biobank uk

Enabling scientific discoveries that improve human health

Whole Genome Sequencing release and multi-omics in UK Biobank

ASHG

November 2nd 2023

Lucy Burkitt-Gray, PhD

Lead Data Analyst

UK Biobank

What is UK Biobank?

biobank uk

Enabling scientific discoveries that improve human health

UK Biobank is a large-scale **biomedical database** and research **resource**, containing in-depth genetic and health information from **half a million** UK participants. The database is **regularly augmented** with additional data and is **globally accessible** to approved researchers undertaking vital research into the most **common and life-threatening diseases**. It is a major contributor to the **advancement of modern medicine** and treatment and has enabled several scientific discoveries that **improve human health**.

Non-profit charity, established by:

MRC Medical Research Council

wellcome

Ongoing core funding and additional funding from:

MRC Medical Research Council

wellcome

CANCER RESEARCH UK

DH Department of Health

NIHR National Institute for Health Research

British Heart Foundation

The Scottish Government

Northwest

DIABETES UK KNOW DIABETES. FIGHT DIABETES.

Llywodraeth Cymru Welsh Government

Link to 'The ground breaking UK Biobank Resource' video:
https://www.youtube.com/watch?v=NGdegXRx8U0&t=152s

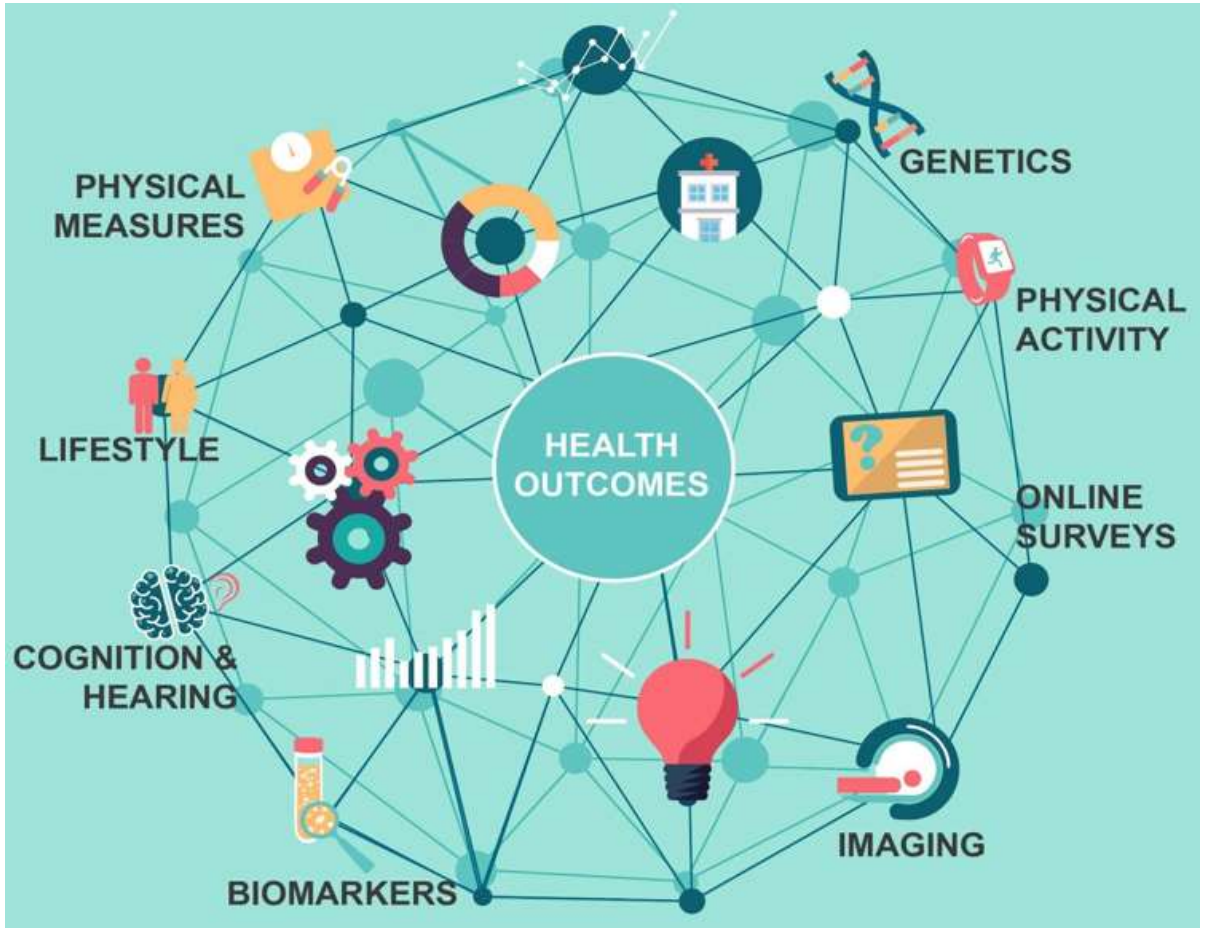
Link to 'What is UK Biobank?' video:
https://www.youtube.com/watch?v=66moI1ZHMYs

Link to 'Celebrating 20 years of UK Biobank':
https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/our-impact

UK Biobank recruitment and baseline assessment

- 500,000 people aged 40 – 69 in 2006-10 from England (89%), Scotland (7%), and Wales (4%)
- 22 assessment centres located around the UK to enhance heterogeneity of the cohort
- Breadth and depth of data
  - Lifestyle and environmental exposures
  - Personal and family medical history
  - Cognitive function, hearing and vision tests
  - Physical measures (BP, lung function, body size)
  - Biological samples (blood, urine, saliva)
- Consent to access health-related records and to re-contact participants for further assessments

Health outcomes and health linkage



Regular updates on participant lifestyle, health and risk factors through focussed questionnaires

1-2 questionnaires per year on specific topic

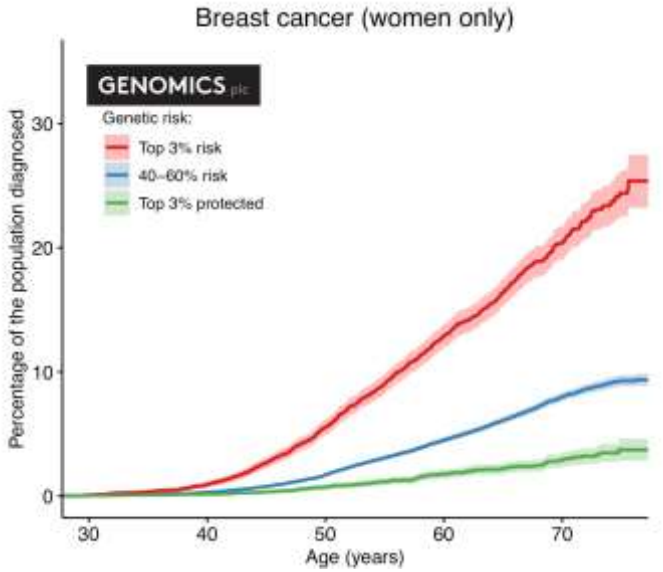
Nutrition, sleep, mental health, pain

Linkage to NHS Digital records provide detailed health information on hospital admissions, GP records, prescriptions, and more

Linkage to death certificates for participant mortality and causes of death

At home physical activity monitoring, cardiac monitoring, COVID antibody testing

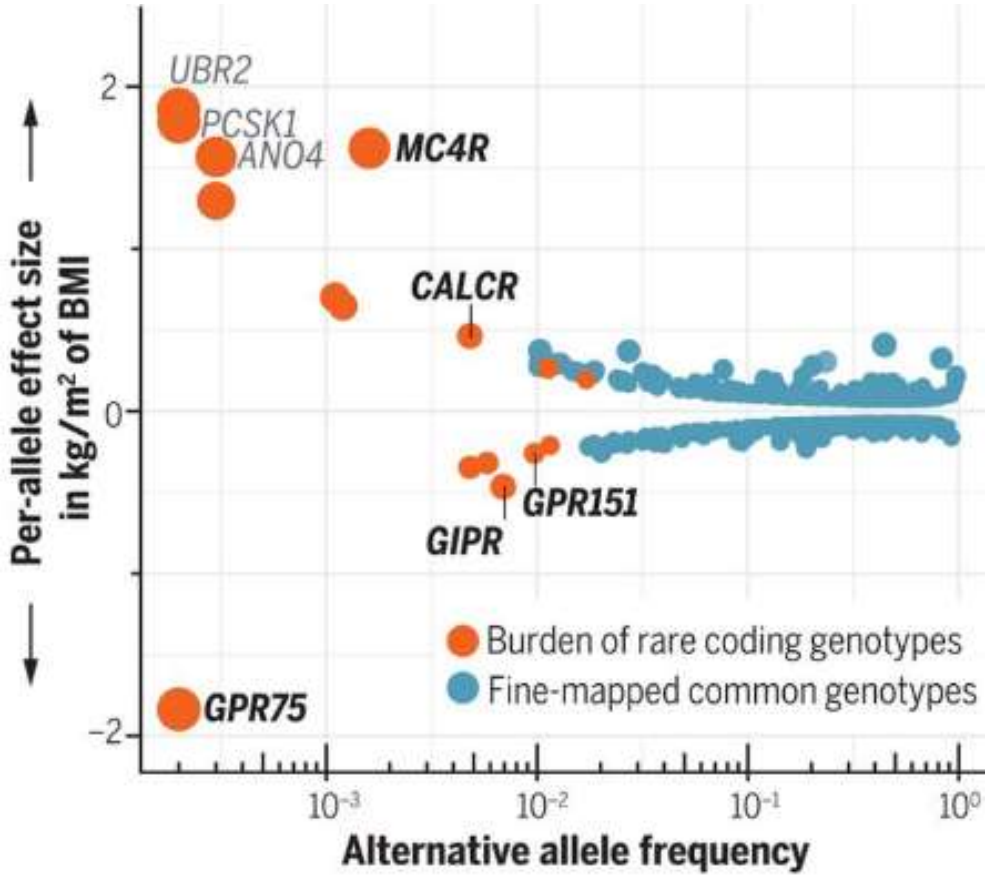
Genotyping of all 500,000 participants



Thompson *et al.*, MedRxiv 2022

- Custom-built genotyping array (850k variants) with imputed measures for 90M+ variants
- Data made available for all 500,000 participants in 2017
- GWAS of hundreds of traits now publicly available
- Polygenic risk scores generated for wide range of conditions
- Made possible due to UK Biobank's large size, standardised assessment and health outcome follow-up through linkage to electronic records

Exome sequencing

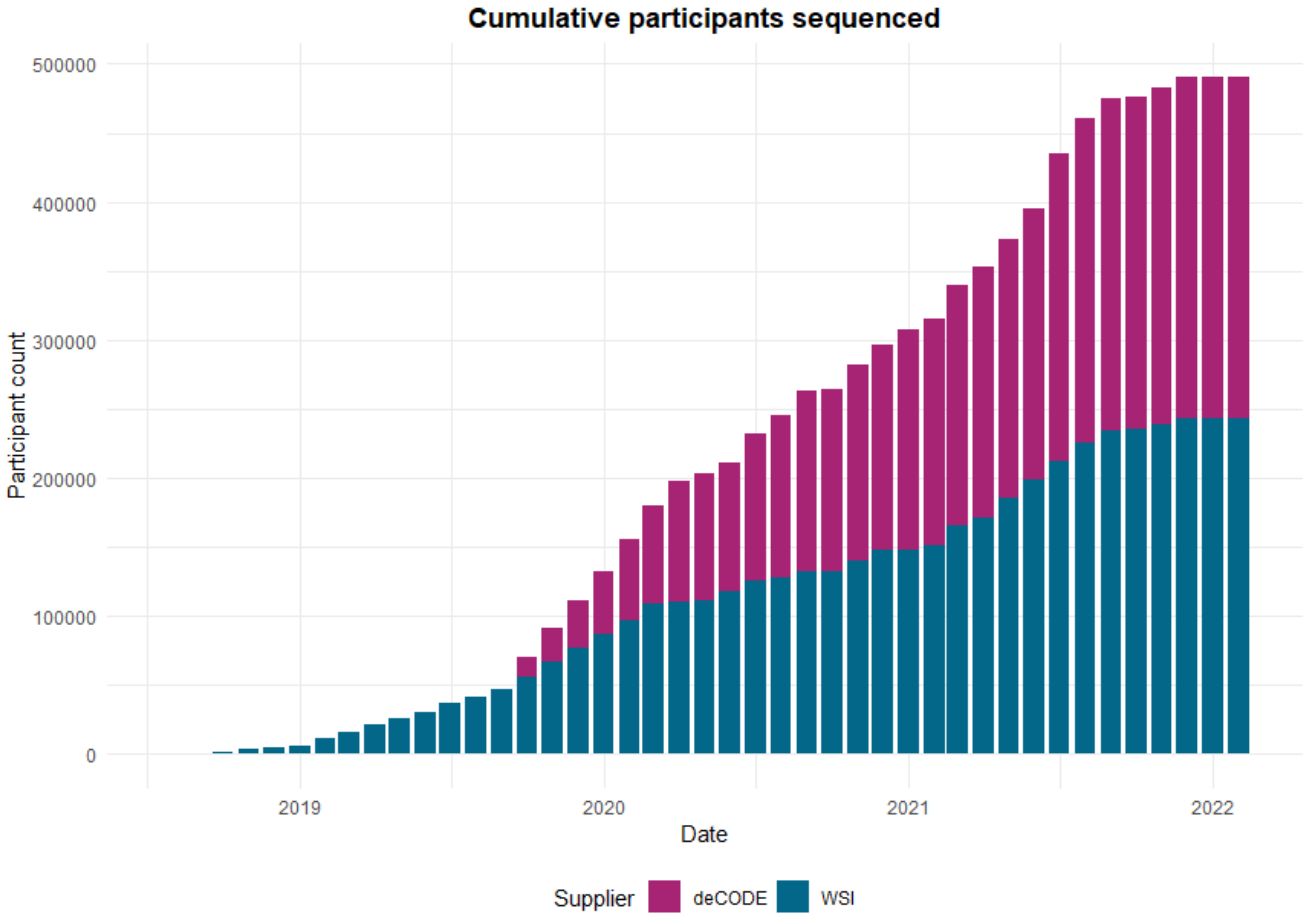


Power of massive-scale exome sequencing to enable discovery-based gene-burden analysis (e.g. GPR75 as obesity therapeutic target)

Akbari *et al.*, Science 2021

- Sequencing of all coding regions of the genome (~2%)
- Regeneron-led commercial funding
- Sequenced with Illumina NovaSeq 6000 platform using S2/S4 flow cells
- First 50,000 sequences released in 2019; final release of whole cohort in mid-2022

Whole genome sequencing

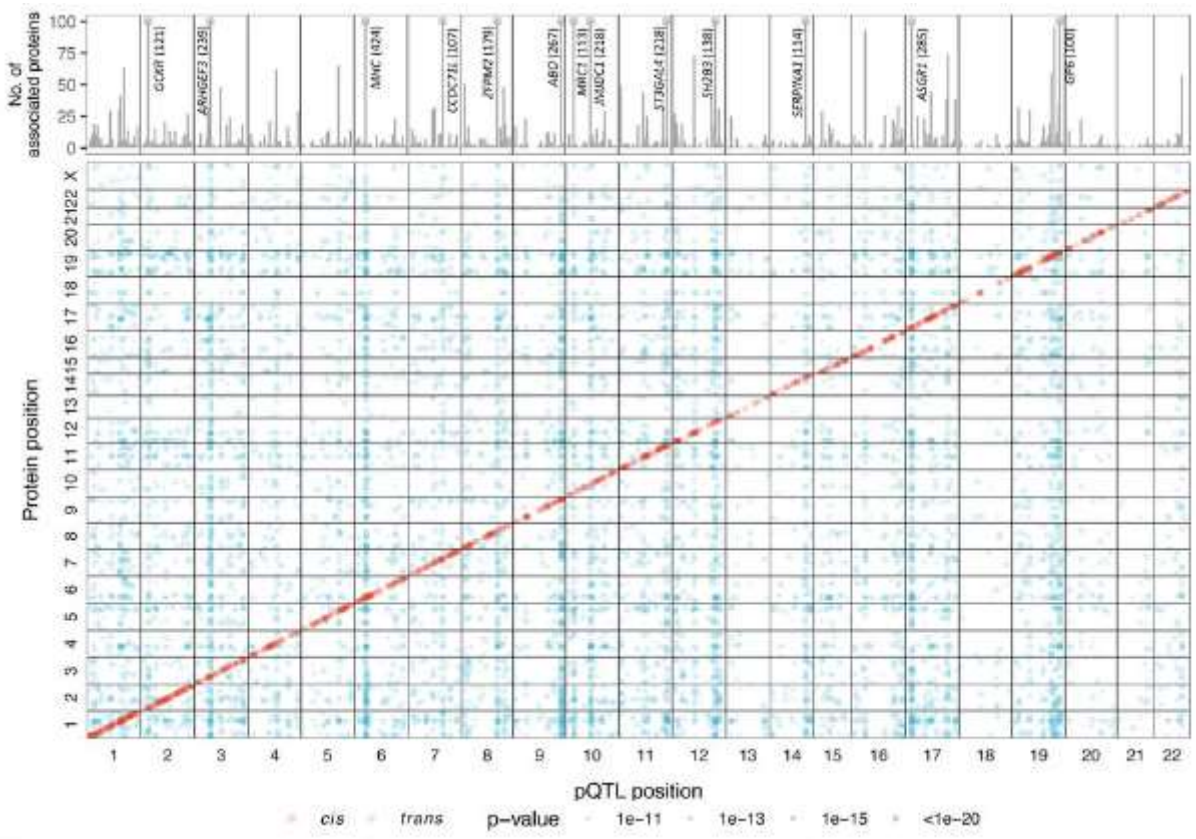


- 30X sequencing across the entire genome
- Government, charity, and industry funding
- Sequencing carried out between Wellcome Sanger Institute and deCODE Genetics using Illumina NovaSeq
- First 200,000 sequences publically available late 2021; full cohort to be released end 2023

UK Research and Innovation

GSK AMGEN Janssen AstraZeneca wellcome

Proteomics



Targeted proteomics on ~60,000 participants began in late 2020 supported by large pharmaceutical consortium

Using Olink platform to measure 3,000 circulating proteins from plasma samples

Already world's largest pQTL collection (10,000+)

Data on 3,000 proteins for 56,000 participants now available

Approved mass spectrometry-based study commencing next year, for full-cohort analysis by Eliptica

ELIPTICA

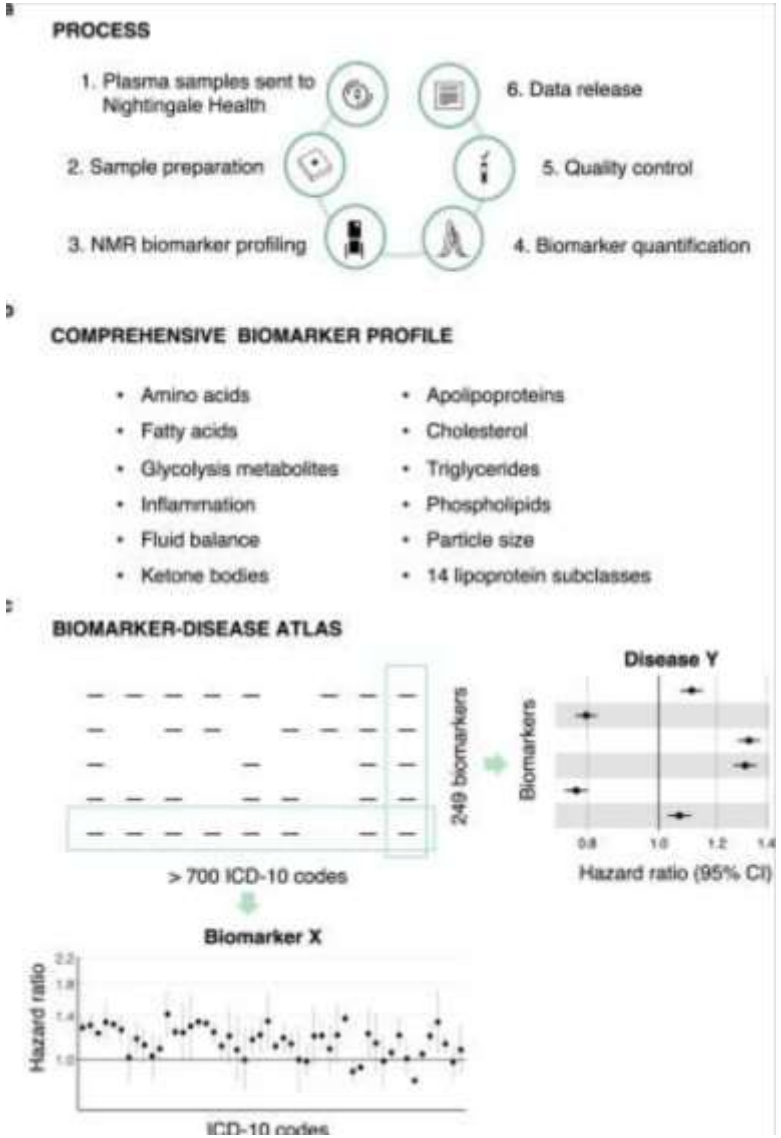
Olink

Sun *et al.*, BioRxiv 2022

Metabolomics

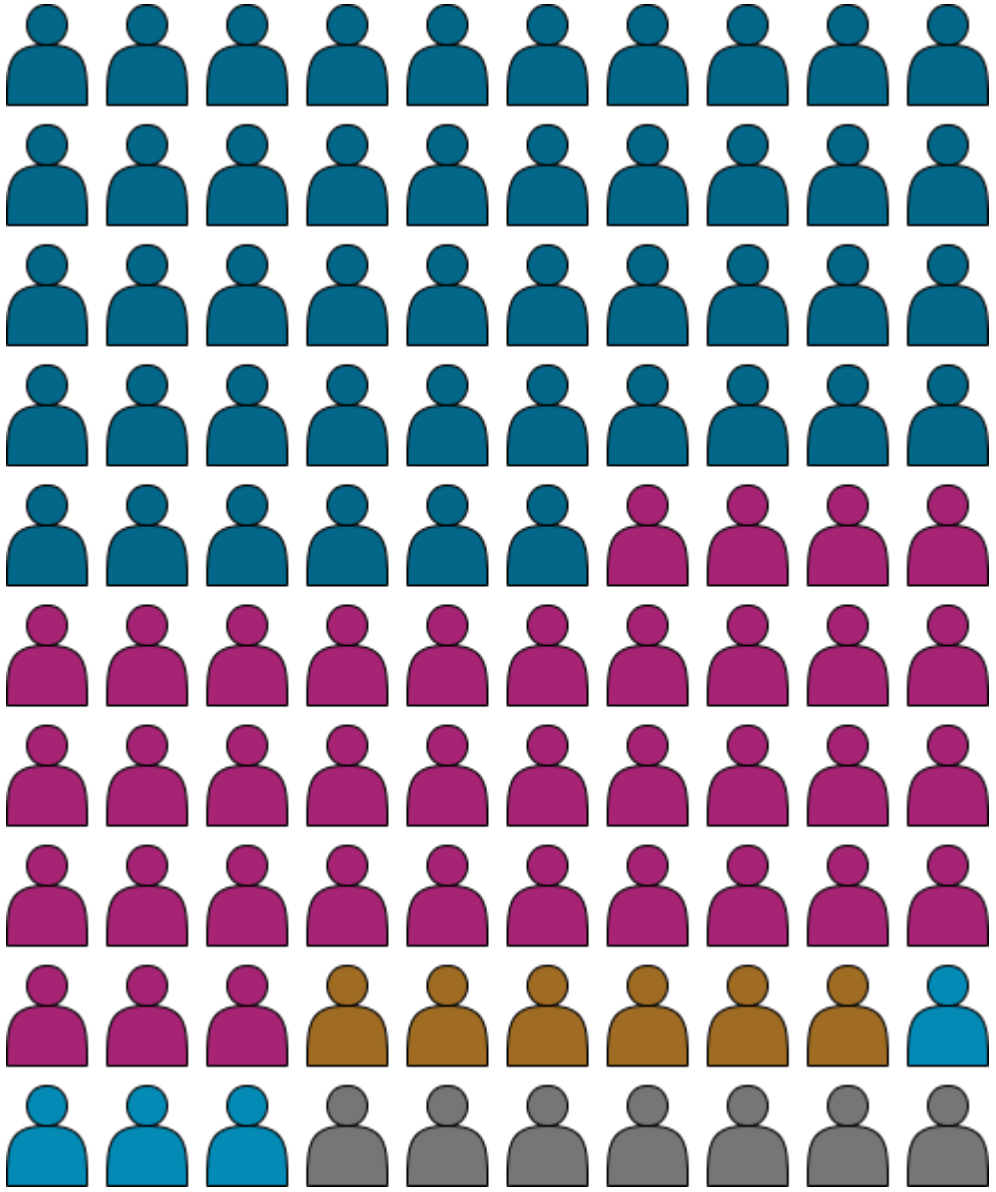
- Metabolomic study carried out with Nightingale Health analysing 249 biomarkers in plasma
- Data for 275,000 participants made available in early 2023
- 16,000 repeat samples analysed, providing two timepoints 2-7 years apart
- Planned release of full cohort metabolomics late 2024/early 2025

Nightingale Health



Julkunen *et al.*, Nat Comm 2023

Potential for multi-omic analysis



As of the end of 2023:

- 46% of the cohort have whole genome, whole exome, genotyping, and metabolomic data
- A further 37% have whole exome, whole genome, and genotyping
- 6% of the cohort have data in all major omic datasets
- 4% have whole genome, whole exome, genotyping and proteomic data
- Overlaps will increase with future proteomic and metabolomic releases in 2024

Acknowledgements

Core funders

UKRI Medical Research Council

wellcome

CANCER RESEARCH UK

British Heart Foundation

Department of Health & Social Care

Collaborators

Lora Boteva

Oliver Gray

Daisy Vinter

Rachael Winkless

Dan Fry

Caroline Clark

Our 500k participants

Apply for access:

ukbiobank.ac.uk/enable-your-research/apply-for-access