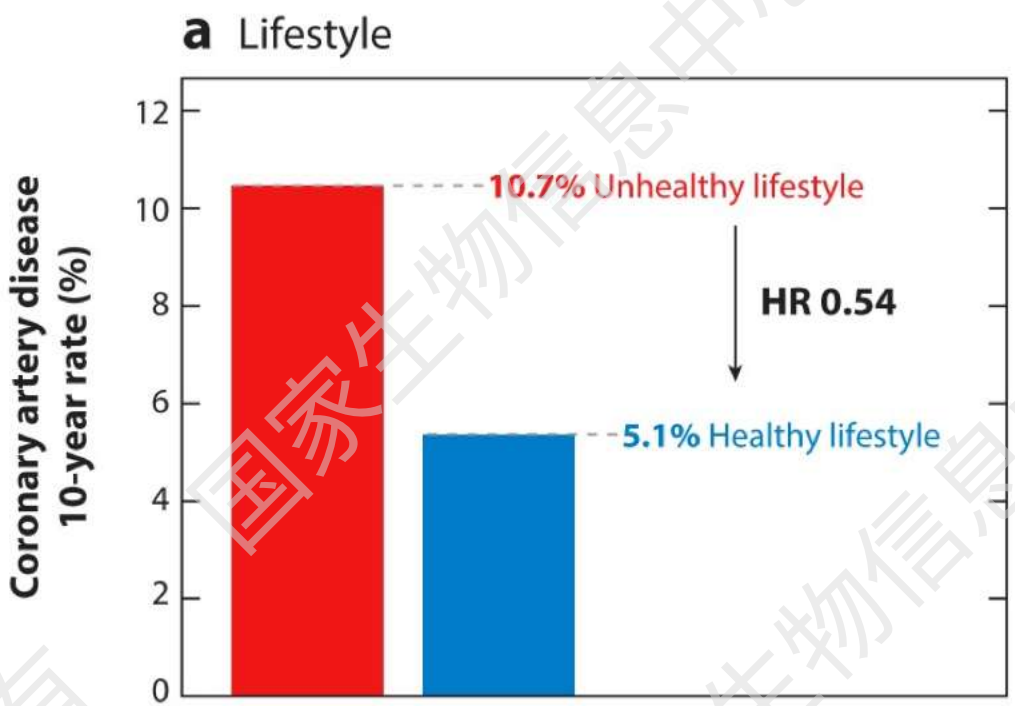


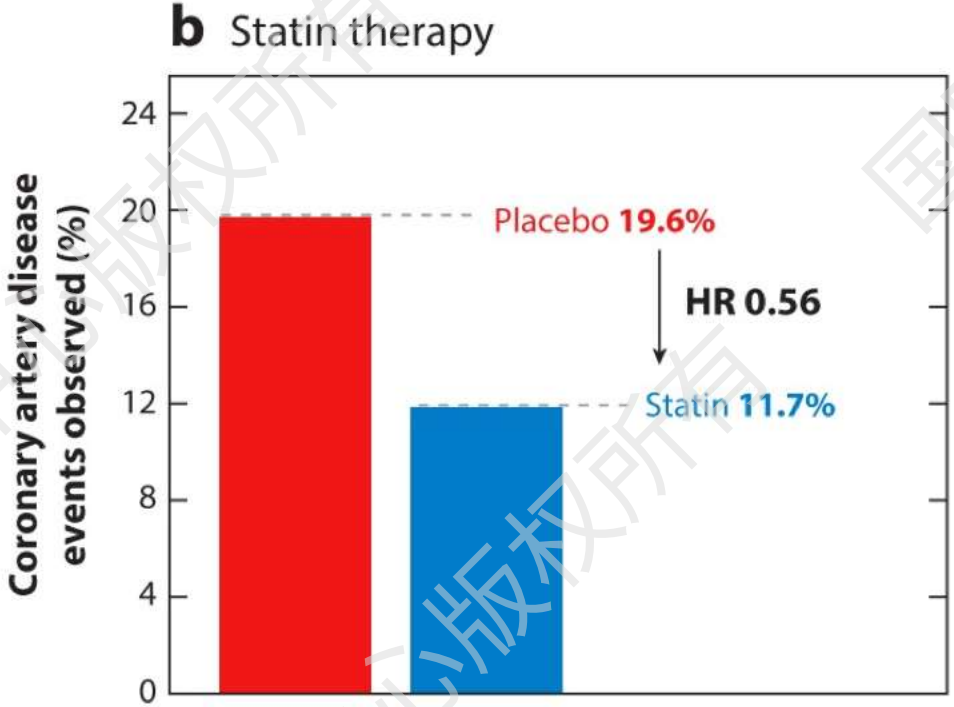
Khera, A. V. *et al.* ***N. Engl. J. Med.*** 2016.

Health lifestyle for cardiovascular disease

- **No smoking, no drinking**
- **Good exercise**
- **Healthy diet**
- **Keep a low BMI**

DNA is not destiny !





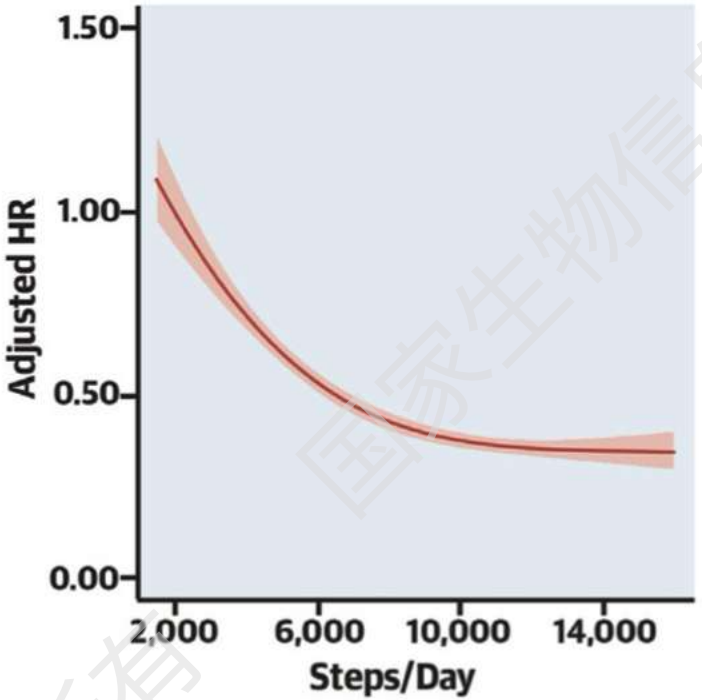
American Heart Association. Healthy for Good

Life's Essential 8

Patel AP, Khera A V. ***Annu Rev Med***. 2022

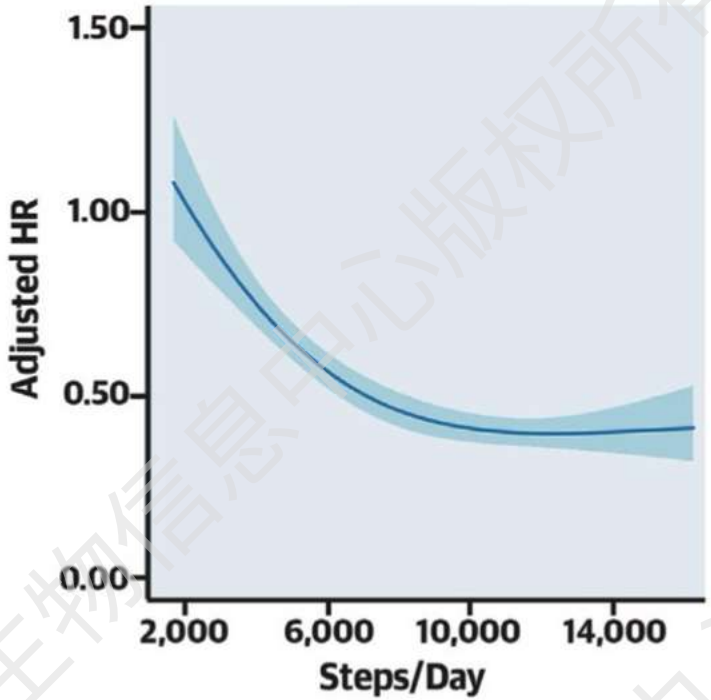
DNA is not destiny -- Offset disease risk is easy!!!

All-cause mortality



Steps	Lower risk
2517	8%
8763	60%
16,000	65%

Cardiovascular disease risk



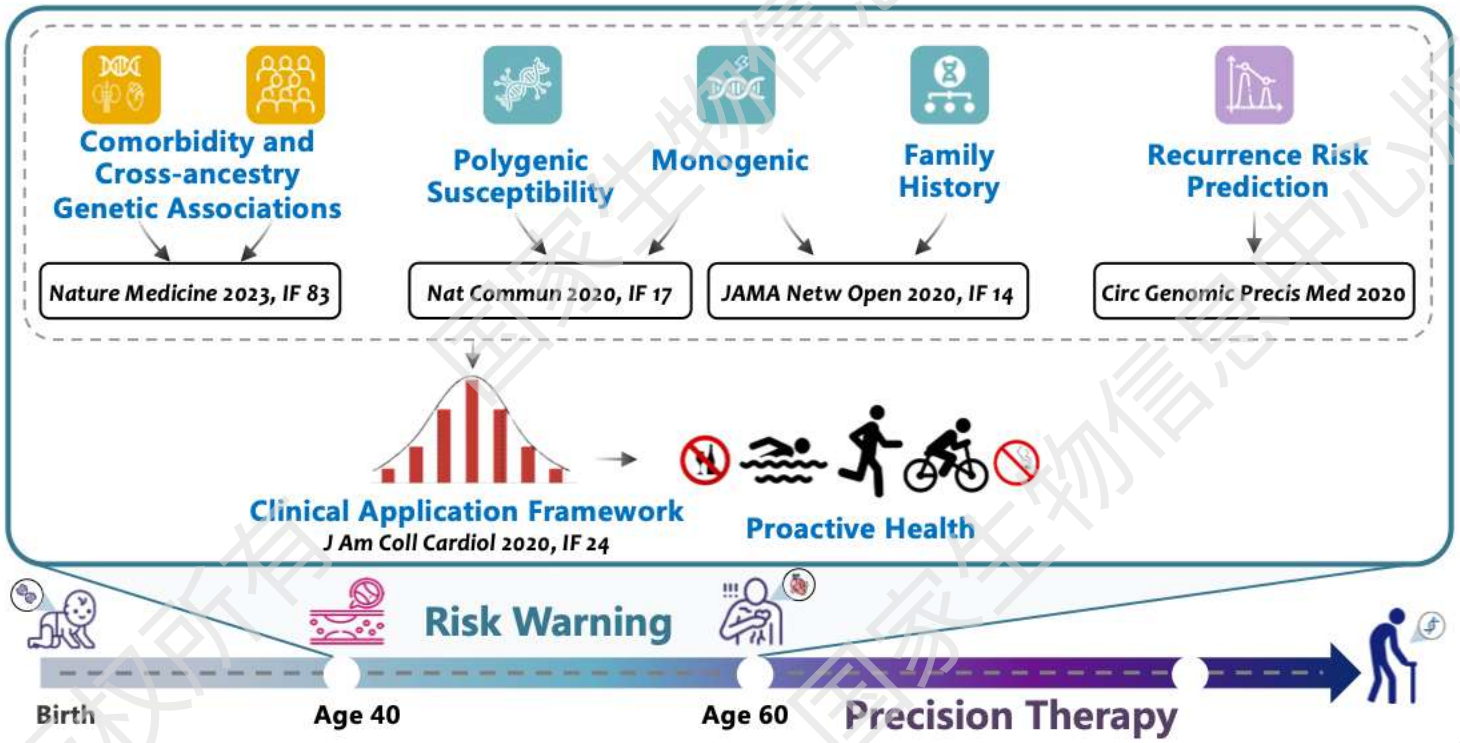
Steps	Lower risk
2735	11%
7126	51%
16,000	58%

12 cohorts
110k individuals
Meta-analysis

Optimal steps:
8K per day

Stens NA, et al. ***J Am Coll Cardiol***. 2023

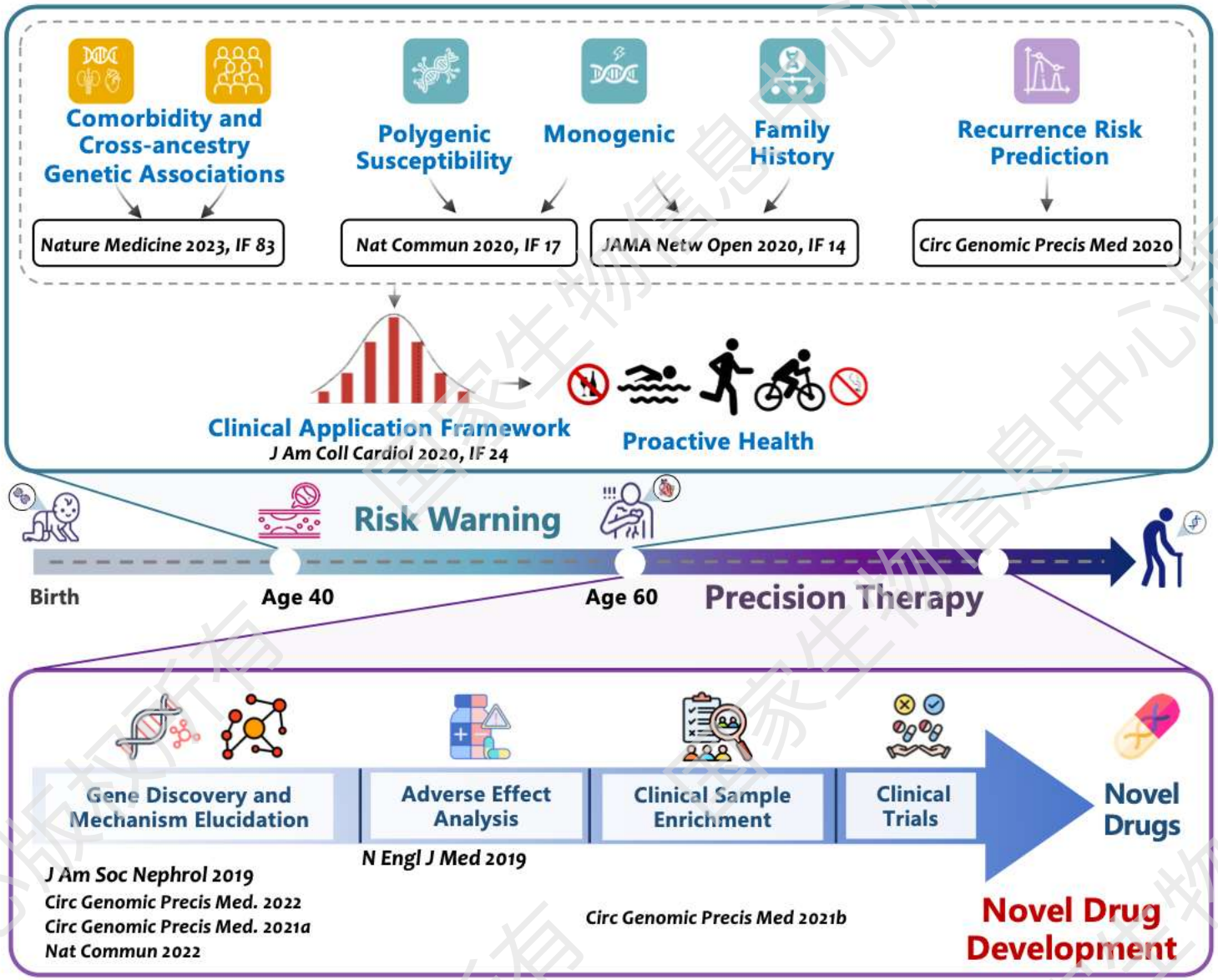
Integrated Multi-Factor Disease Risk Prediction Model



Chronic Disease Risk Management

- Cardiovascular and metabolic diseases are influenced by complex interactions between genetic and environmental factors
- **Systematically integrate** multiple risk factors for the occurrence and development of cardiovascular and metabolic diseases, develop personalized risk assessment and early warning systems, promote proactive health, and reduce disease incidence at the population level

Cohort Genomics and Precision Medicine



Cohort Genomics → Precision Medicine

- Integrate multi-factor risks → personalized risk assessment & early warning. Promote proactive health, reduce population-level disease incidence
- Leveraging large-scale cohort genomic data to **streamline multiple key steps in genetic target discovery**, facilitating innovative drug development for coronary artery disease.

中华心血管病杂志《队列基因组学与冠心病精准医学》

Acknowledgement

- Sekar Kathiresan lab
- Amit Khera lab
- Akl Fahed lab
- Aniruddh Patel
- Krishna Aragam
- Saaket Agrawal

INNOVATION TOGETHER

Wang lab

A data-driven precision medicine lab