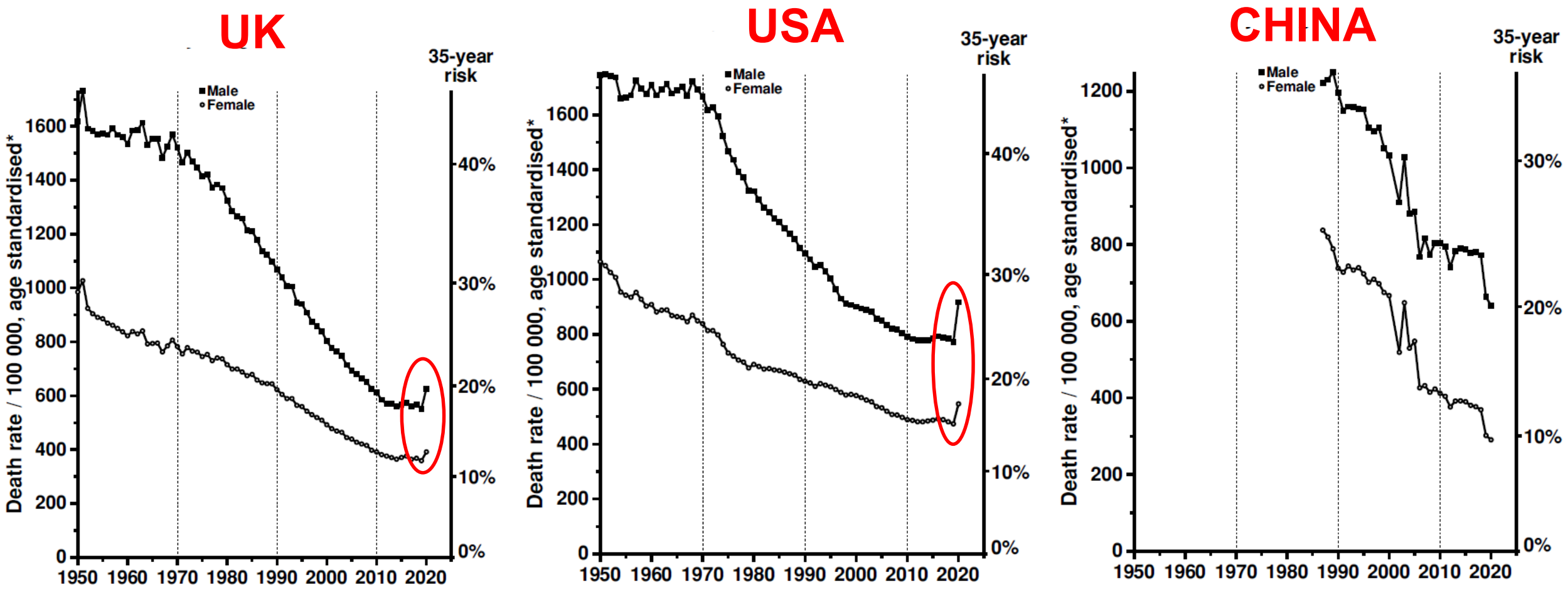
Harnessing the power of proteogenomics in biobanks to advance precision health

Zhengming CHEN
Richard Peto Chair Professor in Epidemiology
Big Data Institute & Oxford Population Health
University of Oxford
Zhengming.chen@ndph.ox.ac.uk

Hong Kong Health Research Symposium, 26th Nov 2024

All-cause mortality at ages 35-69



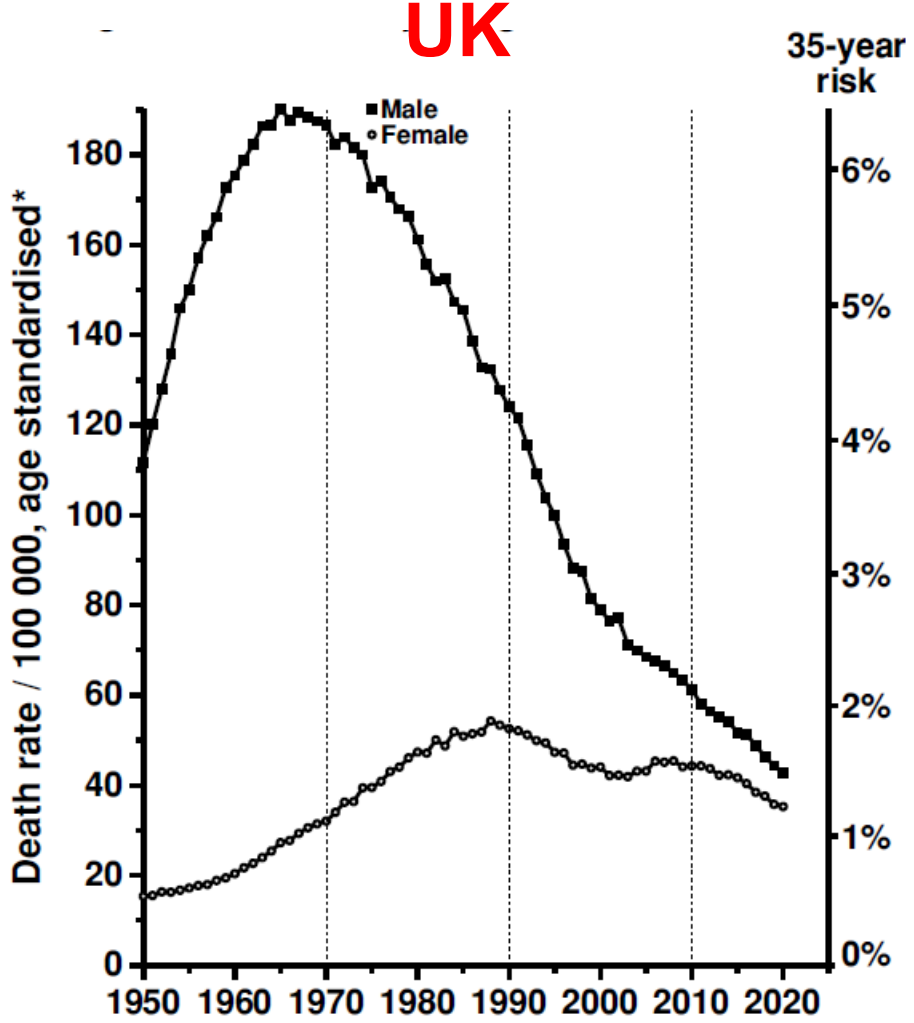
Main COVID deaths in China took place during 12/2022-02/2023

OXFORD POPULATION HEALTH

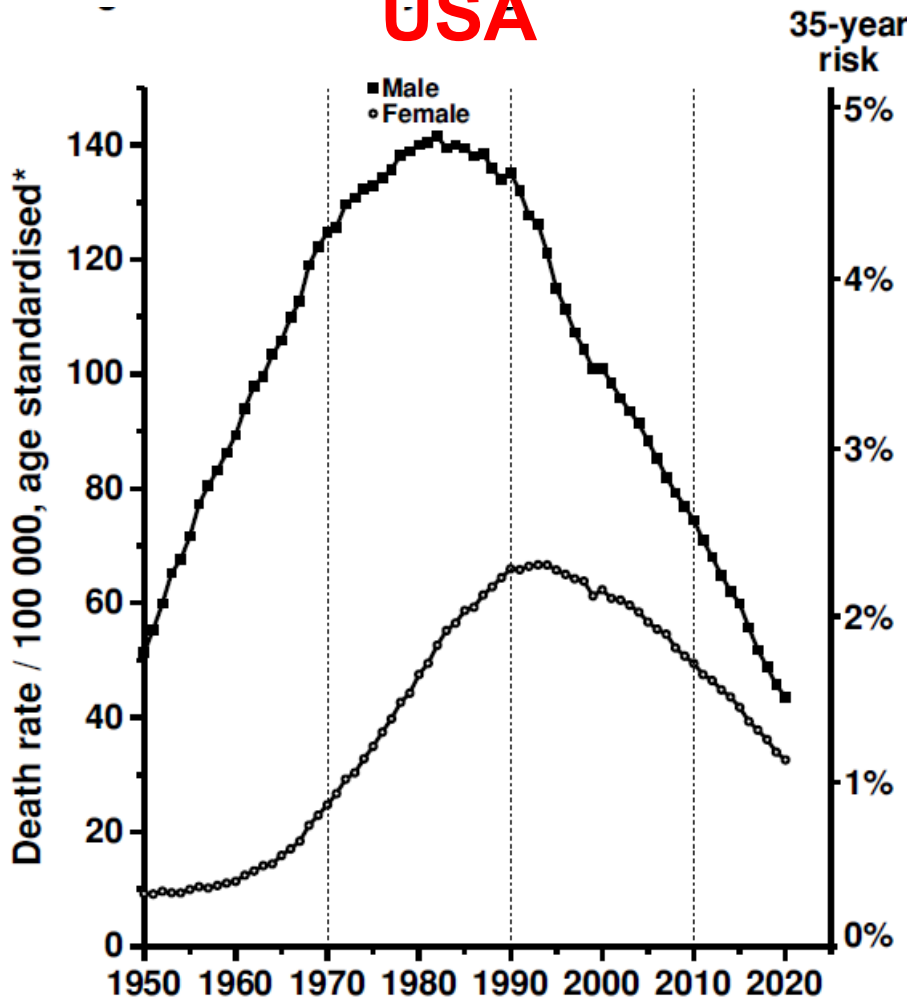
Source: WHO mortality & UN population estimates

Lung cancer mortality at ages 35-69

UK



USA

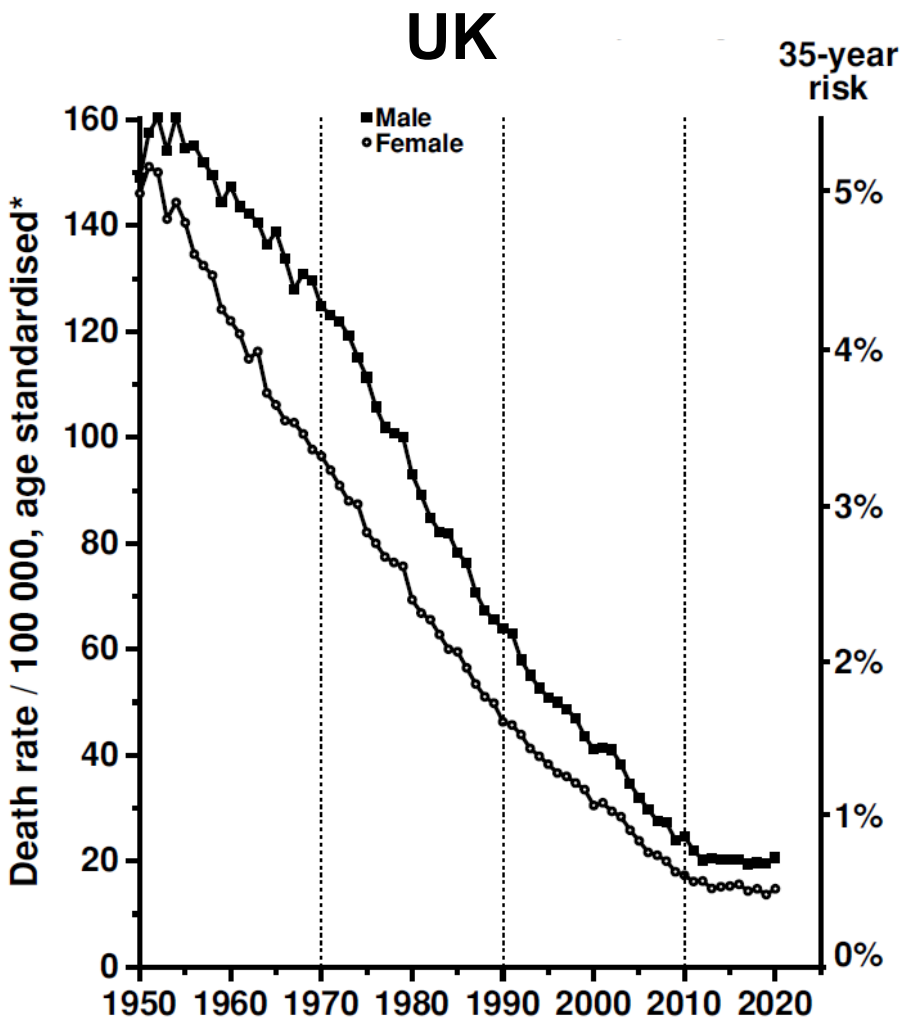


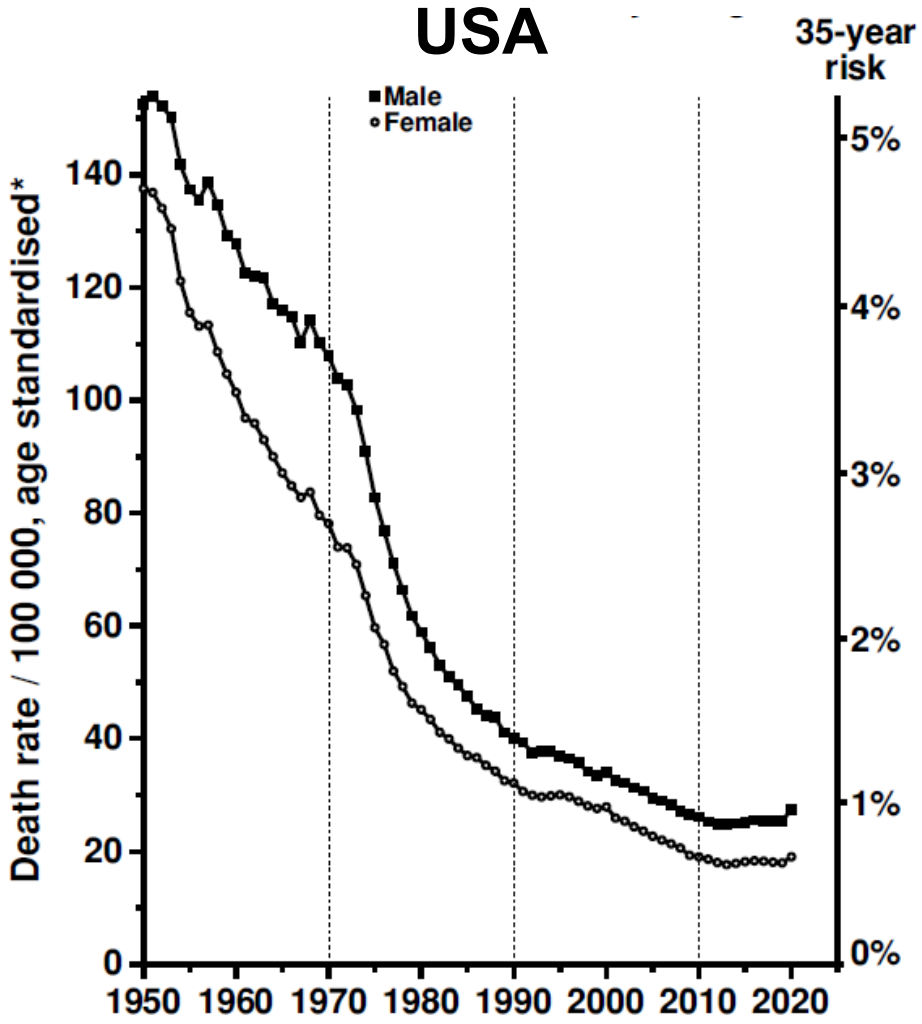
Driven entirely by changes in cigarette consumption

OXFORD POPULATION HEALTH

Source: WHO mortality & UN population estimates

Rapidly falling stroke mortality in UK/USA: not fully explained





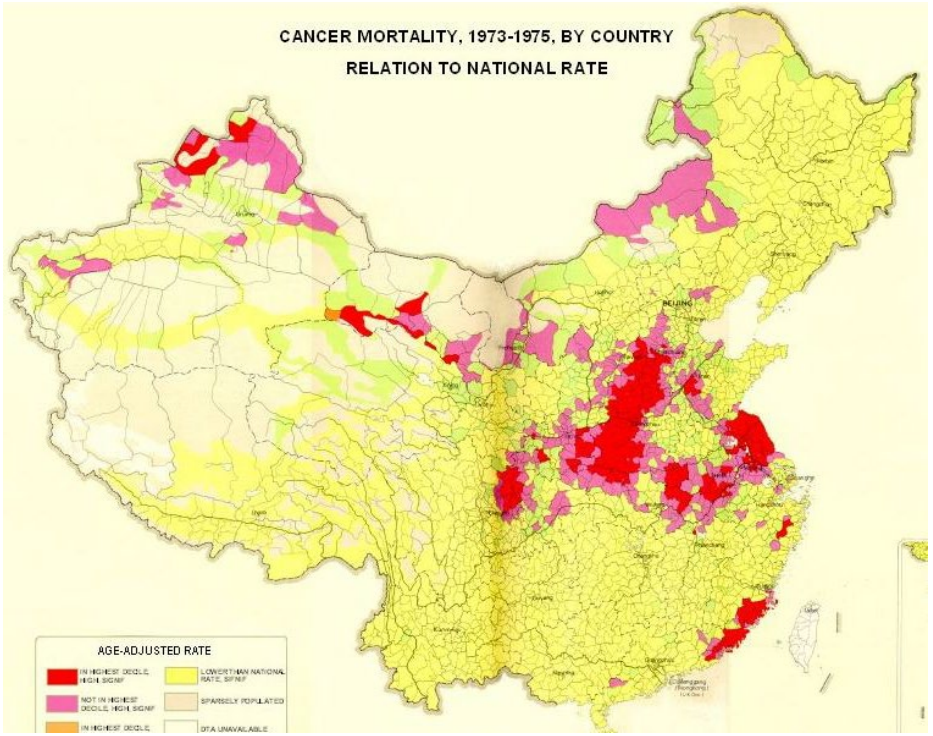
China: >2M stroke deaths/year, with large regional variations

OXFORD POPULATION HEALTH

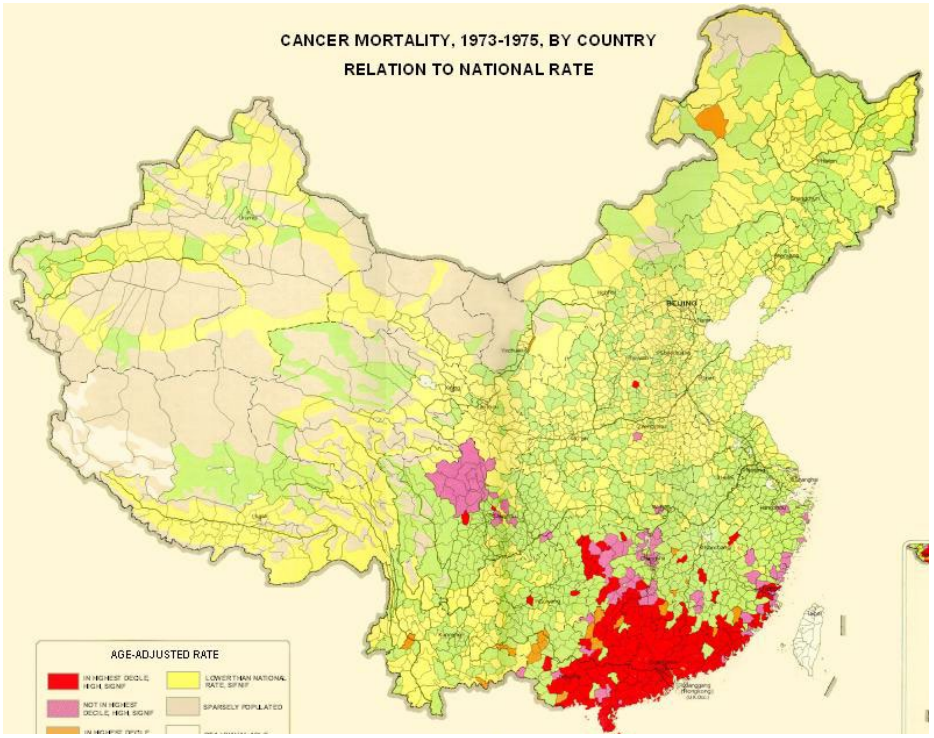
Source: WHO mortality & UN population estimates

Large, unexplained cancer mortality variation

Oesophagus cancer



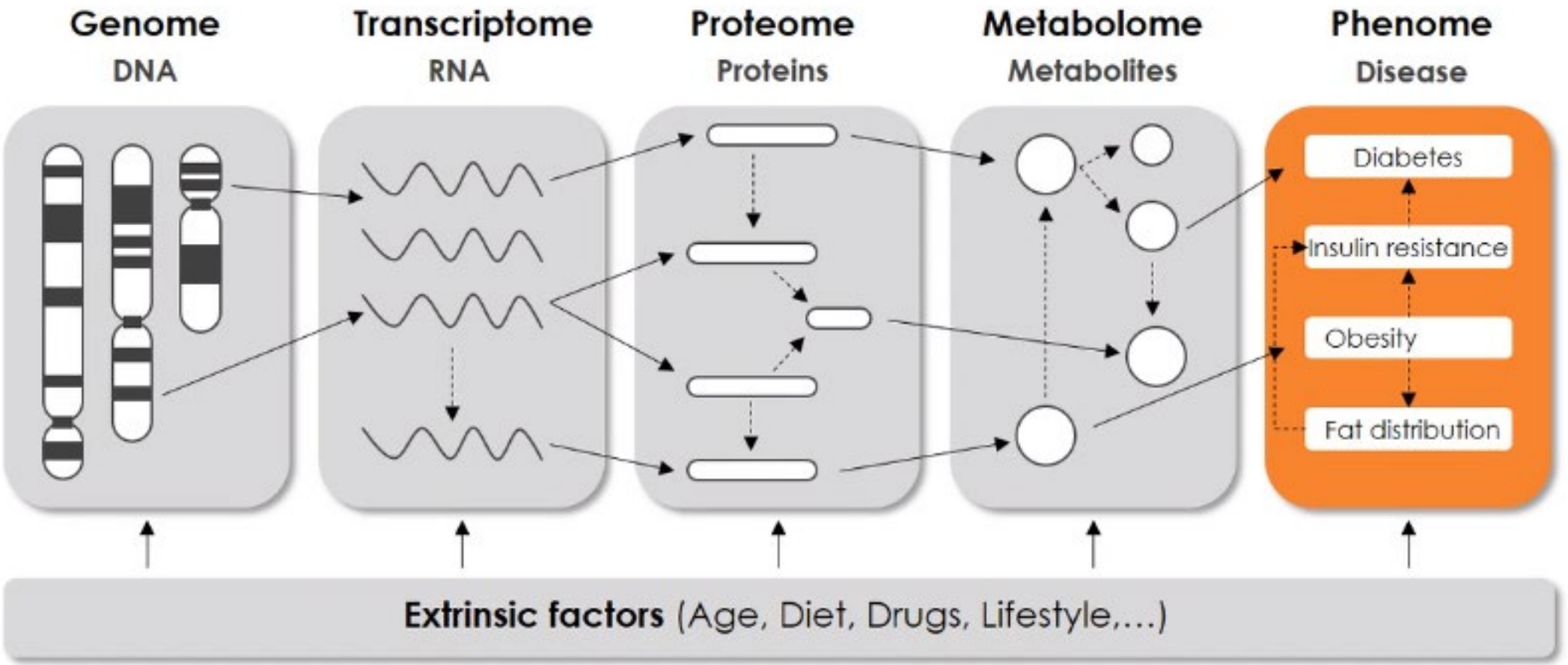
Nasopharynx cancer



Females only, so little effect of tobacco or alcohol
(Red = high mortality >10x Green = low mortality)

OXFORD POPULATION HEALTH

Aetiology of complex human diseases



Complex interplays of “nature” & “nurture”, which may differ between populations

Need prospective cohort studies in diverse populations

OXFORD POPULATION HEALTH

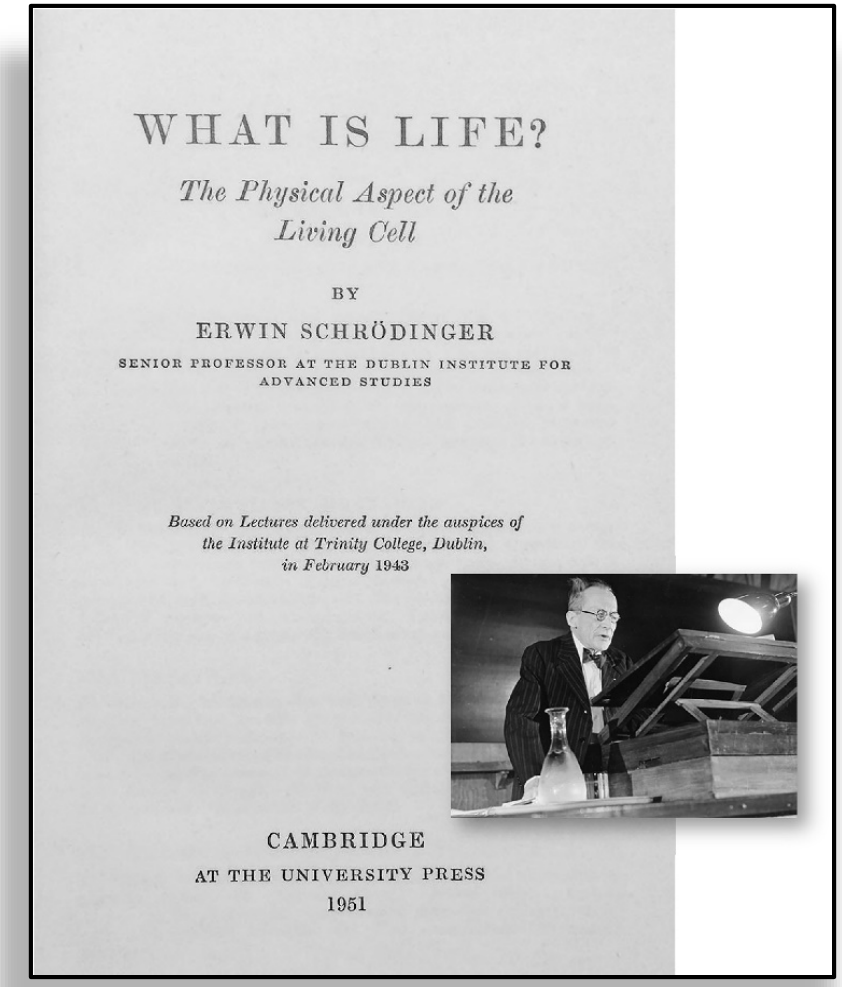
Need new evidence base to advance precision health

- Better understanding of human biology and disease aetiology
- Improved disease prediction, detection and diagnosis
- Identification of new or repurposing drug targets
- Reliable patient stratification for more targeted management

Proteomics will play a key role in empowering such development

OXFORD POPULATION HEALTH

Human Proteome and Health



WHAT IS LIFE?

The Physical Aspect of the Living Cell

BY

ERWIN SCHRÖDINGER

SENIOR PROFESSOR AT THE DUBLIN INSTITUTE FOR ADVANCED STUDIES

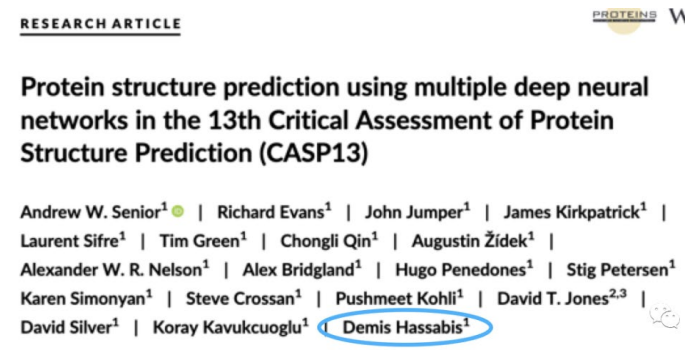
Based on Lectures delivered under the auspices of the Institute at Trinity College, Dublin, in February 1943

CAMBRIDGE

AT THE UNIVERSITY PRESS

1951

Schrodinger's 1944 monograph "What is life" predicted the importance of proteome as a key link between hereditary mechanism, growth and disease.

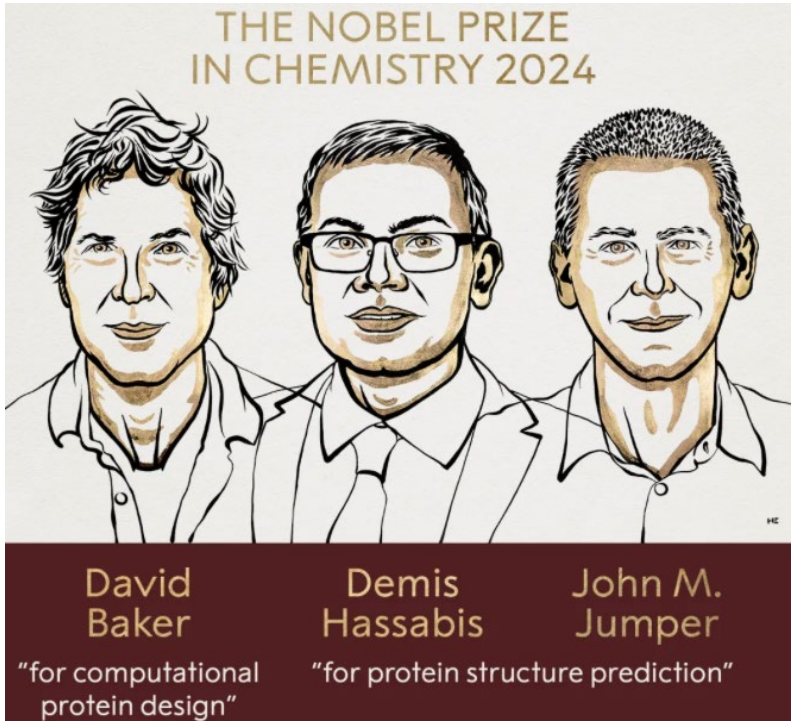


RESEARCH ARTICLE

PROTEINS

Protein structure prediction using multiple deep neural networks in the 13th Critical Assessment of Protein Structure Prediction (CASP13)

Andrew W. Senior[1] | Richard Evans[1] | John Jumper[1] | James Kirkpatrick[1] | Laurent Sifre[1] | Tim Green[1] | Chongli Qin[1] | Augustin Žídek[1] | Alexander W. R. Nelson[1] | Alex Bridgland[1] | Hugo Penedones[1] | Stig Petersen[1] | Karen Simonyan[1] | Steve Crossan[1] | Pushmeet Kohli[1] | David T. Jones[2,3] | David Silver[1] | Koray Kavukcuoglu[1] | Demis Hassabis[1]

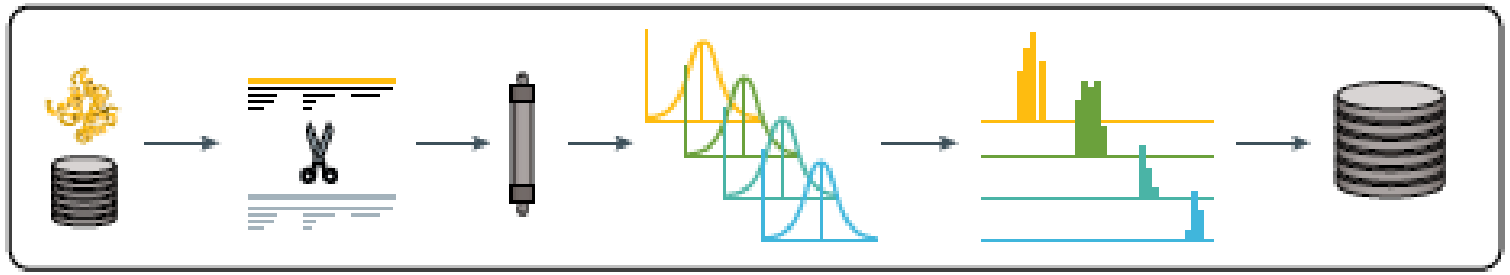


OXFORD POPULATION HEALTH

Integrated proteogenomic analyses will be transformative, facilitated by new technology

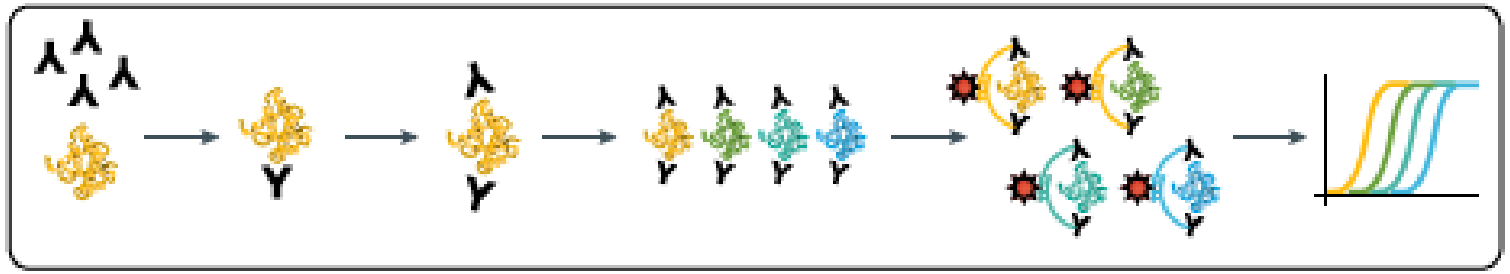
From proteins to proteome

Mass spectrometry-based proteomics



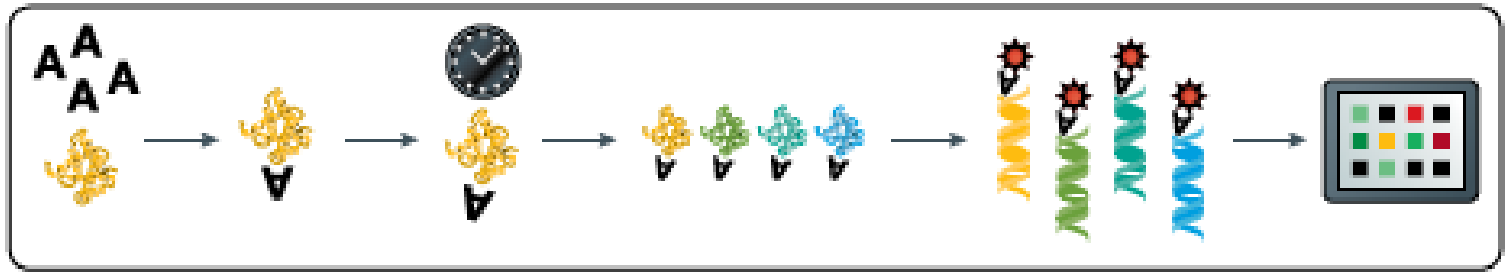
"Gold" standard

Antibody-based affinity proteomics



OLINK Explore

Aptamer-based affinity proteomics



SomaScan

Carrasco-Zanini J. et al. Curr Diab Rep. 2020

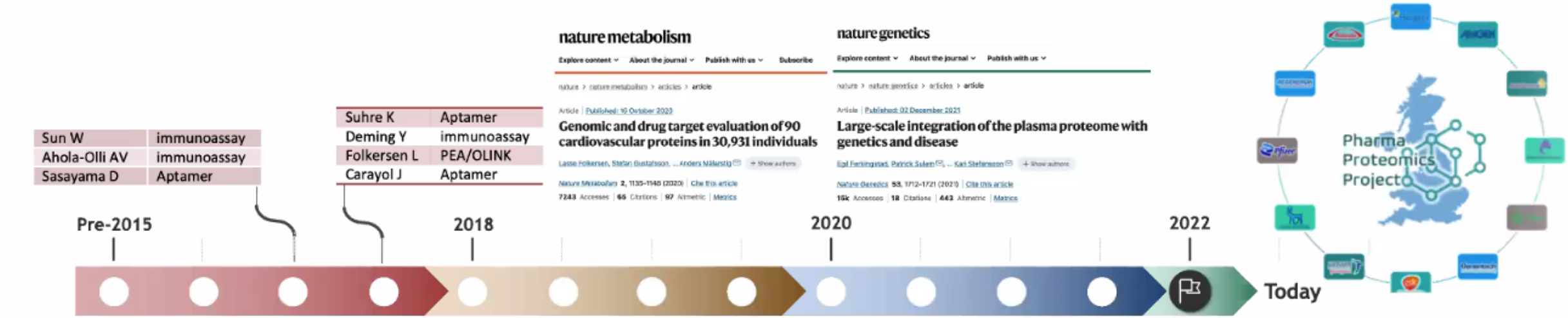
K Suhre, et al. Nature Review (2021)

OXFORD POPULATION HEALTH

Application of proteomics in population studies

Sun W	immunoassay
Ahola-Olli AV	immunoassay
Sasayama D	Aptamer

Suhre K	Aptamer
Deming Y	immunoassay
Folkersen L	PEA/OLINK
Carayol J	Aptamer



Small scale, small sample sizes

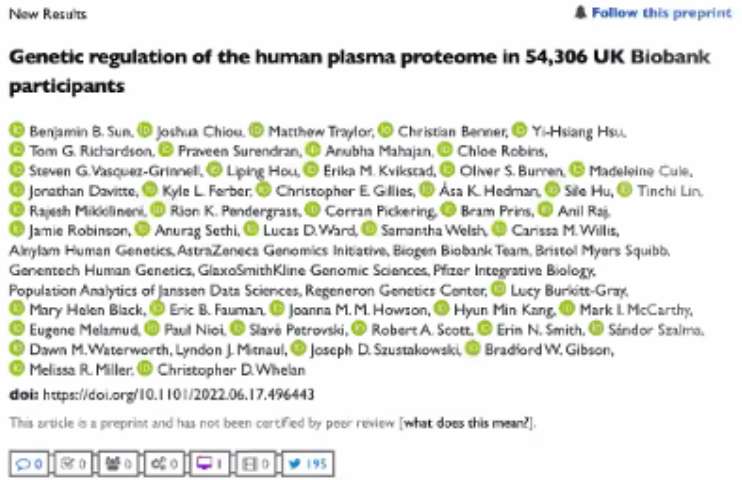
Melzer D	immunoassay
Johansson A	MS
Kim S	immunoassay
Enroth S	PEA/OLINK
Kauwe JS	immunoassay

Scaling up to thousands

Sun BB	Aptamer
Emilsson V	Aptamer
Yao C	immunoassay

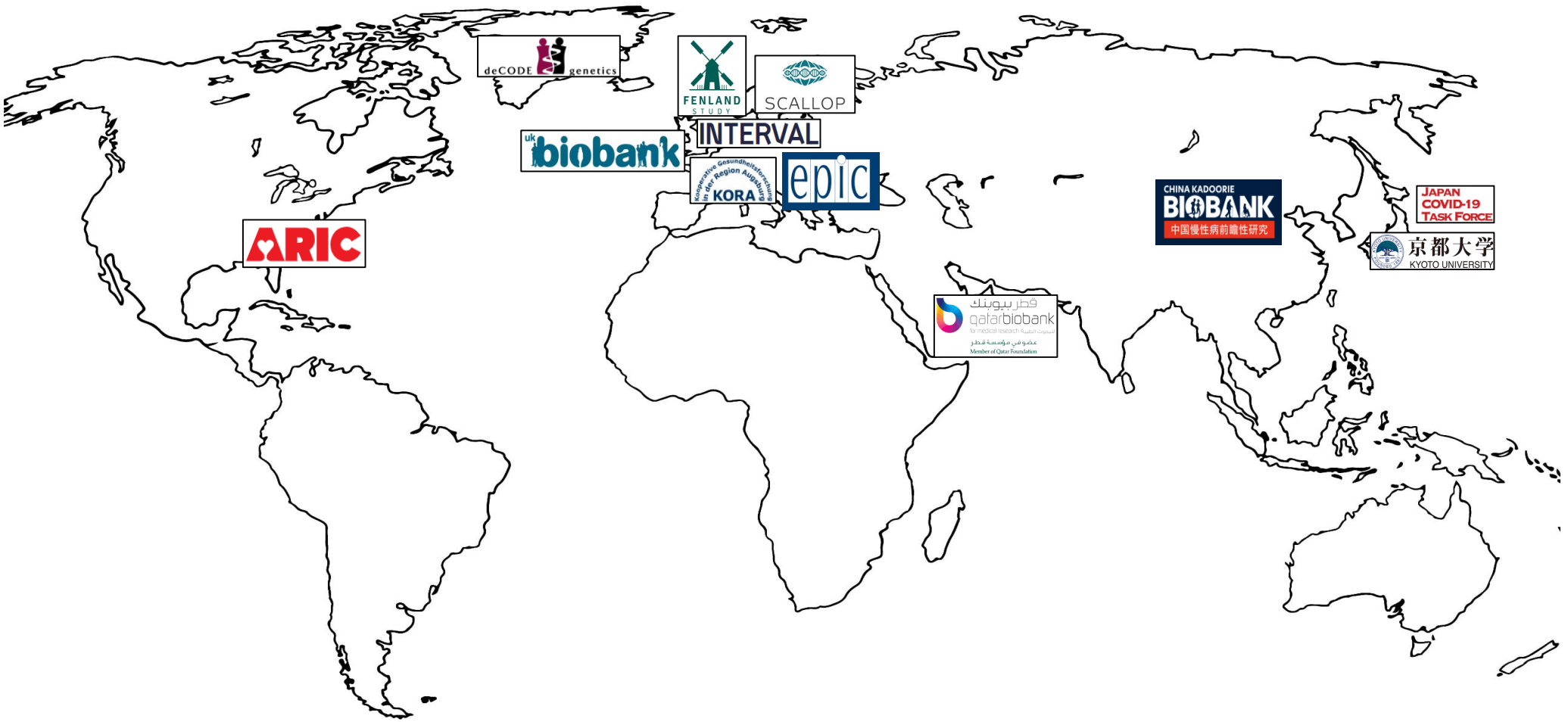
Scaling up to tens of thousands

Folkersen L	PEA/OLINK
Pietzner M	Aptamer
Ferkingstad E	Aptamer



OXFORD POPULATION HEALTH

Overview of proteomics in population biobanks



Olink Explore HT: ~5400 proteins; **SomaScan v5:** ~10,000 proteins

OXFORD POPULATION HEALTH

Characteristics of proteomics assays in biobanks

Study name	Platform	Ancestry	Sample size	No. of proteins
UK Biobank	OLINK	European African Central/South Asian Middle Eastern East Asian Admixed American	34,557 931 920 308 262 97	2,923
deCODE genetics	OLINK & SomaScan	European	35,559	4,719
Fenland	OLINK & SomaScan	European	10,708	4,775
ARIC	OLINK & SomaScan	European African American	7,213 1,871	4,657
Interval study	OLINK & SomaScan	European	3,301	3,622
EPIC-Norfolk	OLINK	European	1,180	2,923
KORA	OLINK & SomaScan	European	997	1,124
Qatar Biobank	SomaScan	Middle Eastern	2,935	1,301
Kyoto University	SomaScan	East Asian	1,823	4,196
Japan Covid-19 task force	OLINK	East Asian	1,384	2,923
China Kadoorie Biobank	**OLINK & SomaScan**	**East Asian**	**3,970**	**7,139**

CKB is in the process of securing funding to scale up to 50-100K participants

OXFORD POPULATION HEALTH

China Kadoorie Biobank (CKB)

A unique combination of scale, breadth and depth

Large

>512,000 men and women aged 35-74 years recruited in 2004-08 from 10 diverse areas across China

Deep

Extensive information on social, lifestyle, environmental, clinical measurements and biological factors, along with reliable disease characterisation and sub-phenotyping

Long

Sufficient duration of follow-up for large numbers of fatal and non-fatal health outcomes to have occurred

Accessible

Data are released periodically to the research community

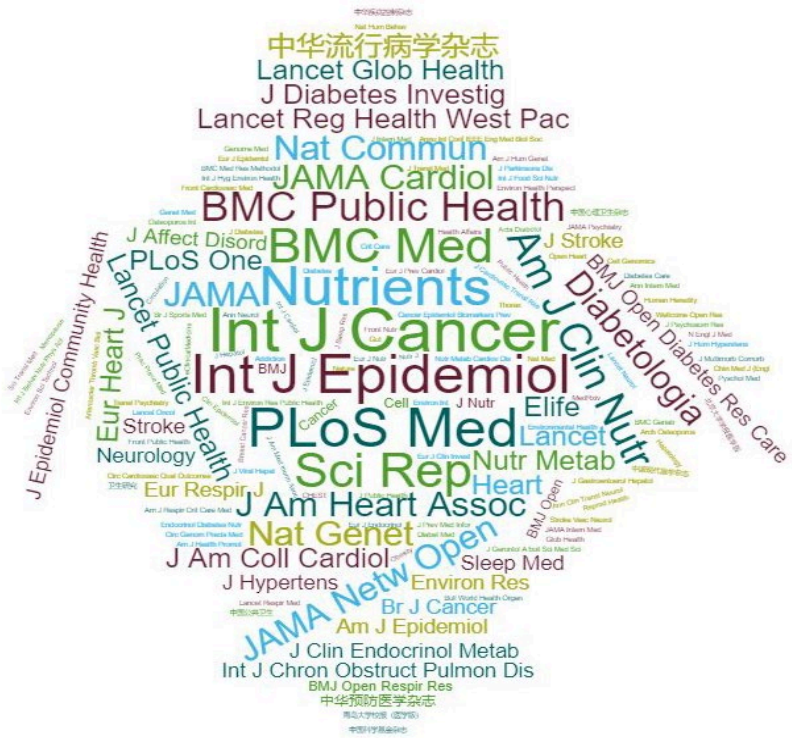
Study website: www.ckbiobank.org

CKB: A uniquely rich and high quality resource

- **Population diversity:** 10 areas, with good representativeness
- **High data quality:** no missing values, no logic errors
- **Complete sample collection:** 99.98%, with no loss, damage or thaw
- **Complete follow-up:** <1% lost to follow up since baseline survey
- **Medical record linkages:** ~100K deaths and >3.5 million ICD-10 coded episodes of hospitalisation for >5000 different disease types
- **Well phenotyped diseases:** e.g. ~95% stroke cases confirmed by brain imaging, with detailed adjudications for several major diseases
- **Three repeat assessments:** >80% responded, with new enhancements

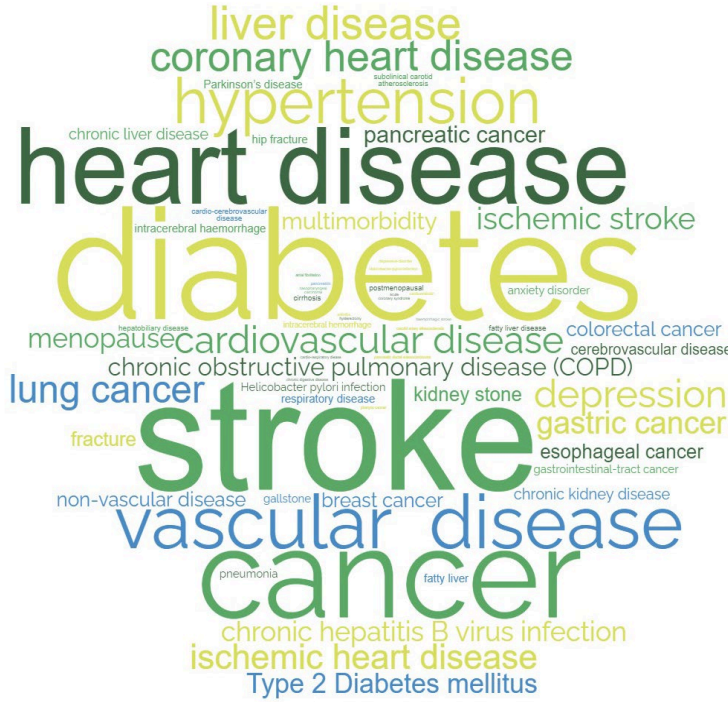
CKB: Scope of research activities and achievements

~600 papers



Risk exposures

Health outcomes

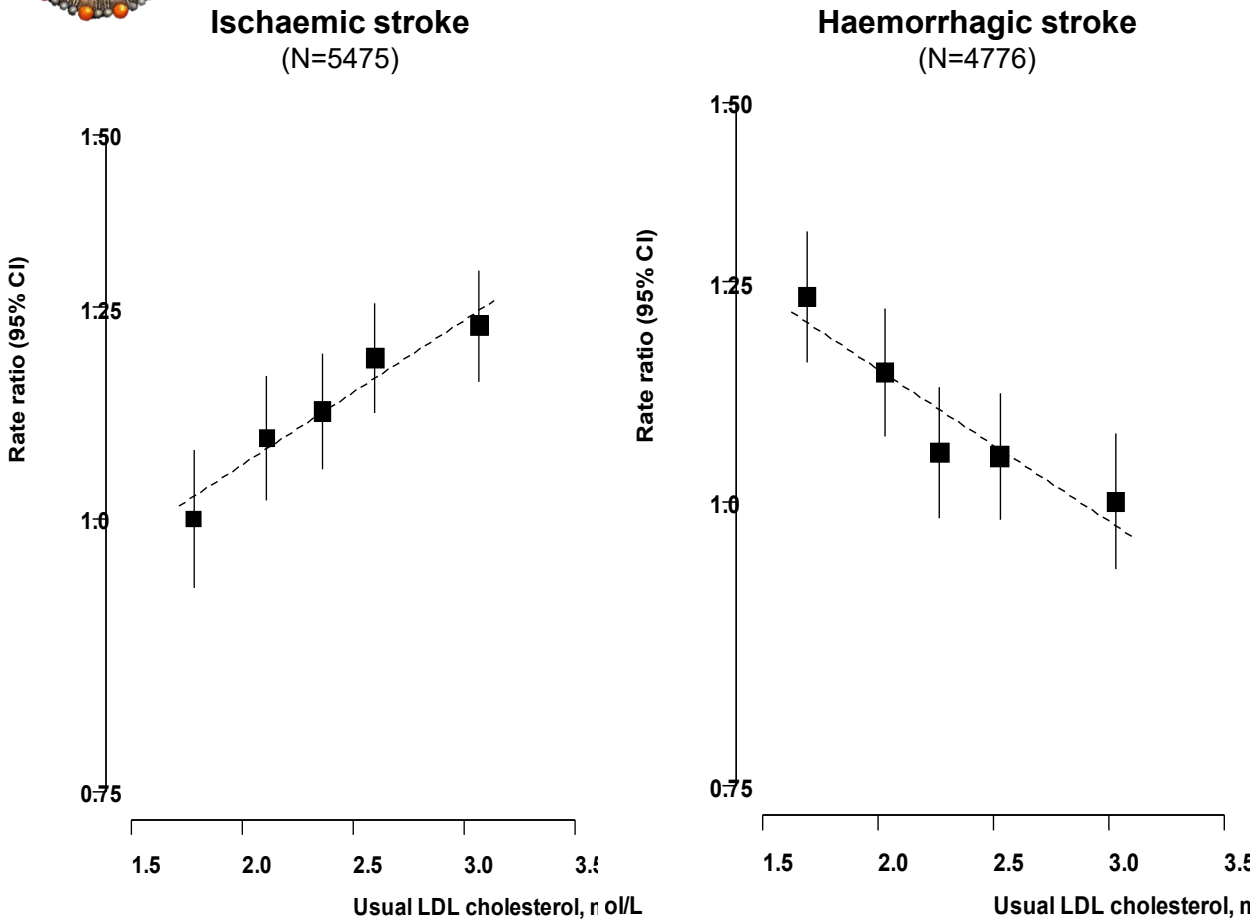


>100 papers led by external collaborators and open access researchers

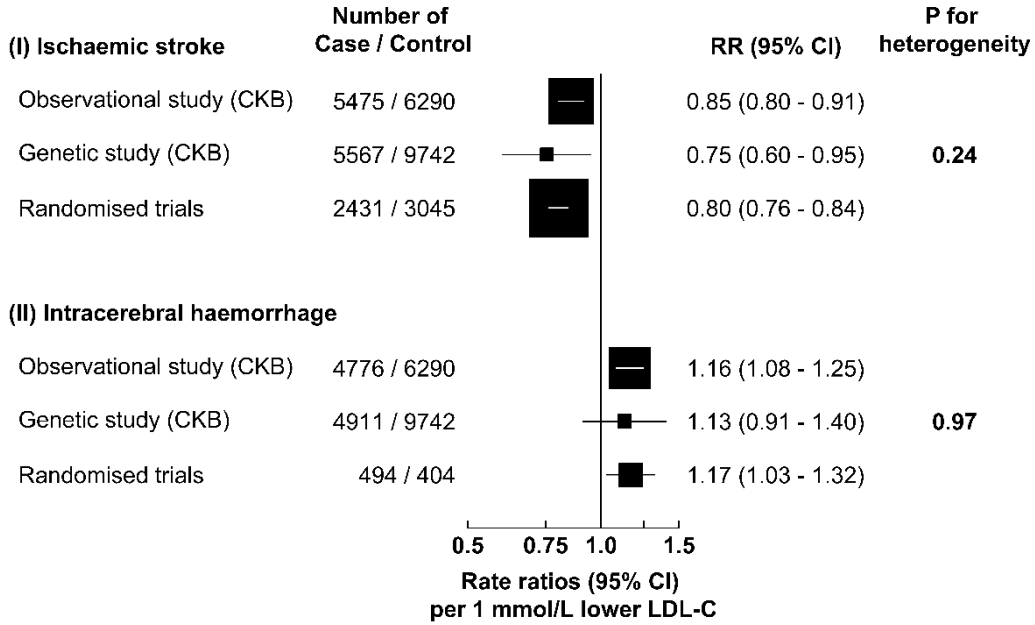
OXFORD POPULATION HEALTH

CHINA KADOORIE BIOBANK 中国慢性病前瞻性研究

CKB: LDL-cholesterol and risk of stroke types



Sun LL, et al Nature Med 2019



nature medicine
LETTERS
https://doi.org/10.1038/s41591-019-0366-x

Causal associations of blood lipids with risk of ischemic stroke and intracerebral hemorrhage in Chinese adults

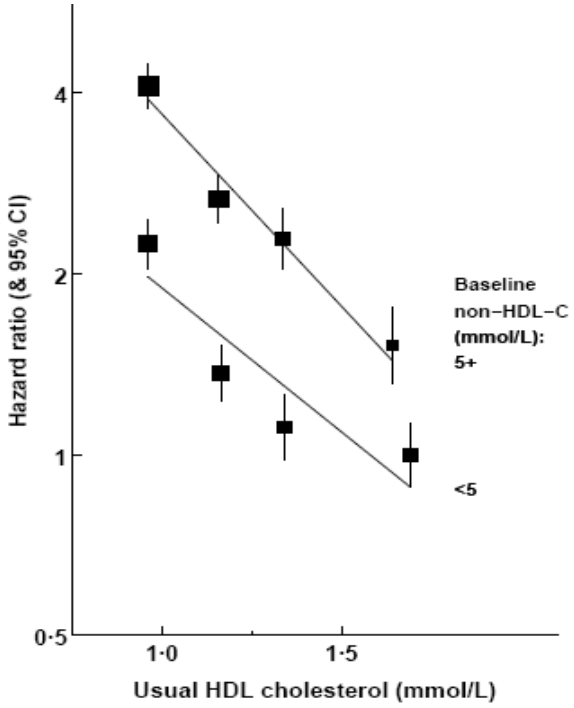
Luanluan Sun[1], Robert Clarke[1*], Derrick Bennett[1], Yu Guo[2], Robin G. Walters[3], Michael Hill[3], Sarah Parish[3], Iona Y. Millwood[3], Zheng Bian[2], Yiping Chen[3], Canqing Yu[4], Jun Lv[4], Rory Collins[1], Junshi Chen[5], Richard Peto[1], Liming Li[4], Zhengming Chen[1*], and on behalf of the China Kadoorie Biobank Collaborative Group[6]

OXFORD POPULATION HEALTH

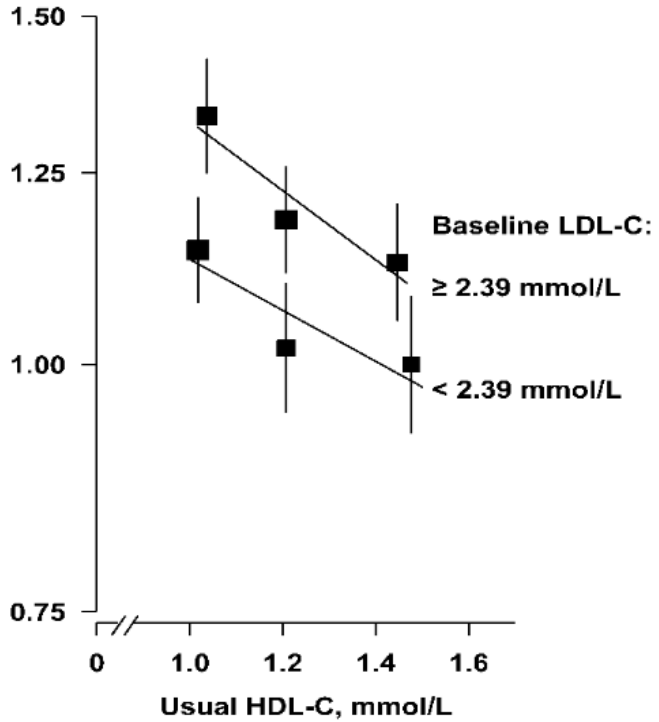
CHINA KADOORIE BIOBANK 中国慢性病前瞻性研究

CKB: HDL-cholesterol, CETP and risk of CVD

HDL-C vs IHD
(PSC)



HDL-C vs IS
(CKB)



CETP PheWAS: **Relation of *CETP* genetic defect *(*rs2303790) with CVD and other diseases**

Vascular disease - **No effect**
(CETP defect vs not: RR=1.0 (0.9-1.1)

Eye disease - **possible side effect**
(CETP defect vs not: RR=1.45 (1.13-1.80)

JAMA Cardiology | Original Investigation

Association of *CETP* Gene Variants With Risk for Vascular and Nonvascular Diseases Among Chinese Adults

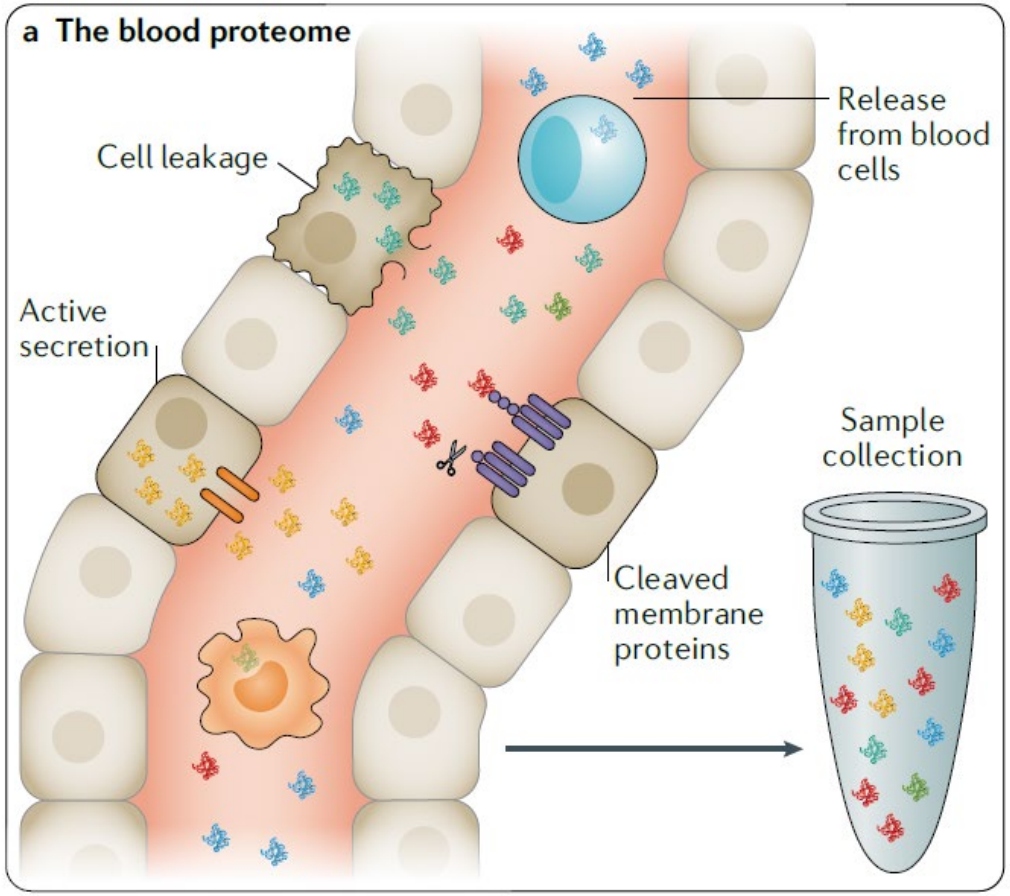
Iona Y. Millwood, DPhil; Derrick A. Bennett, PhD; Michael V. Holmes, PhD; Ruth Boxall, MSc; Yu Guo, MSc; Zheng Bian, MSc; Ling Yang, PhD; Sam Sansome, BSc; Yiping Chen, DPhil; Huaidong Du, PhD; Canqing Yu, PhD; Alex Hacker, MA; Dermot F. Reilly, PhD; Yunlong Tan, MBBS; Michael R. Hill, PhD; Junshi Chen, MD; Richard Peto, FRS; Hongbing Shen, PhD; Rory Collins, FRS; Robert Clarke, MD; Liming Li, MPH; Robin G. Walters, PhD; Zhengming Chen, DPhil; for the China Kadoorie Biobank Collaborative Group

Consistent with **trials** showing **no net effects** of HDL-C raising

OXFORD POPULATION HEALTH

CHINA KADOORIE BIOBANK 中国慢性病前瞻性研究

CKB: Proteomics Pilot Project



Proteomic assays

- ~3000 Olink & >7000 SomaScan proteins measured
- 4000 participants (2000 IHD & 2000 sub-cohort)

Key research areas (>30 active projects)

- Genetic architecture of proteins in diverse populations
- Drug target discovery (eg, IHD, DM, stroke, HF)
- Risk prediction & early diagnosis (eg, IHD, DM)
- Mechanisms linking exposure (eg, BMI, smoking) to specific diseases
- Biological aging clocks, frailty and multi-morbidity
- Comparative study of Olink-SomaScan platforms

Organisation of annual symposium on “Human Proteome and Health”

OXFORD POPULATION HEALTH

CHINA KADOORIE BIOBANK 中国慢性病前瞻性研究

CKB: >15 recently completed proteomics projects

Eur J Epidemiol 2023

GENETIC EPIDEMIOLOGY

Conventional and genetic associations of adiposity with 1463 proteins in relatively lean Chinese adults

Pang Yao[1] · Andri Iona[1] · Christiana Kartsonaki[1,2] · Saredo Said[1] · Neil Wright[1] · Kuang Lin[1] · Alfred Pozarickij[1] · Iona Millwood[1,2] · Hannah Fry[1,2] · Mohsen Mazidi[1] · Yiping Chen[1,2] · Huaidong Du[1,2] · Derrick Bennett[1,2] · Daniel Avery[1,2] · Dan Schmidt[1,2] · Pei Pei[3] · Jun Lv[3,4] · Canqing Yu[3,4] · Michael Hill[1] · Junshi Chen[5] · Richard Peto[1] · Robin Walters[1,2] · Rory Collins[1] · Liming Li[3,4] · Robert Clarke[1] · Zhengming Chen[1,2] on behalf of China Kadoorie

JACC 2023

Plasma Proteomics to Identify Drug Targets for Ischemic Heart Disease

Mohsen Mazidi, PhD,[a,*] Neil Wright, PhD,[a,*] Pang Yao, PhD,[a] Christiana Kartsonaki, DPhil,[a,b] Iona Y. Millwood, DPhil,[a,b] Hannah Fry, BSc,[a,b] Saredo Said, PhD,[a] Alfred Pozarickij, PhD,[a] Pei Pei, MSc,[c] Yiping Chen, DPhil,[a,b] Daniel Avery, MSc,[a] Huaidong Du, PhD,[a,b] Dan Valle Schmidt, MSc,[a] Ling Yang, PhD,[a,b] Jun Lv, PhD,[c,d,e] Canqing Yu, PhD,[c,d,e] Junshi Chen, MD,[f] Michael Hill, DPhil,[a,b] Michael V. Holmes, PhD,[a] Joanna M.M. Howson, PhD,[g] Richard Peto, MD,[a] Rory Collins, MBBS,[a] Derrick A. Bennett, PhD,[a,b] Robin G. Walters, PhD,[a,b] Liming Li, MPH,[c,d,e] Robert Clarke, PhD,[a] Zhengming Chen, DPhil,[a,b]

Diabetes Care 2024

Proteomic analyses in diverse populations improved risk prediction and identified new drug targets for diabetes

Pang Yao[1], Andri Iona[1], Alfred Pozarickij[1], Saredo Said[1], Neil Wright[1], Kuang Lin[1], Iona Millwood[1,2], Hannah Fry[1,2], Christiana Kartsonaki[1,2], Mohsen Mazidi[1], Yiping Chen[1,2], Bowen Liu[1], Ling Yang[1,2], Daniel Avery[1,2], Dan Schmidt[1,2], Dianjianyi Sun,[3,4] Pei Pei[3], Jun Lv[3,4], Canqing Yu[3,4], Michael Hill[1], Derrick Bennett[1,2], Robin Walters[1,2], Liming Li[3,4], Robert Clarke[1], Huaidong Du[1,2], Zhengming Chen[1,2] on behalf of China Kadoorie Biobank

Nature Med 2024

Proteomic aging clock predicts mortality and risk of common age-related diseases in diverse populations

M. Austin Argentieri[1,2], Sihao Xiao[1], Derrick Bennett[1], Laura Winchester[3], Alejo J. Nevado-Holgado[3], Ashwag Albukhari[4], Pang Yao[1], Mohsen Mazidi[1], Jun Lv[5], Liming Li[6], Cassandra J. Adams[7], Robert Clarke[1], Najaf Amin[1], Zhengming Chen[1*], Cornelia M. van Duijn[1*]

Eur J Epidemiol 2024

Risk prediction of ischemic heart disease using plasma proteomics, conventional risk factors and polygenic scores in Chinese and European adults

Mohsen Mazidi[1] · Neil Wright[1] · Pang Yao[1] · Christiana Kartsonaki[1] · Iona Y. Millwood[1] · Hannah Fry[1] · Saredo Said[1] · Alfred Pozarickij[1] · Pei Pei[2] · Yiping Chen[1] · Baihan Wang[1] · Daniel Avery[1] · Huaidong Du[1] · Dan Valle Schmidt[1] · Ling Yang[1] · Jun Lv[2,3,4] · Canqing Yu[2,3,4] · DianJianYi Sun[2,3,4] · Junshi Chen[5] · Michael Hill[1] · Richard Peto[1] · Rory Collins[1] · Derrick A. Bennett[1] · Robin G. Walters[1] · Liming Li[2,3,4] · Robert Clarke[1] · Zhengming Chen[1] on behalf of China Kadoorie Biobank Collaborative Group

Commun Biology 2024

Conventional and genetic associations of general and central adiposity with 2944 proteins in relatively lean Chinese adults

Andri Iona[1*], Pang Yao[1*], Alfred Pozarickij[1], Saredo Said[1], Neil Wright[1], Kuang Lin[1], Iona Millwood[1,2], Hannah Fry[1,2], Christiana Kartsonaki[1,2], Mohsen Mazidi[1], Yiping Chen[1,2], Fiona Bragg[1], Bowen Liu[1], Ling Yang[1,2], Daniel Avery[1,2], Dan Schmidt[1,2], Dianjianyi Sun,[3,4,5] Pei Pei[3], Jun Lv[3,4,5], Canqing Yu[3,4,5], Michael Hill[1], Derrick Bennett[1,2], Robin Walters[1,2], Liming Li[3,4,5], Robert Clarke[1], Huaidong Du[1,2], Zhengming Chen[1,2] on behalf of China Kadoorie Biobank Collaborative Group ‡

Nature Commun 2024 (accepted)

Comparative studies of genetic and phenotypic associations for 2,168 plasma proteins measured by two affinity-based platforms in 4,000 Chinese adults

Baihan Wang[1], Alfred Pozarickij[1], Mohsen Mazidi[1], Neil Wright[1], Pang Yao[1], Saredo Said[1], Andri Iona[1], Christiana Kartsonaki[1,2], Hannah Fry[1,2], Kuang Lin[1], Yiping Chen[1,2], Huaidong Du[1,2], Daniel Avery[1,2], Dan Valle Schmidt[1,2], Canqing Yu[3,4,5], Dianjianyi Sun[3,4,5], Jun Lv[3,4,5], Michael Hill[1], Liming Li[3,4,5], Derrick A Bennett[1,2], Rory Collins[1], Robin G Walters[1,2], Robert Clarke[1], Iona Y Millwood[1,2], Zhengming Chen[1,2], on behalf of China Kadoorie Biobank

Nature Genetics (under revision)

Ancestry diversity in the genetic determinants of the human plasma proteome and associated new drug targets

Saredo Said[1], Alfred Pozarickij[1], Kuang Lin[1], Sam Morris[1], Christiana Kartsonaki[1,2], Neil Wright[1], Hannah Fry[1,2], Yiping Chen[1,2], Huaidong Du[1,2], Derrick Bennett[1,2], Daniel Avery[1,2], Dan Valle Schmidt[1,2], Liming Li[3,4,5], Jun Lv[3,4,5], Canqing Yu[3,4,5], Dianjianyi Sun[3,4,5], Pei Pei[4], Junshi Chen[6], Michael Hill[1], Richard Peto[1], Rory Collins[1], Robert Clarke[1], Iona Y Millwood[1,2], Zhengming Chen[1,2], Robin G Walters[1,2], on behalf of China Kadoorie Biobank

OXFORD POPULATION HEALTH

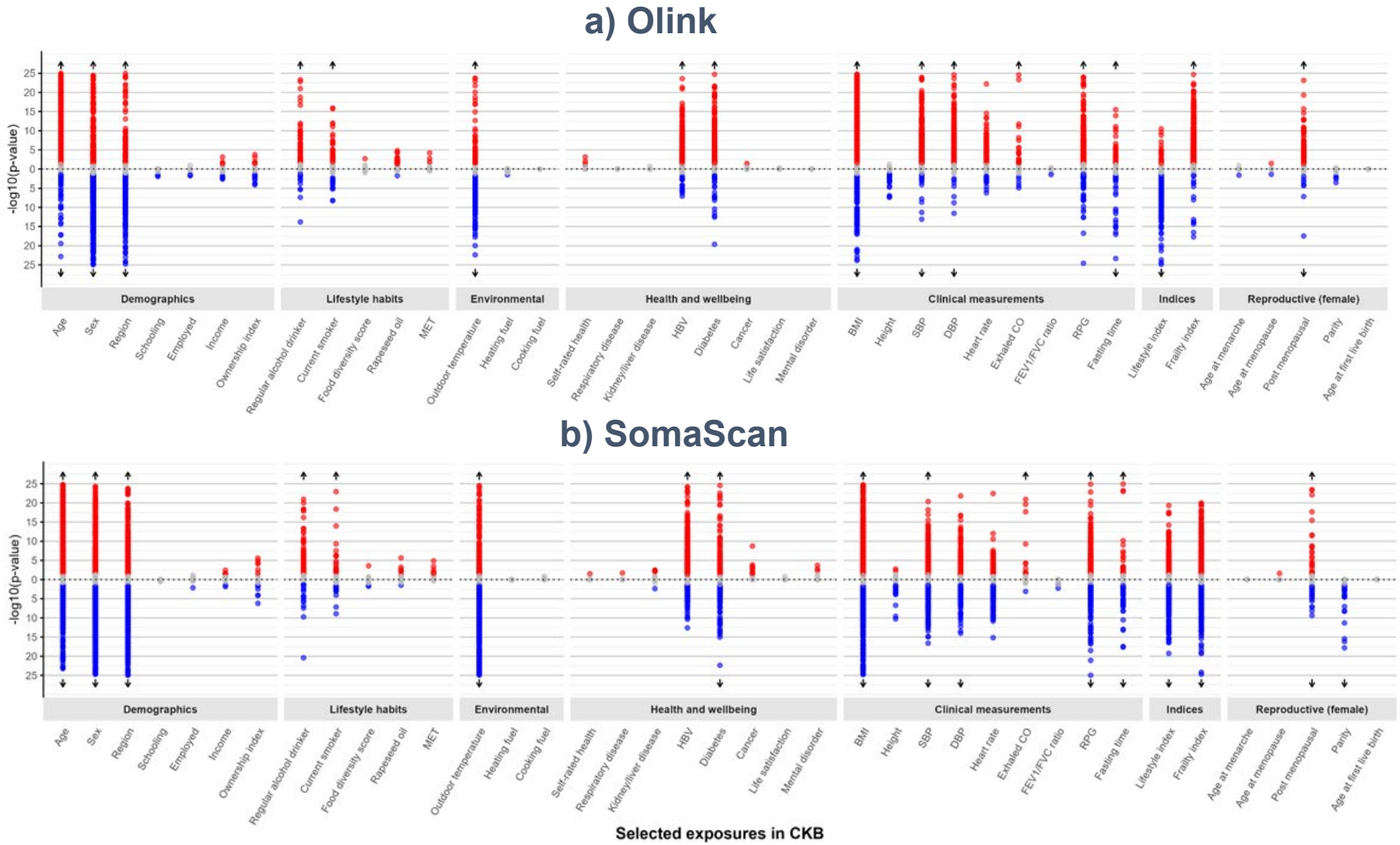
CHINA KADOORIE BIOBANK 中国慢性病前瞻性研究

CKB: Key proteomics-related research findings

- Levels of plasma proteins affected by genetic, non-genetic, as well as environmental factors
- Proteins greatly outperformed PGS for prediction of many diseases
- In CKB, ~50% of pQTLs signals are distinct from those in Europeans
- Genetic (MR, colocalisation) analyses helped establish causality and identify potential drug targets for IHD, stroke, T2D, HF and obesity
- Protein-based ageing clocks reliably predicted age-related traits and risks of multiple diseases and all-cause mortality
- Complementarity of the two affinity-based assay platforms

Using conventional, genetic, ML and downstream analyses, with replication and/or combined analyses in UKB and other studies

CKB: Exposome profiles of Olink and SomaScan proteins

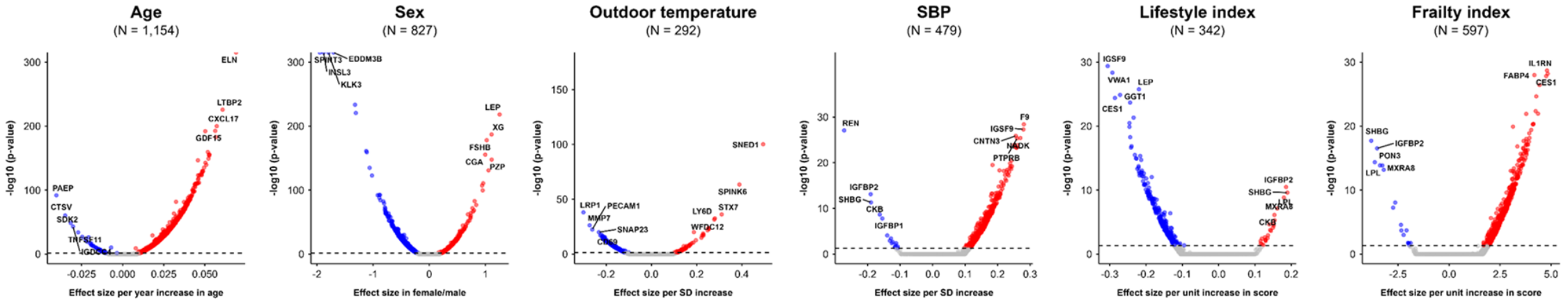


Source: Chan, K et al doi: https://doi.org/10.1101/2024.10.24.24316041 (pre-print)

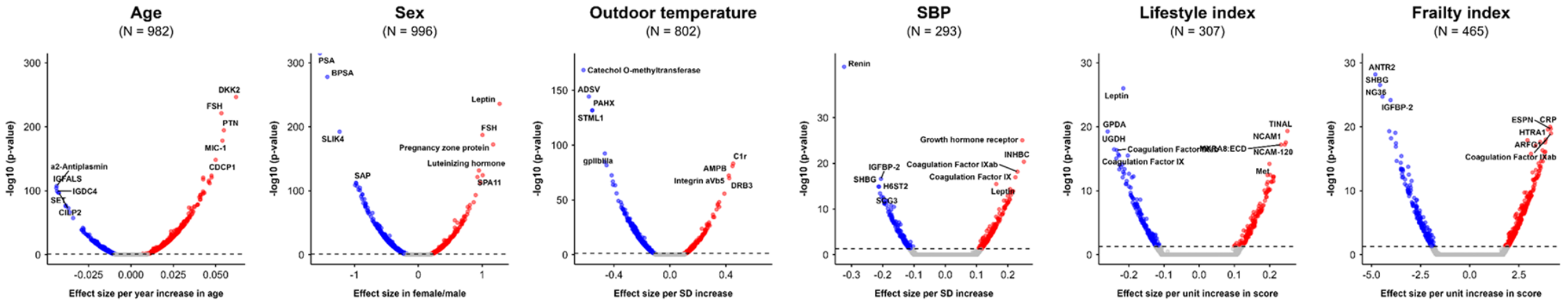
Iona, A et al doi: https://doi.org/10.1101/2024.10.23.24315975 (pre-print)

CKB: Selected exposome characteristics by panel

Olink



SomaScan



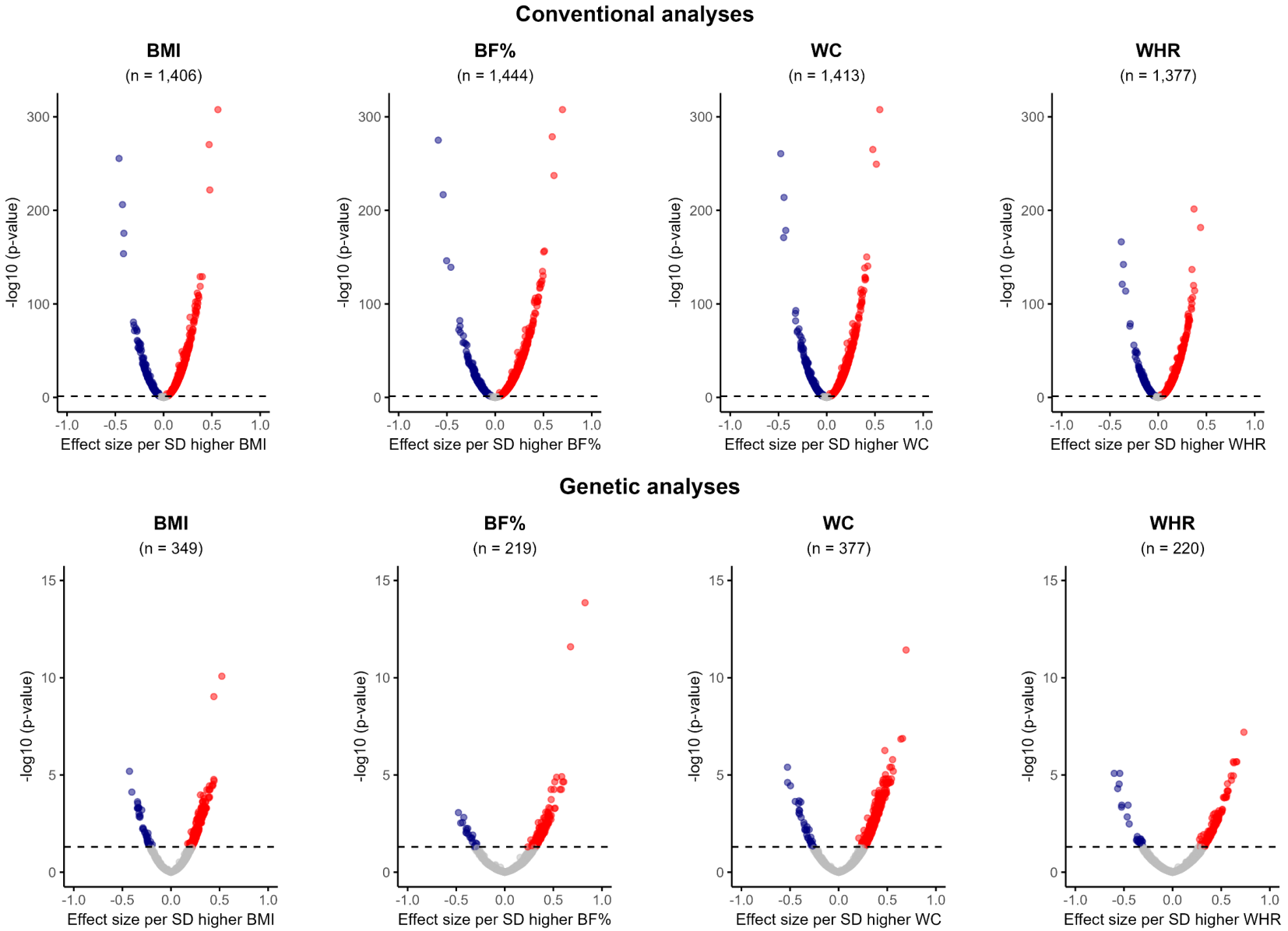
OXFORD POPULATION HEALTH

Source: Chan, K et al doi: https://doi.org/10.1101/2024.10.24.24316041 (pre-print)

Iona, A et al doi: https://doi.org/10.1101/2024.10.23.24315975 (pre-print)

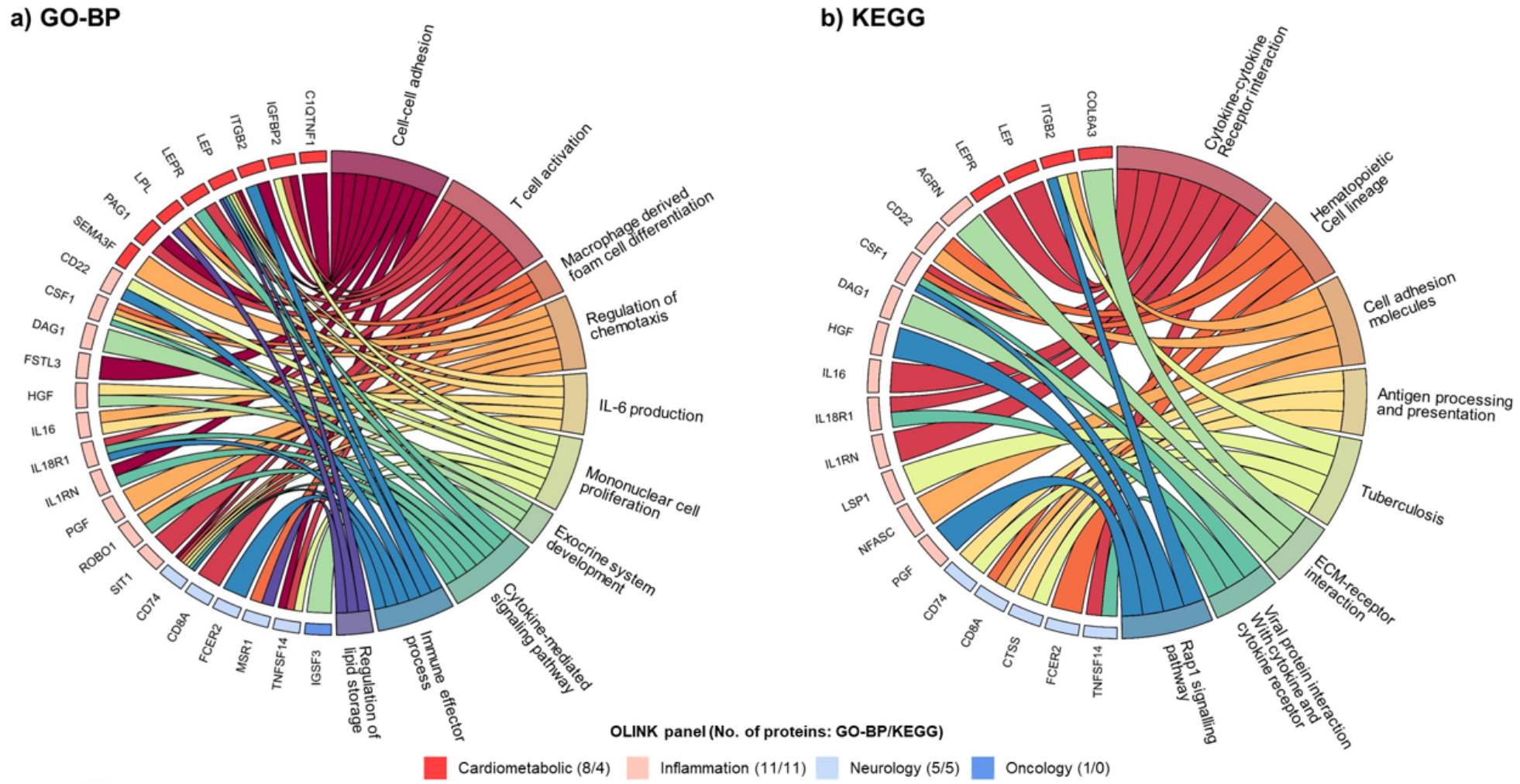
CHINA KADOORIE BIOBANK 中国慢性病前瞻性研究

CKB: Adiposity and protein levels, Olink panel



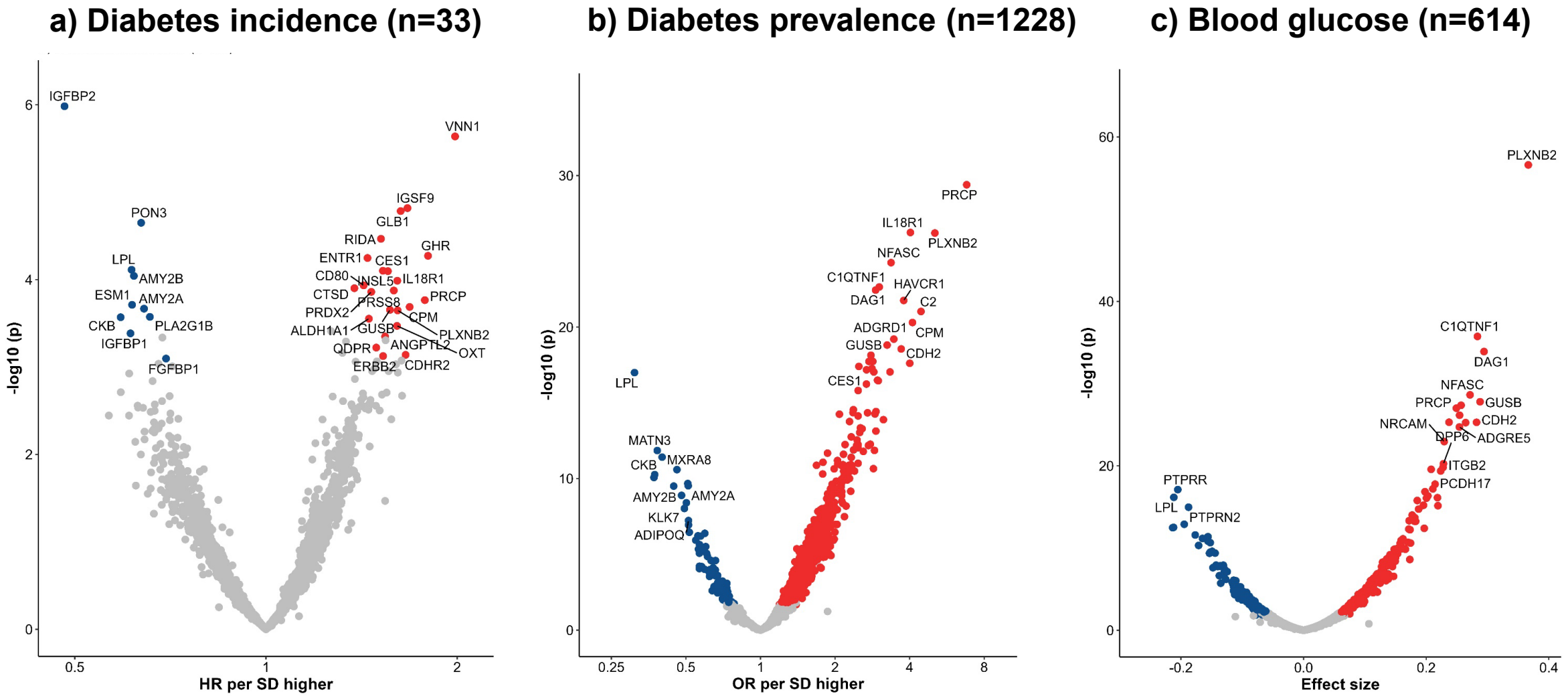
Iona A, Yao P, et al. *Commun Biol* 2024; 1327

CKB: Selected functions of top BMI-related proteins



P Yao, et al. Eur J Epidemiol 2023

CKB: Olink proteins significantly associated with T2DM



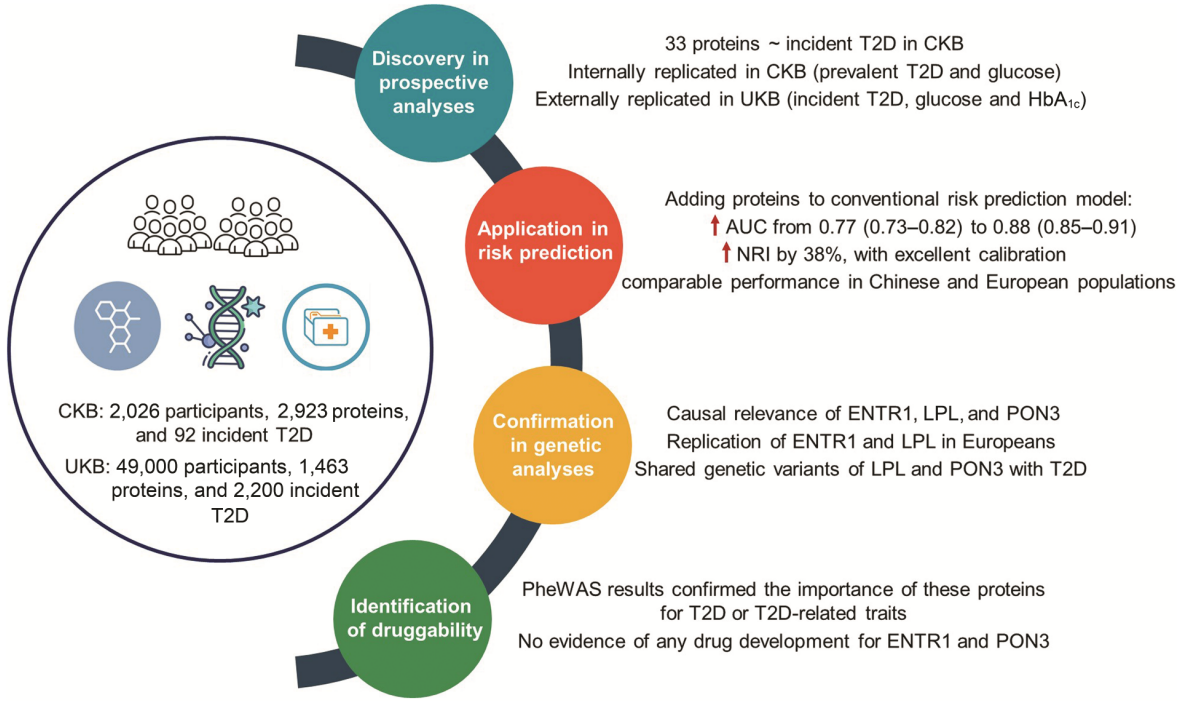
With great overlap of **33 proteins** across three outcome measures

CKB: Discovery of novel drug targets for T2D

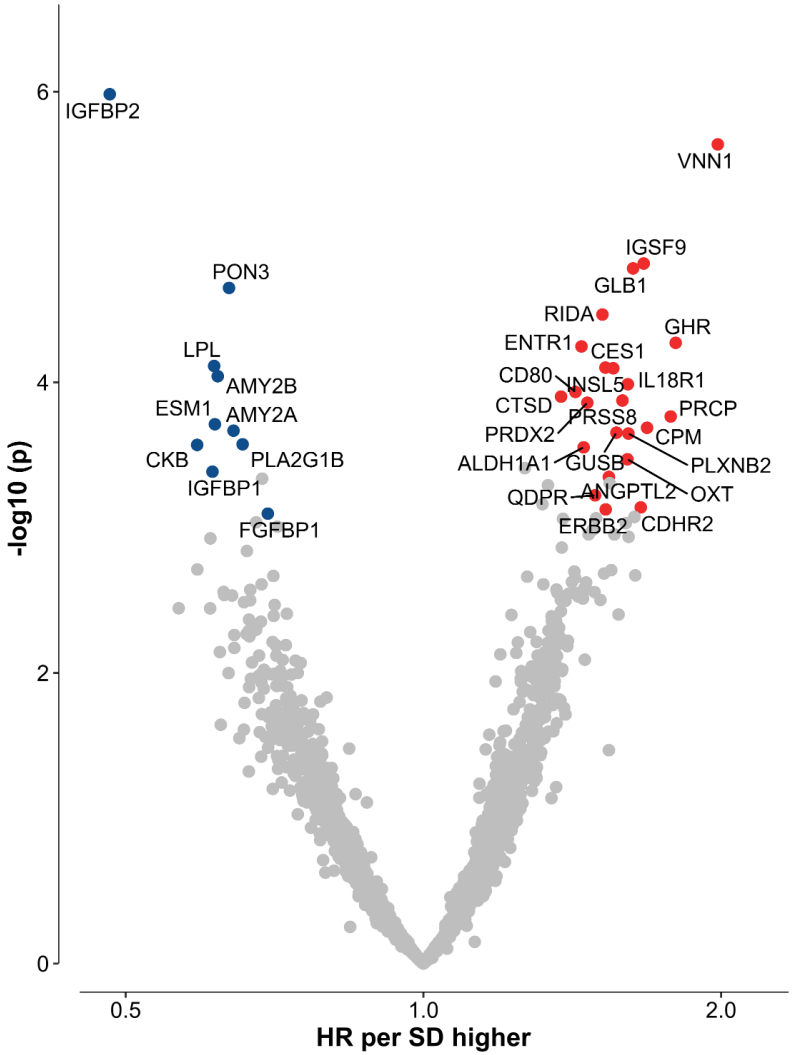
Diabetes Care
American Diabetes Association

Proteomic Analyses in Diverse Populations Improved Risk Prediction and Identified New Drug Targets for Type 2 Diabetes

Pang Yao, Andri Iona, Alfred Pozarickij, Saredo Said, Neil Wright, Kuang Lin, Iona Millwood, Hannah Fry, Christiana Kartsonaki, Mohsen Mazidi, Yiping Chen, Fiona Bragg, Bowen Liu, Ling Yang, Junxi Liu, Daniel Avery,

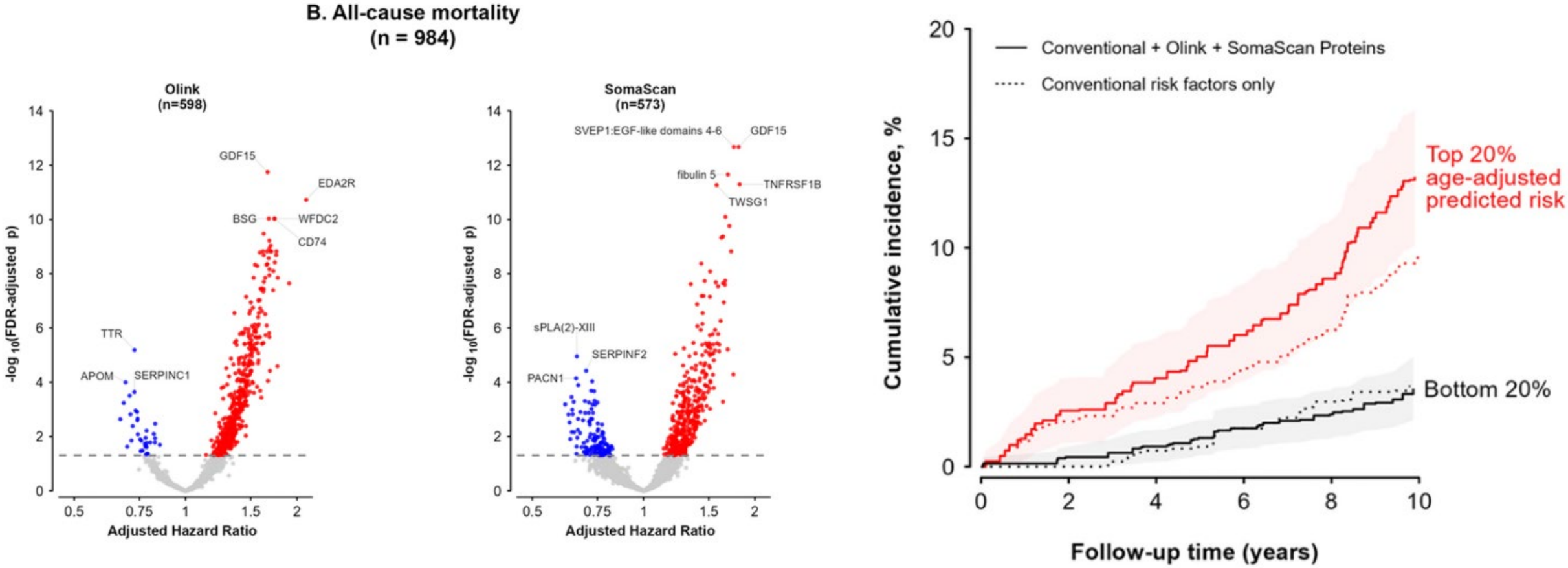


AUC, area under the curve; CKB, China Kadoorie Biobank; LPL, lipoprotein lipase; NRI, net reclassification index; PheWAS, phenome-wide association study; UKB, UK Biobank.



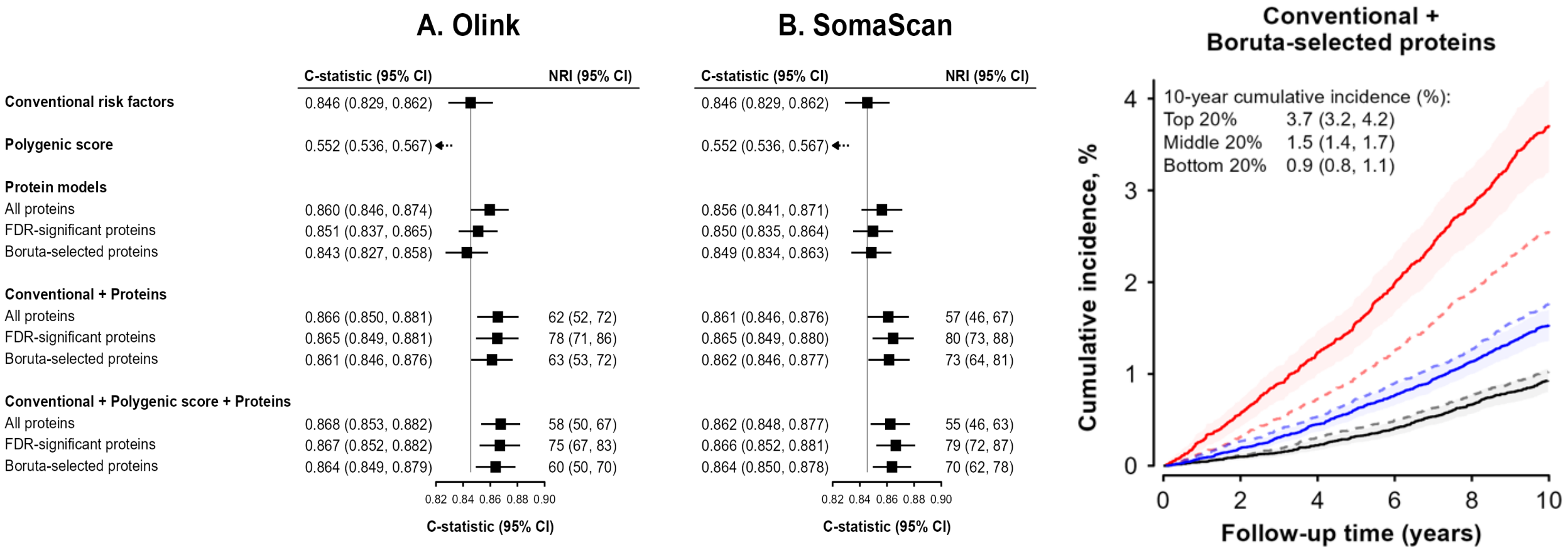
CKB: Proteomics improve risk prediction of all-cause mortality

(analyses restricted to 2000 subcohort participants)



Bennett D, et al. *Nat Ageing* (under review)

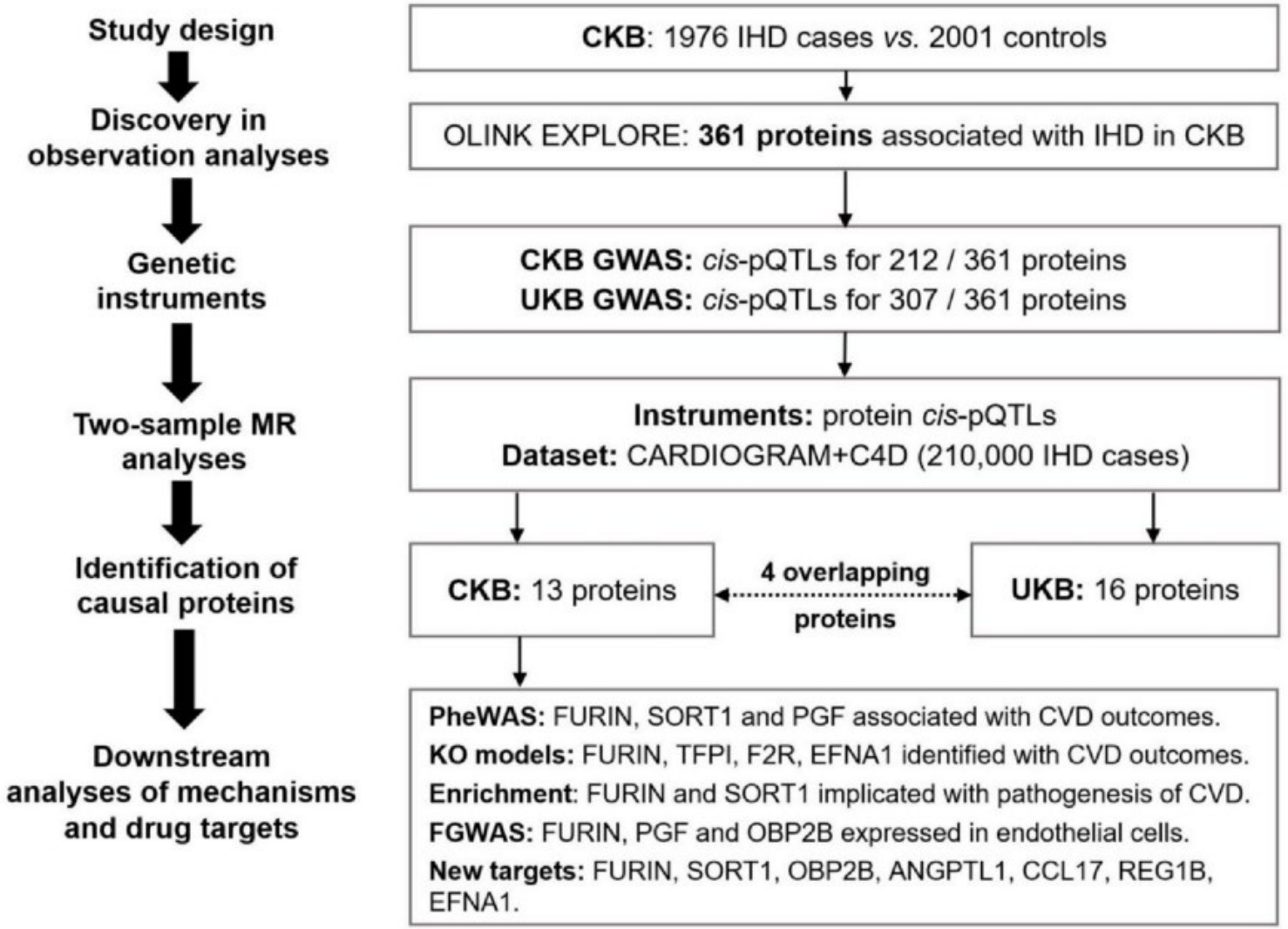
CKB: Proteomics improve risk prediction of IHD



For Olink proteins, the associations were externally replicated in the UKB

Mazidi M, et al. *Eur J Epidemiol* 2024

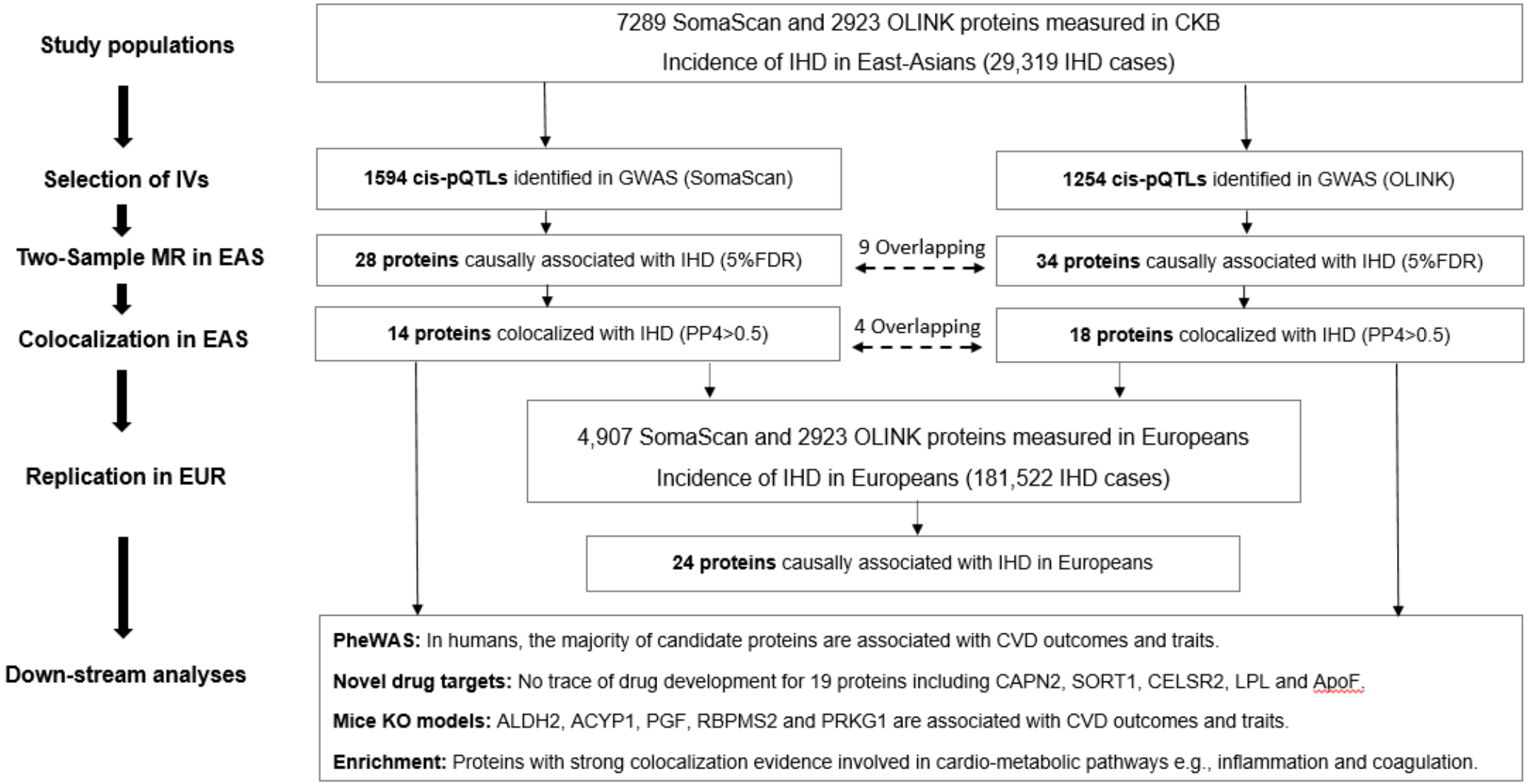
CKB: Discovery of drug targets for IHD, Olink panel



FURIN
ASGRI
MMP3
F2R

Mazidi M, et al. JACC 2023

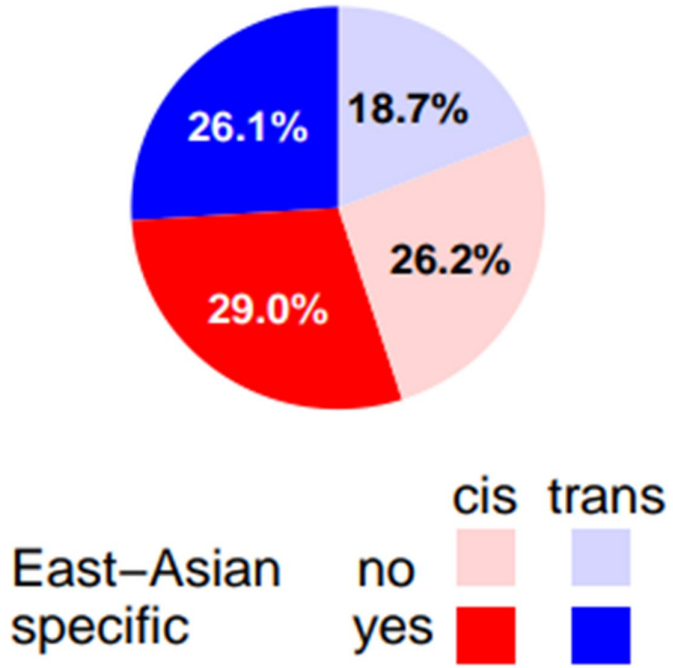
CKB: Assessing causal associations of proteins with IHD



Manuscript under preparation

CKB: Distinct pQTLs identified in East Asians

- No *cis*-pQTLs identified in UKB for ALDH2 and GLIPR1
- Lead variant within the structural gene in CKB but not UKB for 155 other proteins
- >50% of pQTL signals appear distinct from those in UKB
  - 18% were excluded in UKB due to low MAF
  - Potentially included a more common variant (D' > 0.95)



UKB / CKB： Protein-based biological ageing clock

Nature Med 2024

Article

https://doi.org/10.1038/s41591-024-03164-7

Proteomic aging clock predicts mortality and risk of common age-related diseases in diverse populations

Received: 14 September 2023

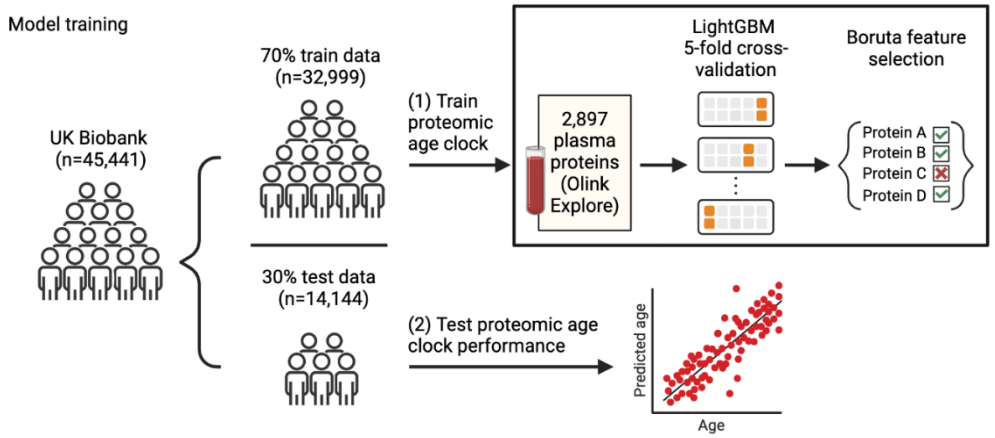
Accepted: 27 June 2024

Published online: 08 August 2024

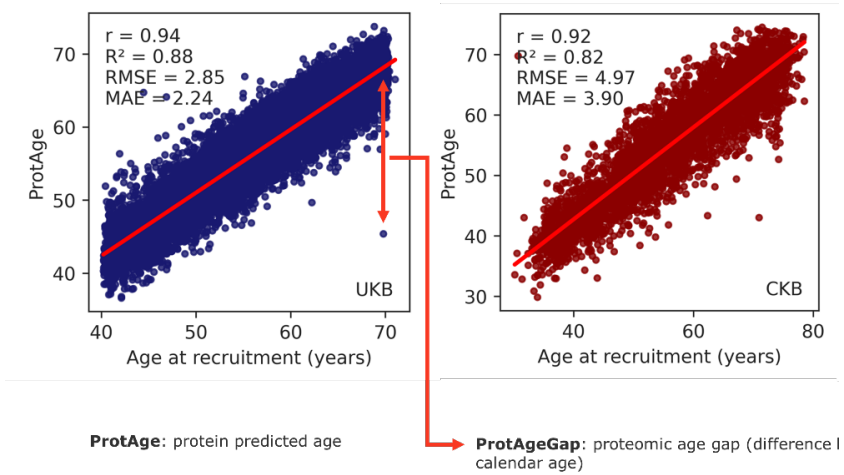
A list of authors and their affiliations appears at the end of the paper

Circulating plasma proteins play key roles in human health and can

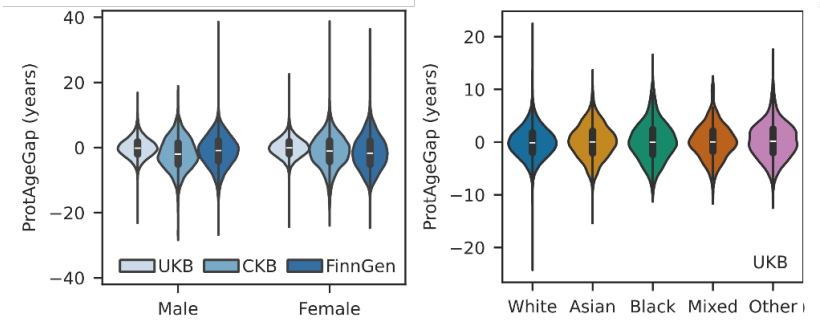
1) Main analytic approaches



2) Proteomics ageing clocks

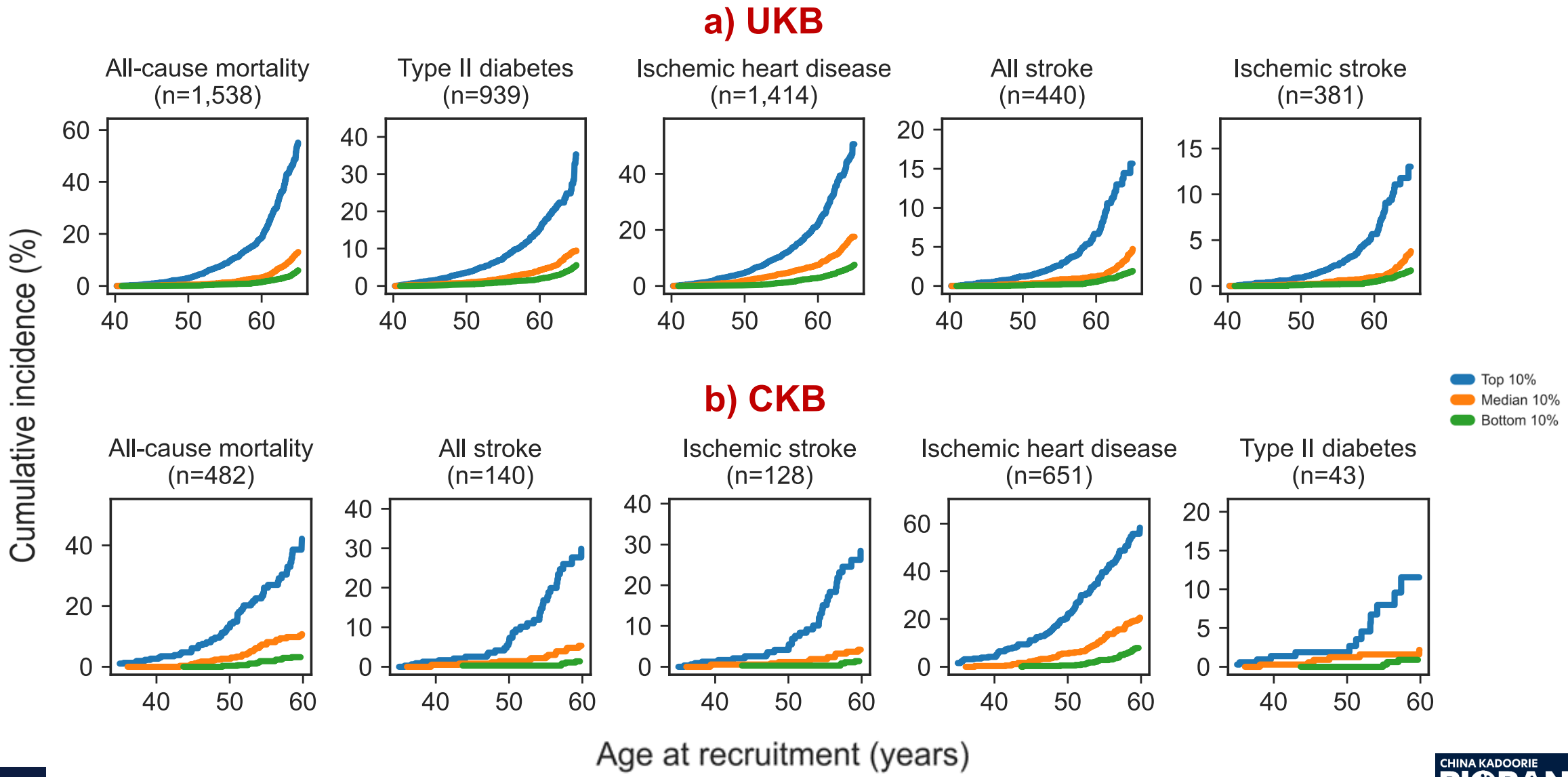


3) ProtAgeGap distributions



ProtAgeGap can predict >20 age-related traits (eg, bone density, renal function, telomere length)

Protein ageing clock can predict multiple disease risks



OXFORD POPULATION HEALTH

CHINA KADOORIE BIOBANK 中国慢性病前瞻性研究

Promises of proteomic profiling in biobanks

(in combination with genetic and other phenotypic data)

- Improve understanding of human **biology** and disease **aetiology**
- Identify **novel** pathways representing potential **drug targets**
- Enhance **risk prediction** and **early diagnosis** of diseases
- Reveal **mechanisms** linking lifestyles and traits with diseases
- Clarify **molecular basis** of genetic variants associated with disease
- Improve patient **stratification**, incl. **drug response** prediction, for more targeted management

KEY NEEDS: Turn samples into data

OXFORD POPULATION HEALTH

CHINA KADOORIE BIOBANK 中国慢性病前瞻性研究

CKB Research Partners and Funders

嘉道理 KADOORIE

MRC Clinical Trial Service Unit

British Heart Foundation

CANCER RESEARCH UK

NSFC

wellcome trust

BAYER

novo nordisk®

Novo Nordisk Logo PNG

中华人民共和国科学技术部 Science and Technology of the People's Rep

gsk

MERCK