HDRUK
Health Data Research UK

Molecular science and electronic health records: what does the future hold?

Dr Adam Butterworth
HDRUK Cambridge
23rd June 2021

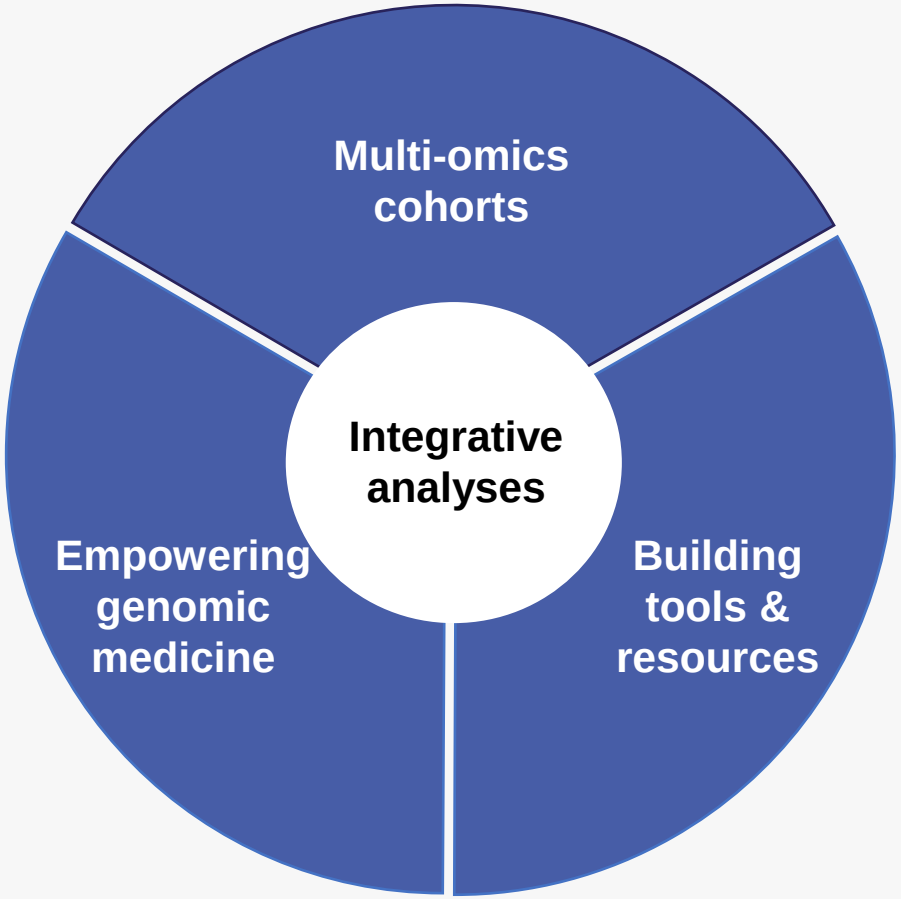
The challenge and opportunity

HDRUK
Health Data Research UK

Bridging the gap between genotype and phenotype by linking multiple layers of molecular and other data with e-health records

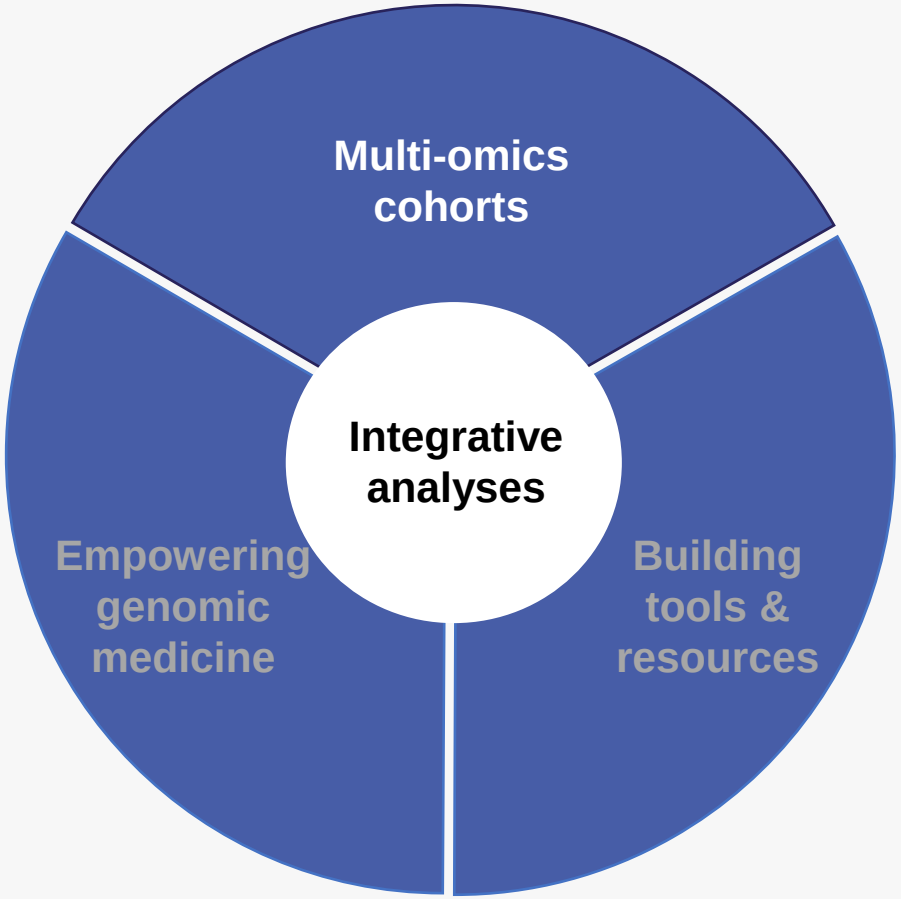
Understanding Causes of Disease: from molecules to electronic health records

HDRUK
Health Data Research UK



Understanding Causes of Disease: from molecules to electronic health records

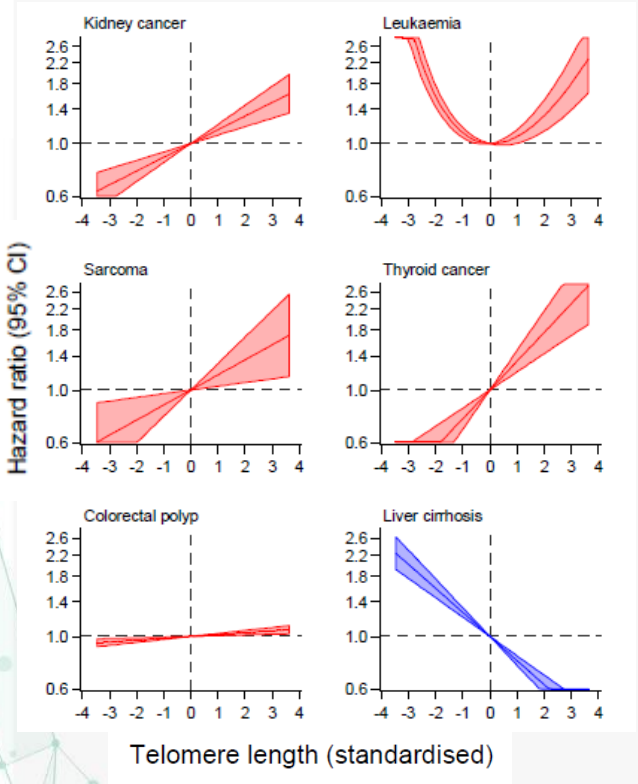
HDRUK
Health Data Research UK



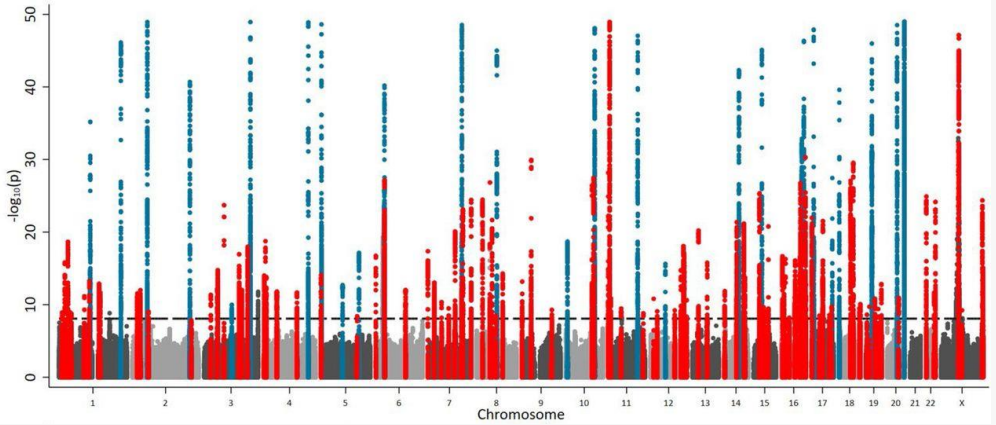
Integrating genomes, health records & telomere length in 500,000 people

HDRUK
Health Data Research UK

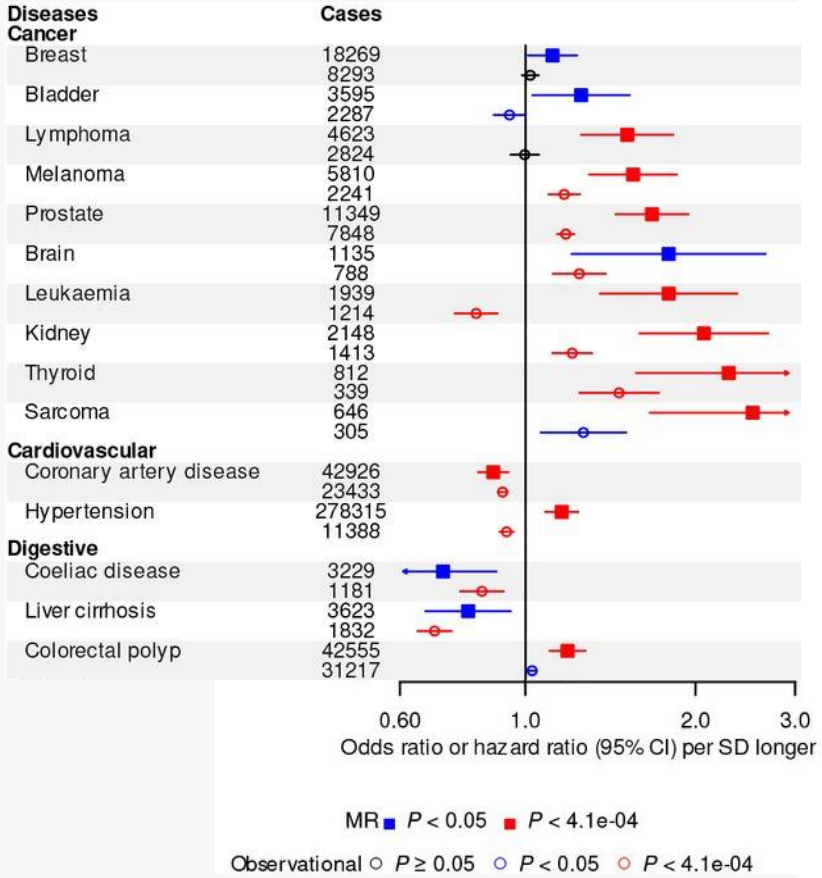
Observational associations with 123 diseases



197 genetic signals for telomere length



Causal associations suggested for 31 diseases

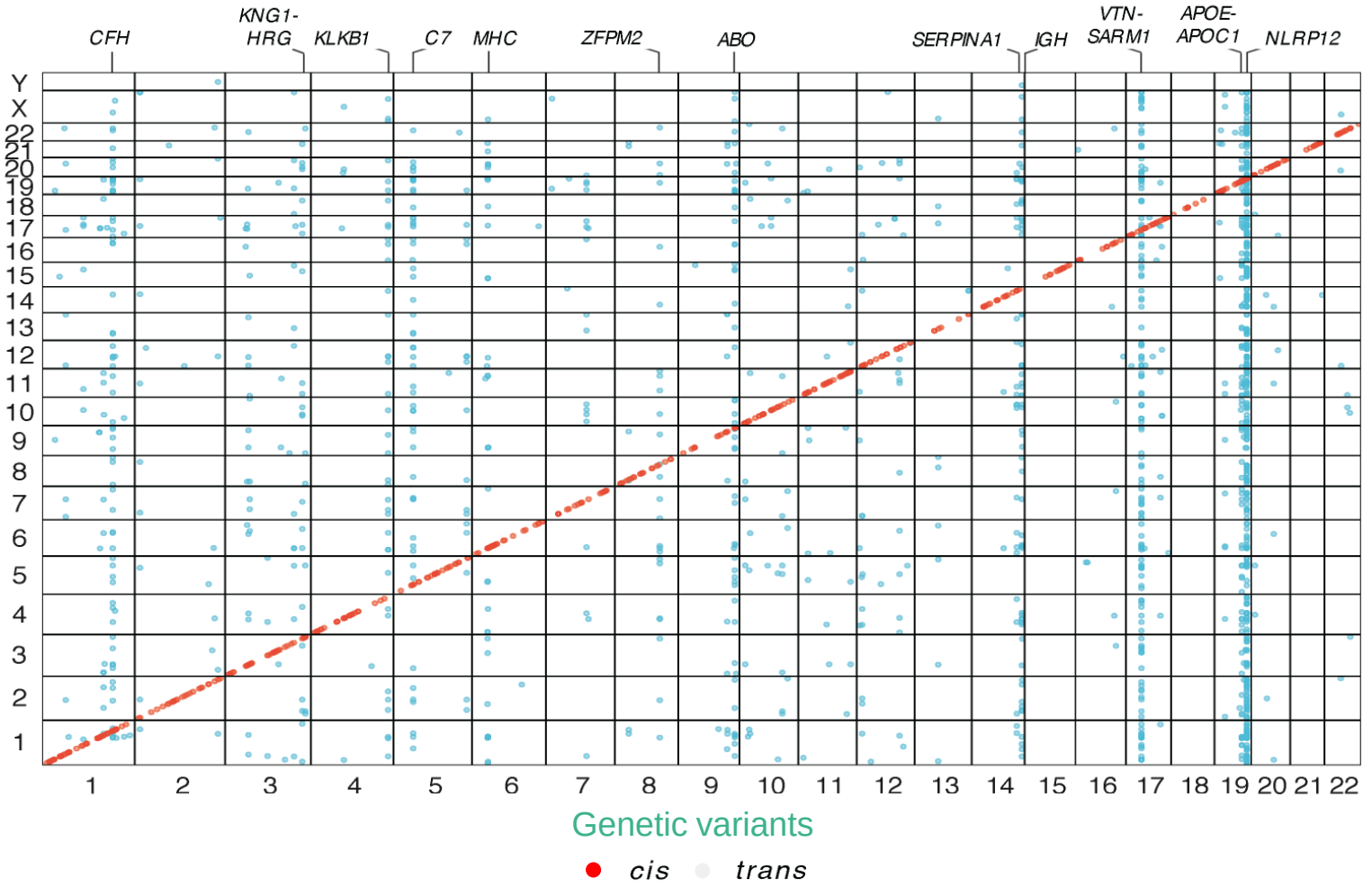


Codd *et al, Nature Genetics, in press*

An atlas of ~2000 genotype-protein associations

HDRUK
Health Data Research UK

Genes encoding target proteins

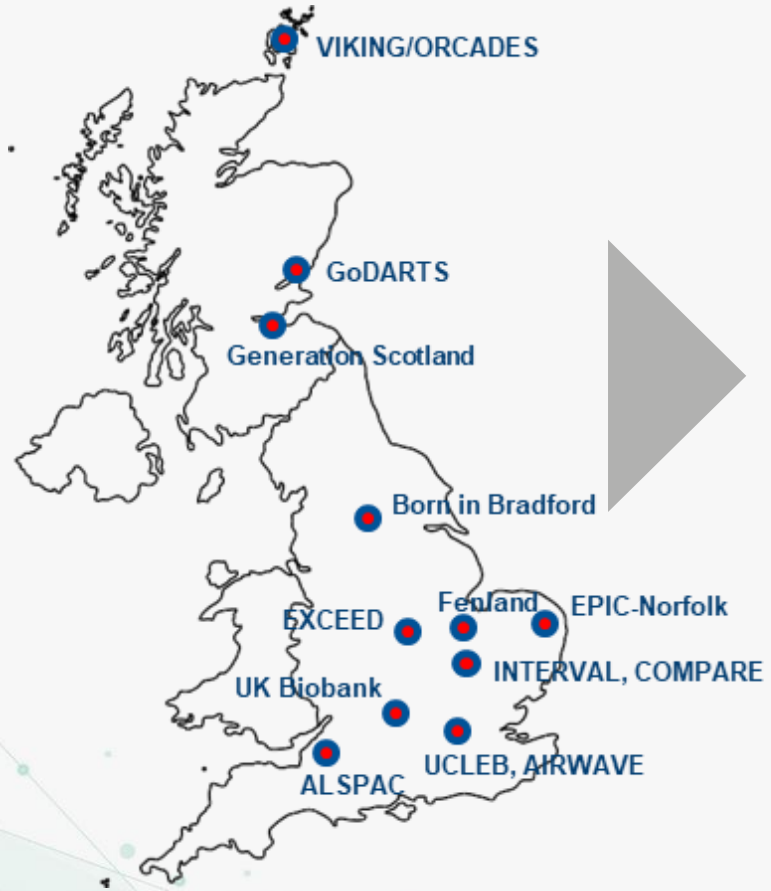


Sun *et al, Nature* 2018

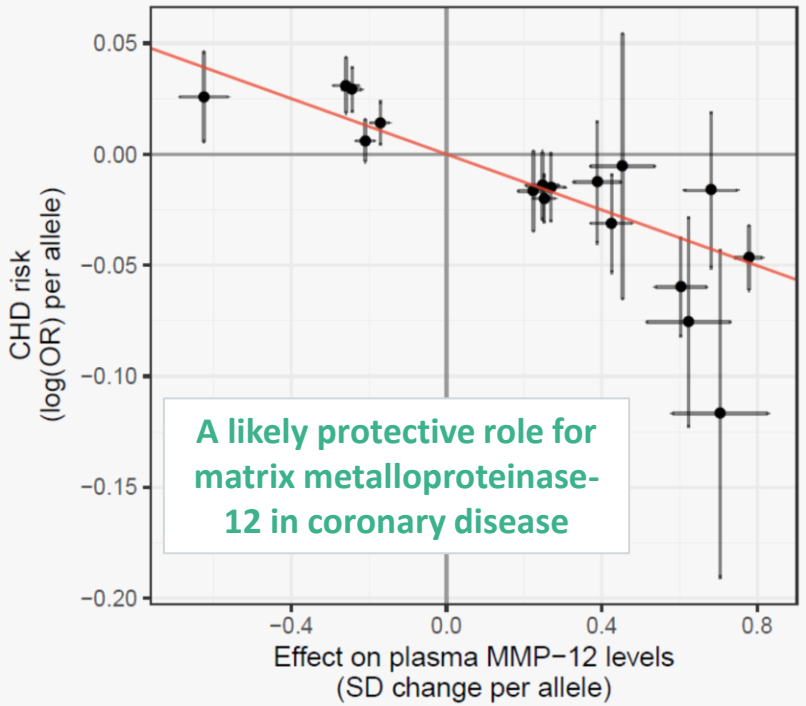
HDRUK Multi-omics Cohorts Consortium

HDRUK
Health Data Research UK

13 cohorts, >800K participants



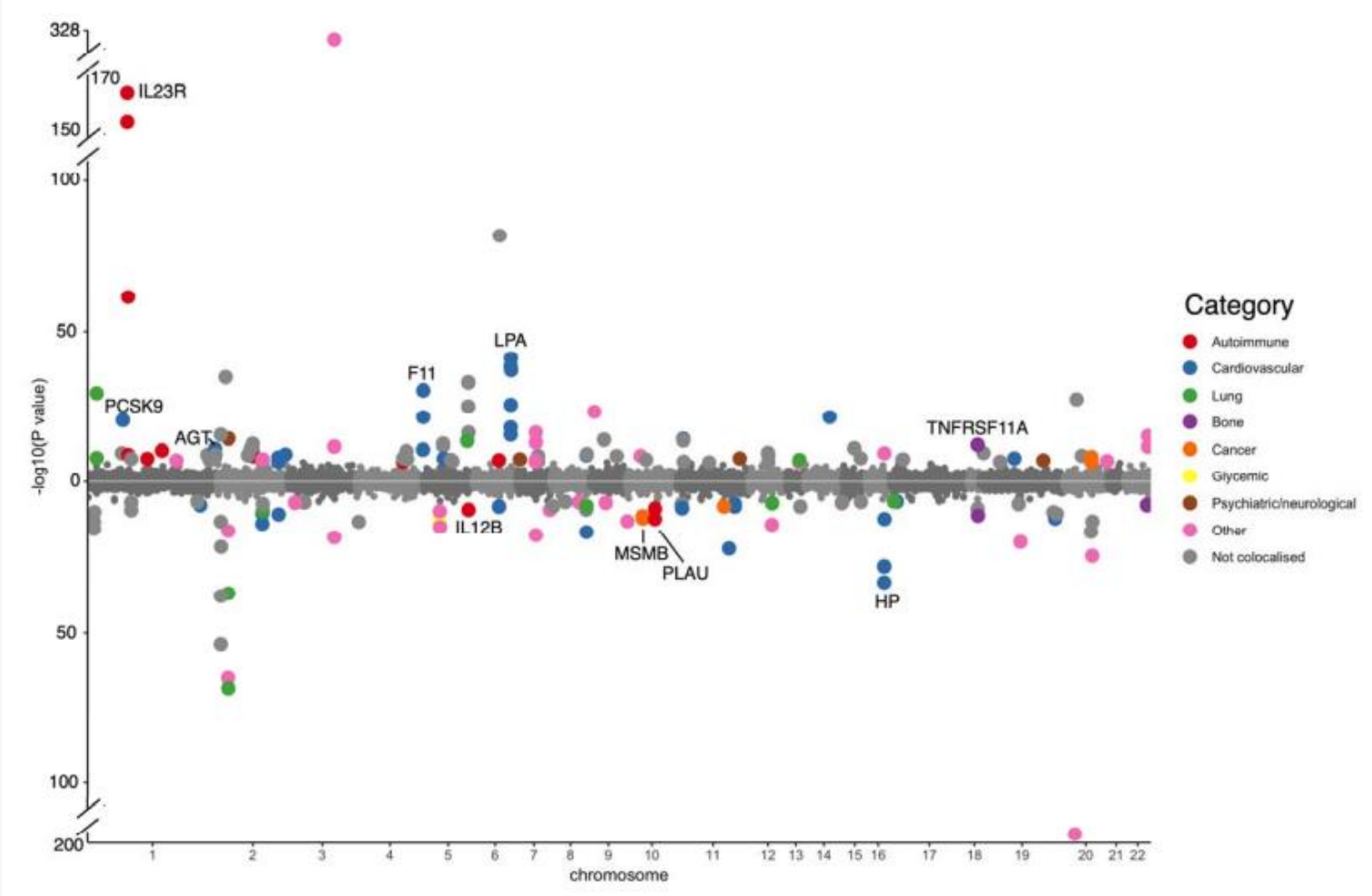
Integrative analyses of multi-omic, genomic and EHR data



- Multi-disease aetiology
- Therapeutic target prioritisation
- Risk prediction

Testing ~1000 proteins for causal effects on >200 diseases and traits

HDRUK
Health Data Research UK

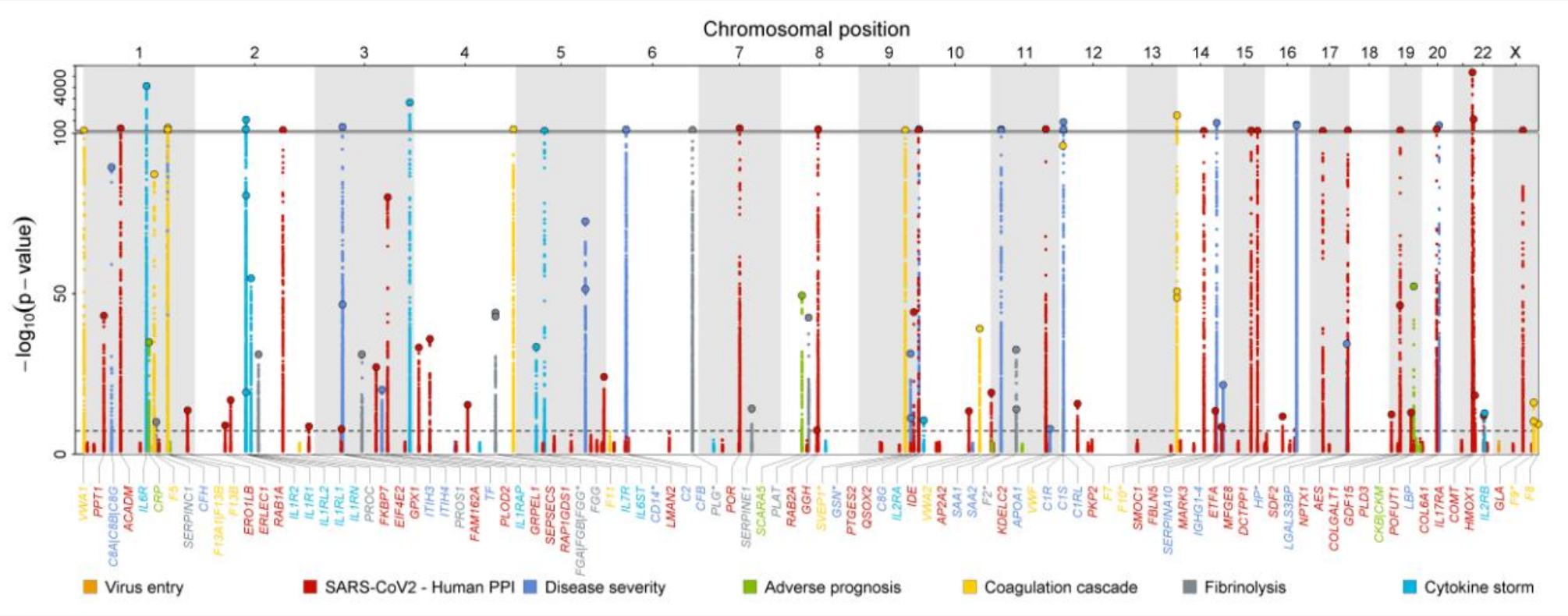


gsk GlaxoSmithKline

Zheng *et al, Nature Genetics*, 2020

Genetic signals for ~100 proteins associated with SARS-CoV2 or COVID-19 response

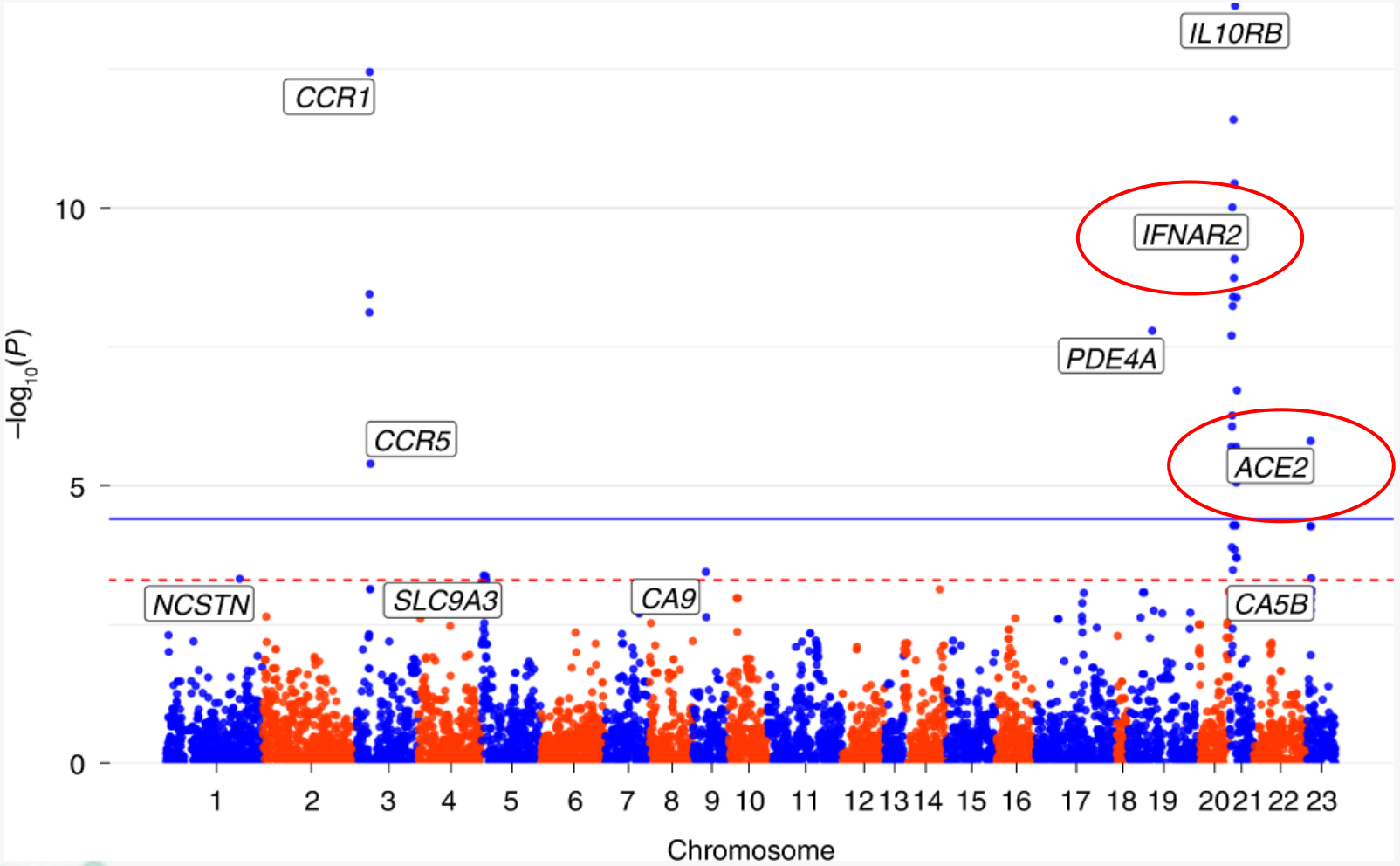
HDRUK
Health Data Research UK



Pietzner *et al, bioRxiv*, 2020

Integrating genomics, transcriptomics, proteomics and COVID-19 outcomes to identify potential treatments

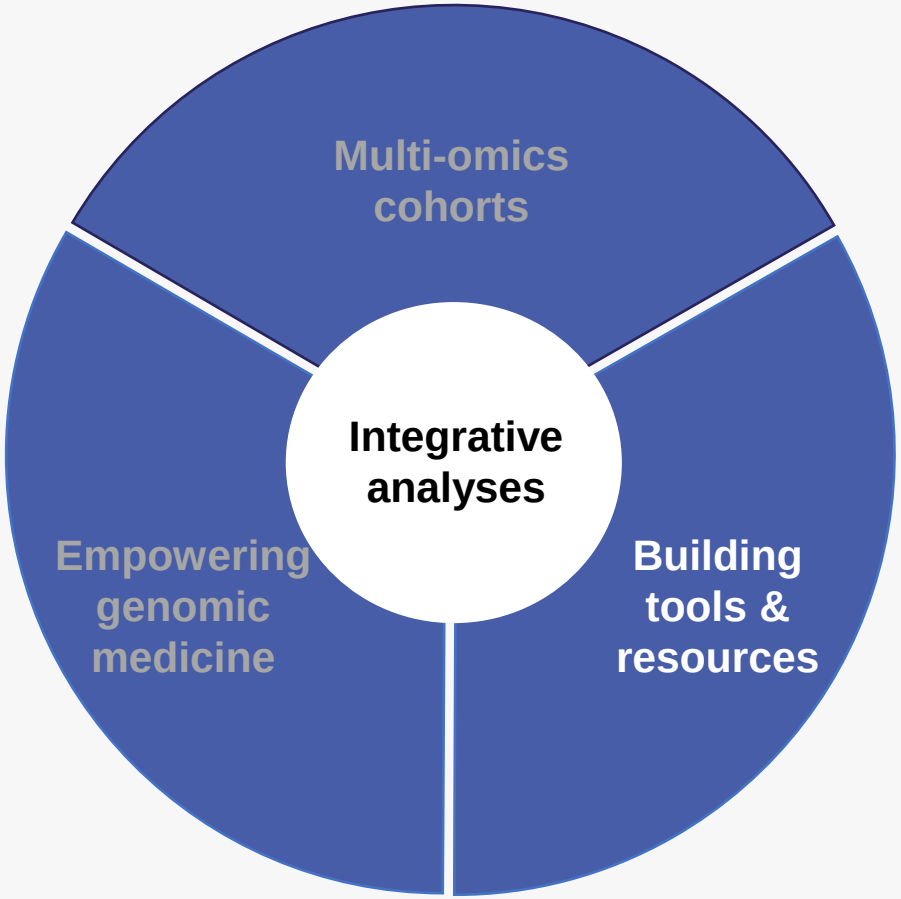
HDRUK
Health Data Research UK



Gaziano *et al, Nature Medicine*, 2021

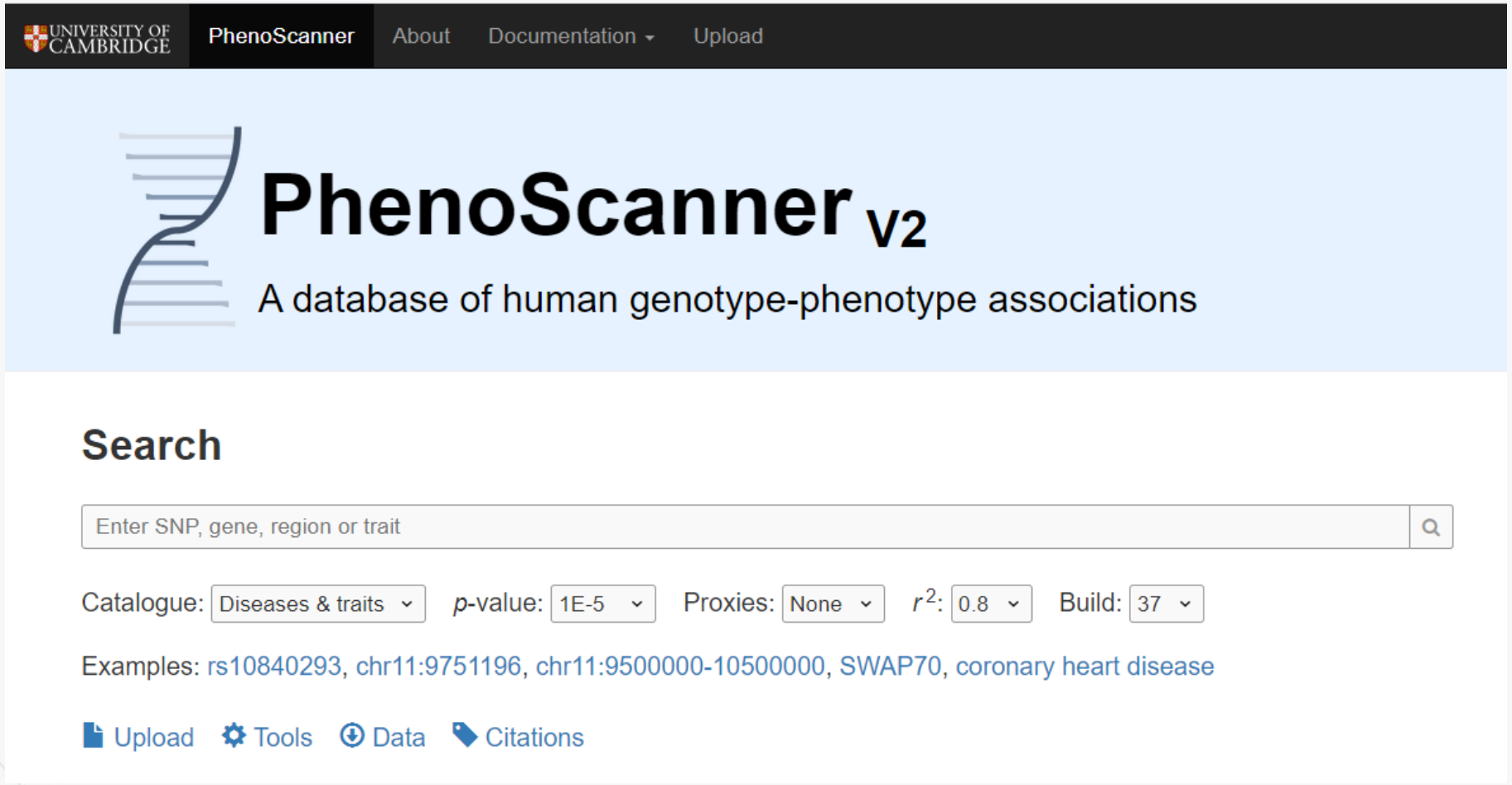
Understanding Causes of Disease: from molecules to electronic health records

HDRUK
Health Data Research UK



Publicly accessible community resources

HDRUK
Health Data Research UK



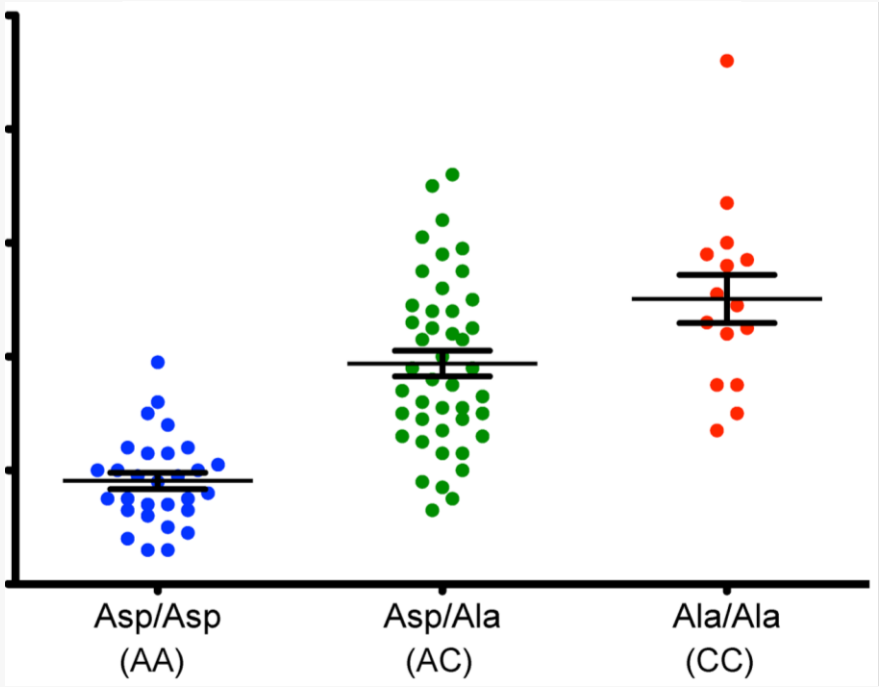
- 65 billion associations
- ~20,000 users from 100 countries
- Millions of database queries
- Cited in 610 papers

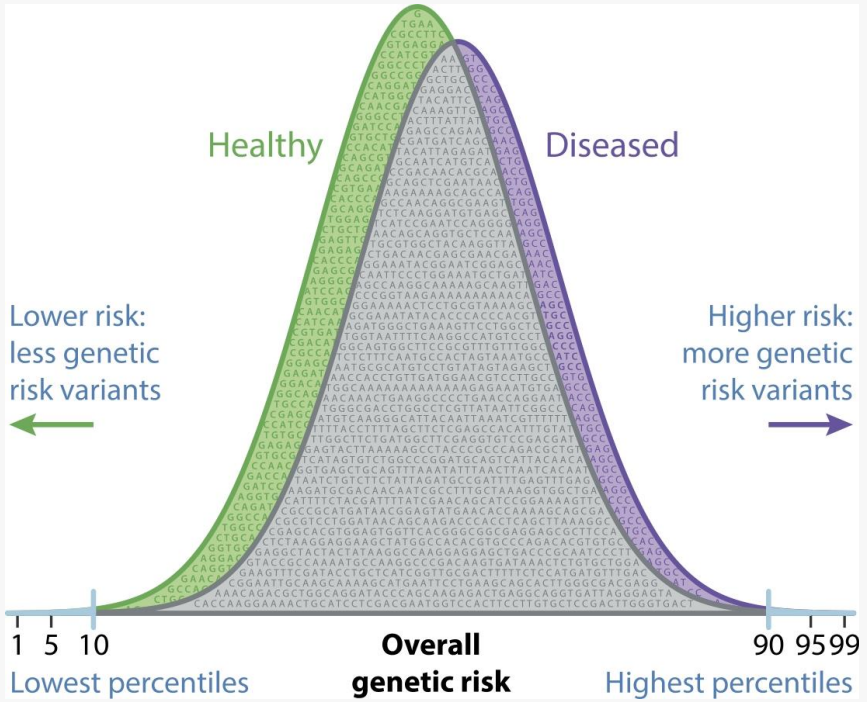
Kamat et al, *Bioinformatics,* 2018

Polygenic risk scores

HDRUK
Health Data Research UK

A weighted sum of many disease-associated genetic variants



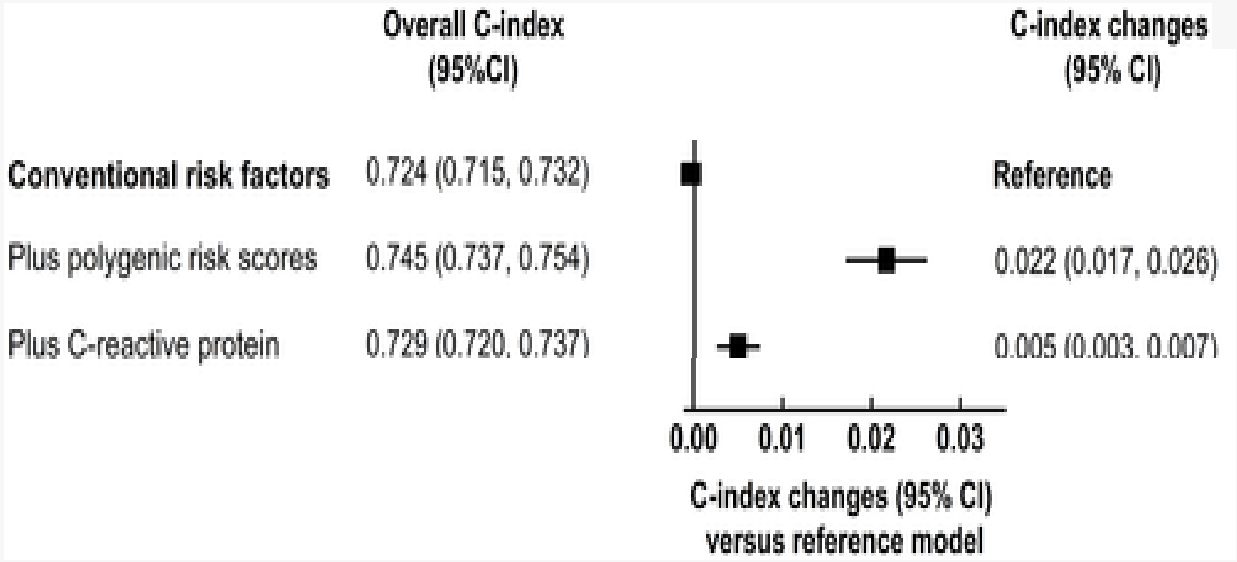


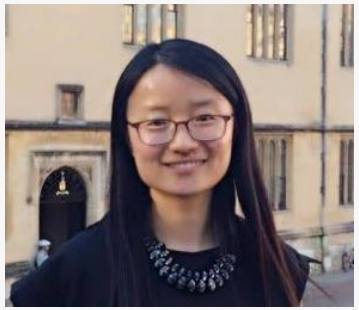
One variant → Thousands of variants

Polygenic risk scores improve prediction of coronary heart disease in UK Biobank

HDRUK
Health Data Research UK

"Polygenic risk scores……could translate into meaningful clinical benefit if applied at scale, and lead to the prevention of 7% more CVD events than conventional risk factors alone."

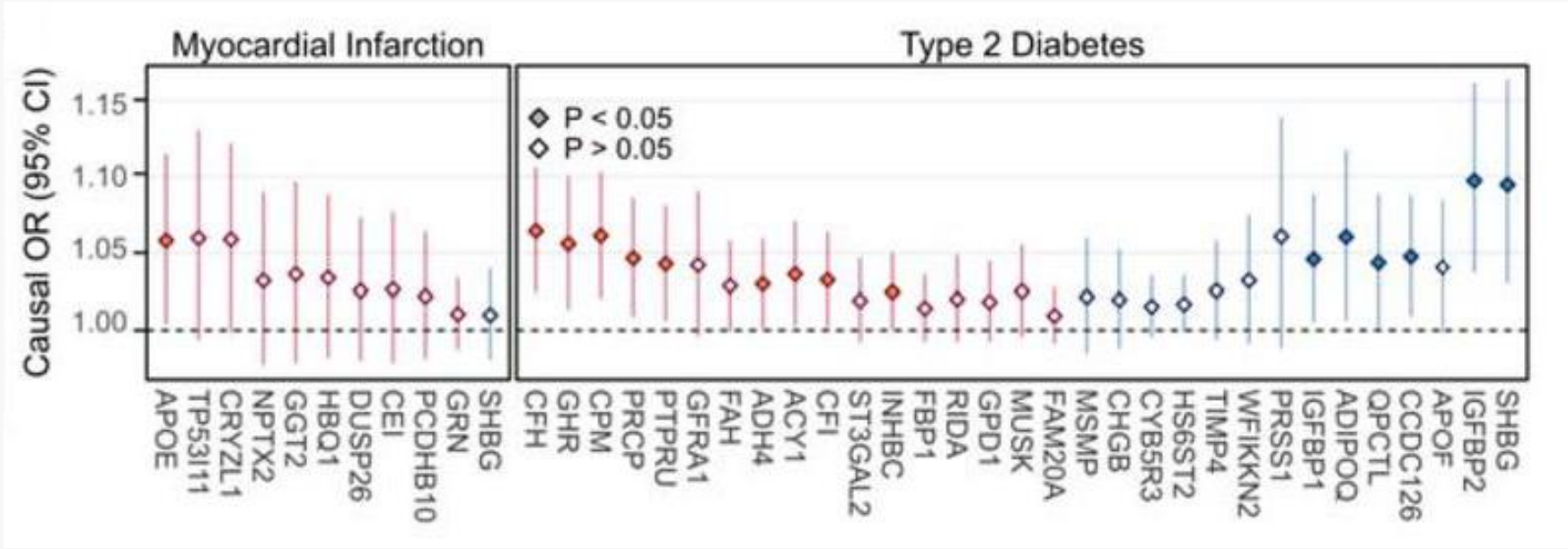




Sun et al, *PLoS Medicine*, 2021

Combining PGS with proteomics to identify aetiological pathways

HDRUK
Health Data Research UK



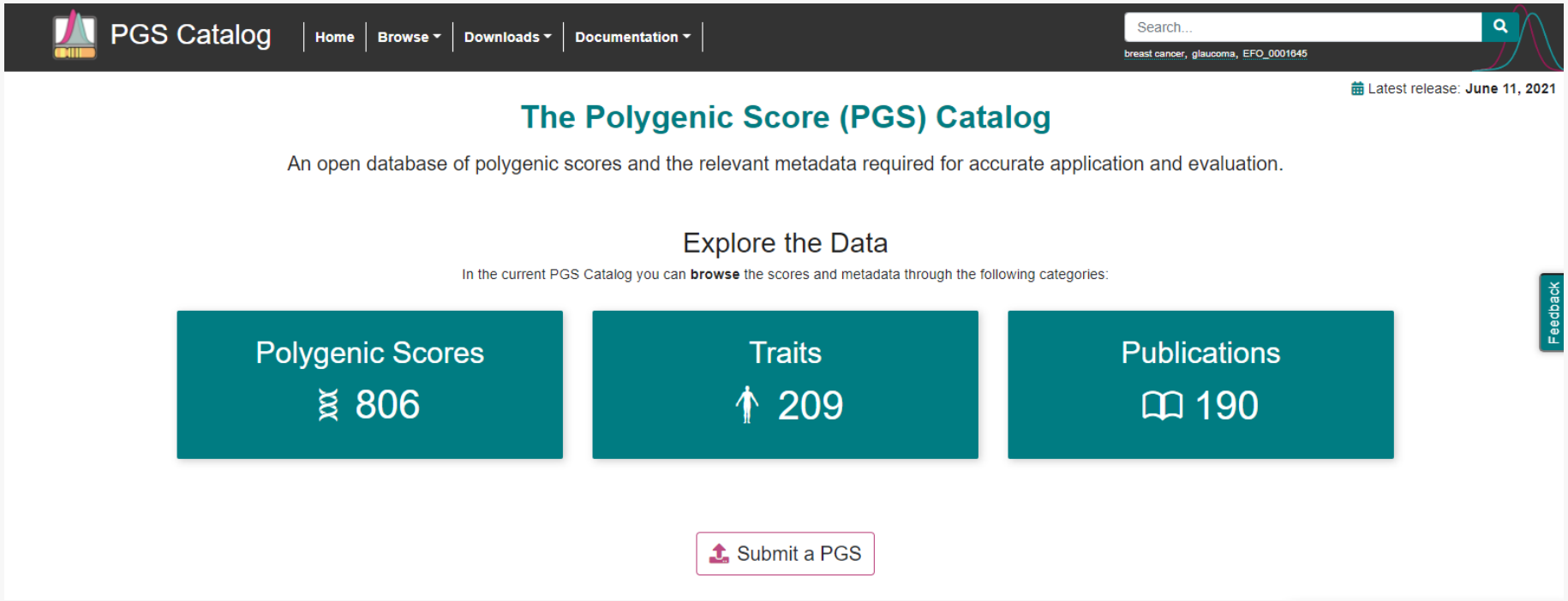
Ritchie et al, *medRxiv,* 2021

www.pgscatalog.org

Polygenic Score Catalog

HDRUK
Health Data Research UK

An open database created in partnership with EMBL-EBI, NHGRI and ClinGen



All published polygenic scores, systematic benchmarking

Lambert et al, *Nature Genetics* 2021

Meta data about each polygenic score

www.pgscatalog.org

HDRUK
Health Data Research UK

PGS Catalog / Polygenic Scores / PGS000018

Polygenic Score (PGS) ID: **PGS000018**

Download Score | FTP directory

Terms and Licenses

Predicted Trait	
Reported Trait	**Coronary artery disease**
Mapped Trait(s)	**coronary artery disease** (EFO_0001645)

Released in PGS Catalog: **Oct. 14, 2019**

Score Details

Score Construction	
PGS Name	**metaGRS_CAD**
Variants	
Original Genome Build	**hg19**
Number of Variants	1,745,179
Development Method	
Name	metaGRS
Parameters	metaGRS log(HR) mixing weights: GRS46K=0.1278, FDR202=0.2359 and 1000Genomes=0.2400

PGS Source	
PGS Catalog Publication (PGP) ID	PGP000007
Citation (*link to publication*)	Inouye M *et al. J Am Coll Cardiol* (2018)

Ancestry Distribution

Source of Variant Associations (**G**WAS)	Multi-ancestry (including European): 50.9% European: 37% South Asian: 6.7% East Asian: 3% Hispanic or Latin American: 1.1% African: 0.8% Greater Middle Eastern: 0.6% 382,026 individuals (100%)
Score **D**evelopment/Training	Multi-ancestry (including European): 100% 3,000 individuals (100%)
PGS **E**valuation	European: 68.8% African: 12.5% Hispanic or Latin American: 12.5% Multi-ancestry (including European): 6.2% 16 Sample Sets

Guidelines for reporting PGS

HDRUK
Health Data Research UK

Reporting standards necessary for PGS reproducibility, translation and public trust

Perspective

Improving reporting standards for polygenic scores in risk prediction studies

https://doi.org/10.1038/s41586-021-03243-6

Received: 20 April 2020

Accepted: 15 January 2021

Published online: 10 March 2021

Check for updates

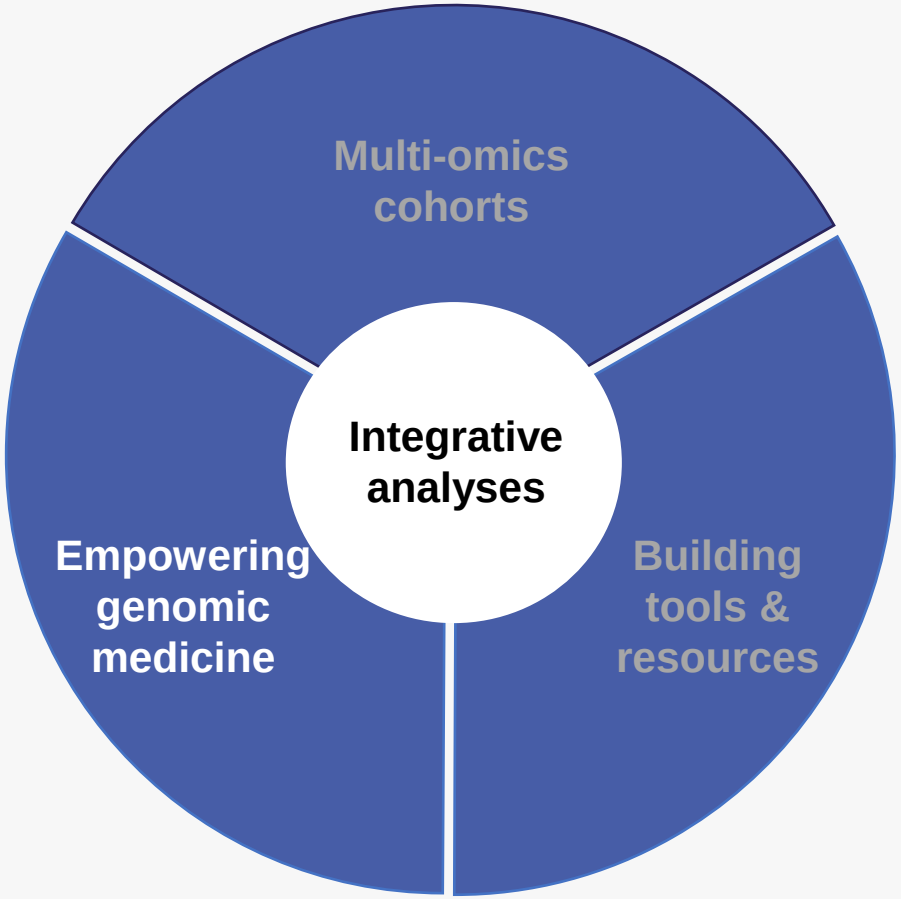
Hannah Wand[1,2,31], Samuel A. Lambert[3,4,5,6,7,31], Cecelia Tamburro[8], Michael A. Iacocca[1], Jack W. O'Sullivan[1,2], Catherine Sillari[8], Iftikhar J. Kullo[9], Robb Rowley[8], Jacqueline S. Dron[10,11], Deanna Brockman[10], Eric Venner[12], Mark I. McCarthy[13,14], Antonis C. Antoniou[15], Douglas F. Easton[15], Robert A. Hegele[11], Amit V. Khera[10], Nilanjan Chatterjee[16,17], Charles Kooperberg[18], Karen Edwards[19], Katherine Vlessis[20], Kim Kinnear[20], John N. Danesh[5,6,21], Helen Parkinson[6,7], Erin M. Ramos[8], Megan C. Roberts[22], Kelly E. Ormond[20,23], Muin J. Khoury[24], A. Cecile J. W. Janssens[25], Katrina A. B. Goddard[26,27], Peter Kraft[28], Jaqueline A. L. MacArthur[7], Michael Inouye[3,4,5,6,21,29,32] & Genevieve L. Wojcik[30,32]

Polygenic risk scores (PRSs), which often aggregate results from genome-wide association studies, can bridge the gap between initial discovery efforts and clinical applications for the estimation of disease risk using genetics. However, there is notable heterogeneity in the application and reporting of these risk scores, which hinders the translation of PRSs into clinical care. Here, in a collaboration between the Clinical Genome Resource (ClinGen) Complex Disease Working Group and the

Wand et al, *Nature* 2021

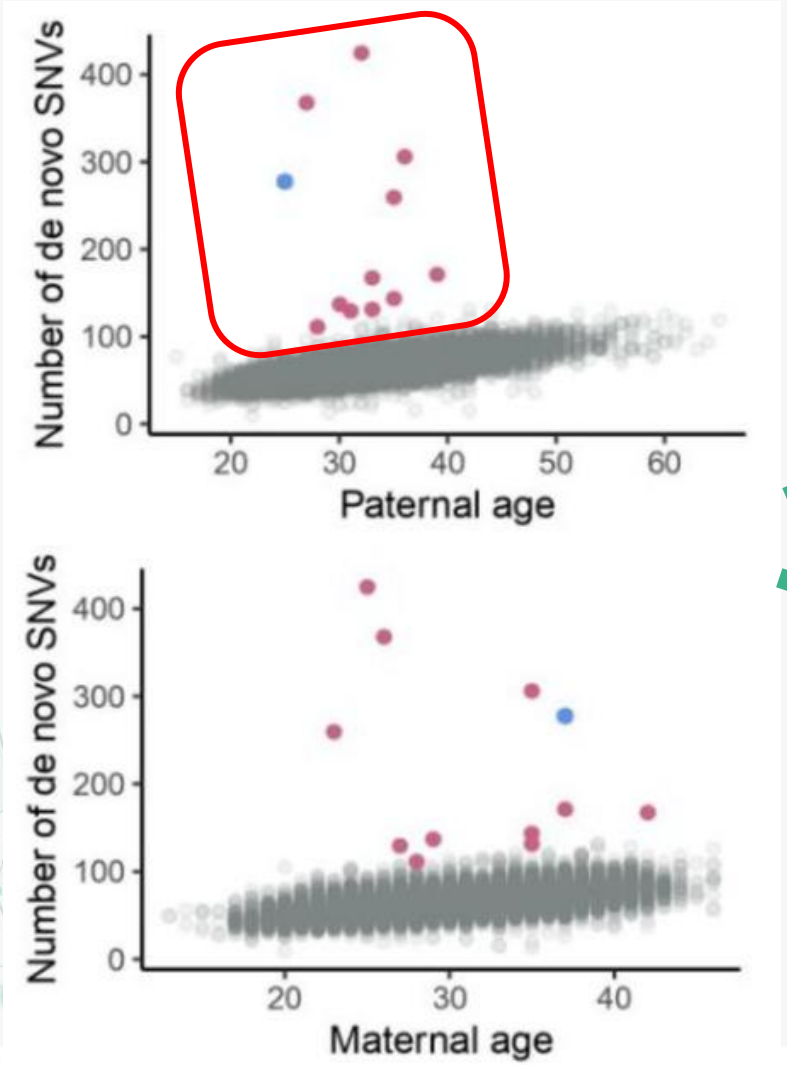
Understanding Causes of Disease: from molecules to electronic health records

HDRUK
Health Data Research UK



Linking genomes and longitudinal EHRs to identify causes of extreme mutation rates

HDRUK
Health Data Research UK



4 with likely genetic explanations

5 with paternal cancer and chemotherapy prior to conception

Kaplanis *et al, bioRxiv*, 2021

The future

HDRUK
Health Data Research UK

1. Enhanced diversity

BELIEVE study

Kadoorie Biobank

Mexico City Prospective Study

Million Veteran Program

The future

HDRUK
Health Data Research UK

1. Enhanced diversity

2. Deeper data linkages

British Heart Foundation
Data Science Centre
Led by Health Data Research UK

The future

1. Enhanced diversity
2. Deeper data linkages
3. Multimorbidity

Funders


MRC
Medical
Research
Council


NHS
National Institute for
Health Research


gsk
GlaxoSmithKline


AstraZeneca


NOVARTIS


Pfizer


wellcome trust


DH
Department
of Health


HDRUK
Health Data Research UK


MERCK


British Heart
Foundation


NHS
Blood and Transplant


REGENERON


Engineering and
Physical Sciences
Research Council


SomaLogic


Biogen


UK Research
and Innovation


Economic
and Social
Research Council


olink
PROTEOMICS


Nightingale


BIOMARIN

Acknowledgements

Understanding Causes of Disease theme members

Multi-omics Cohorts Consortium

Polygenic Score Catalog team

HDR UK Cambridge Hub team and Executive committee

Patient and public panels, representatives and Steering Committee members

Study participants

HDRUK team