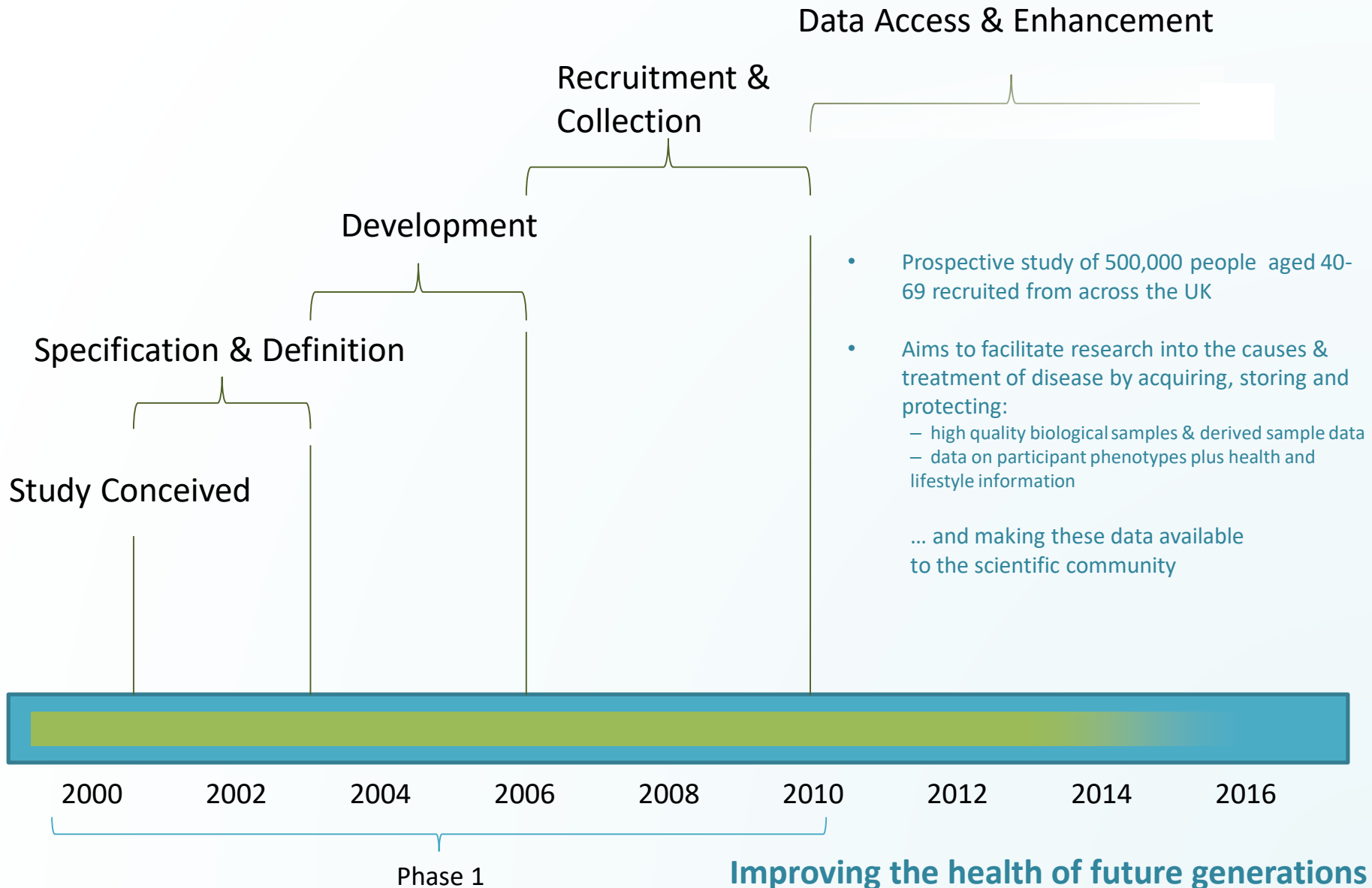




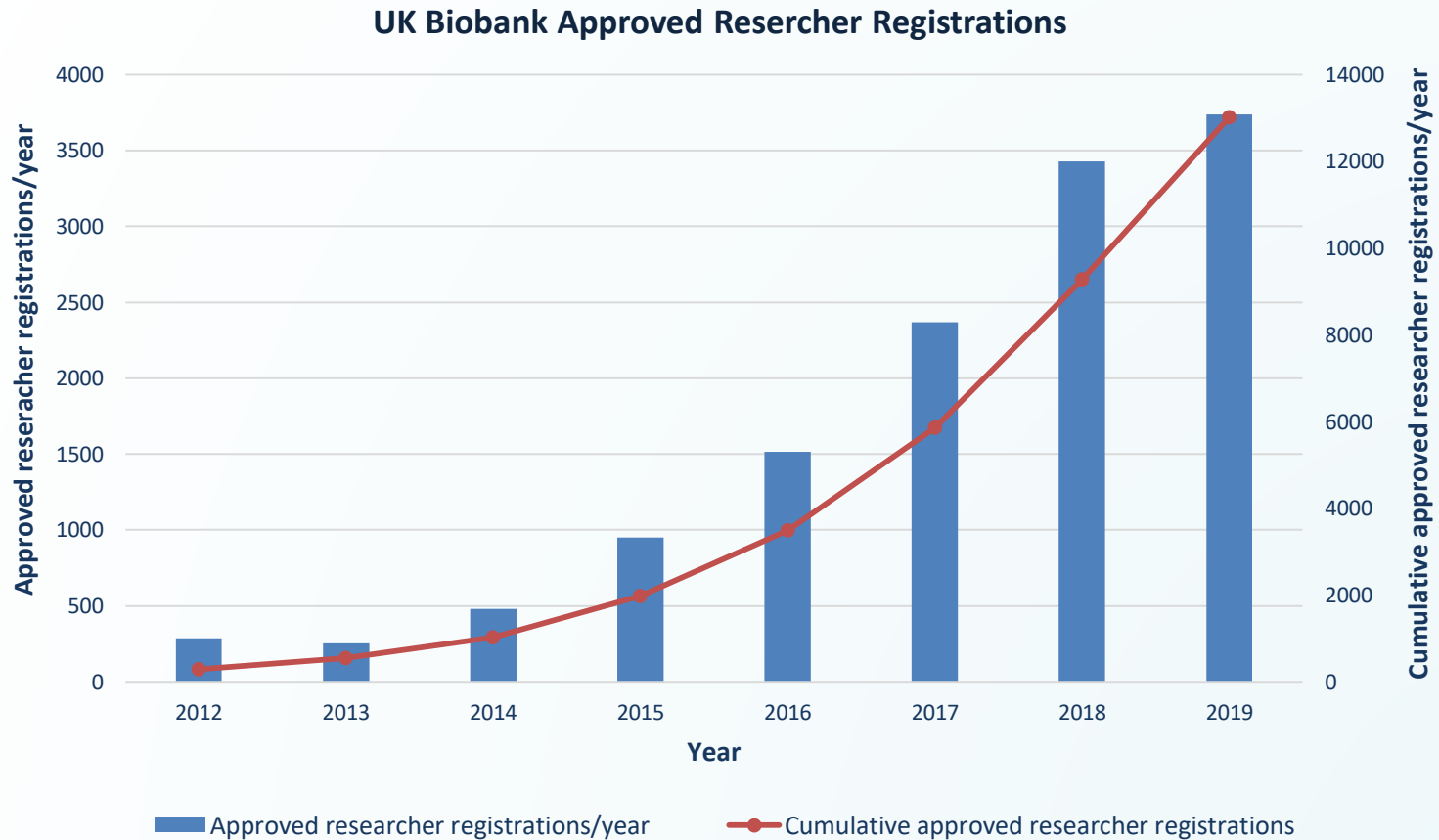
UK Biobank Past, Present, and Future

Chris Boulton
November 2019



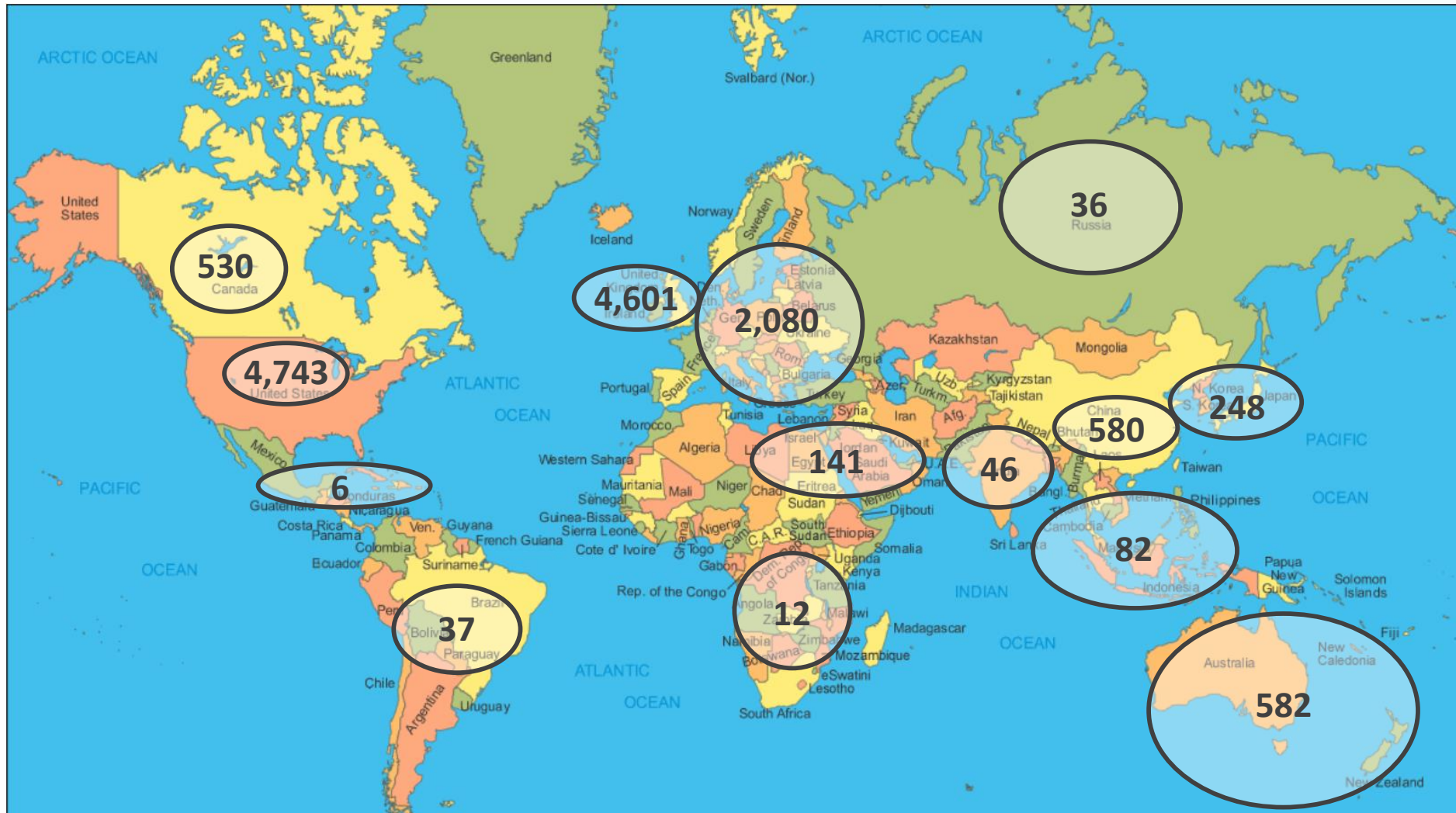
Use of the UK Biobank Resource

Researchers



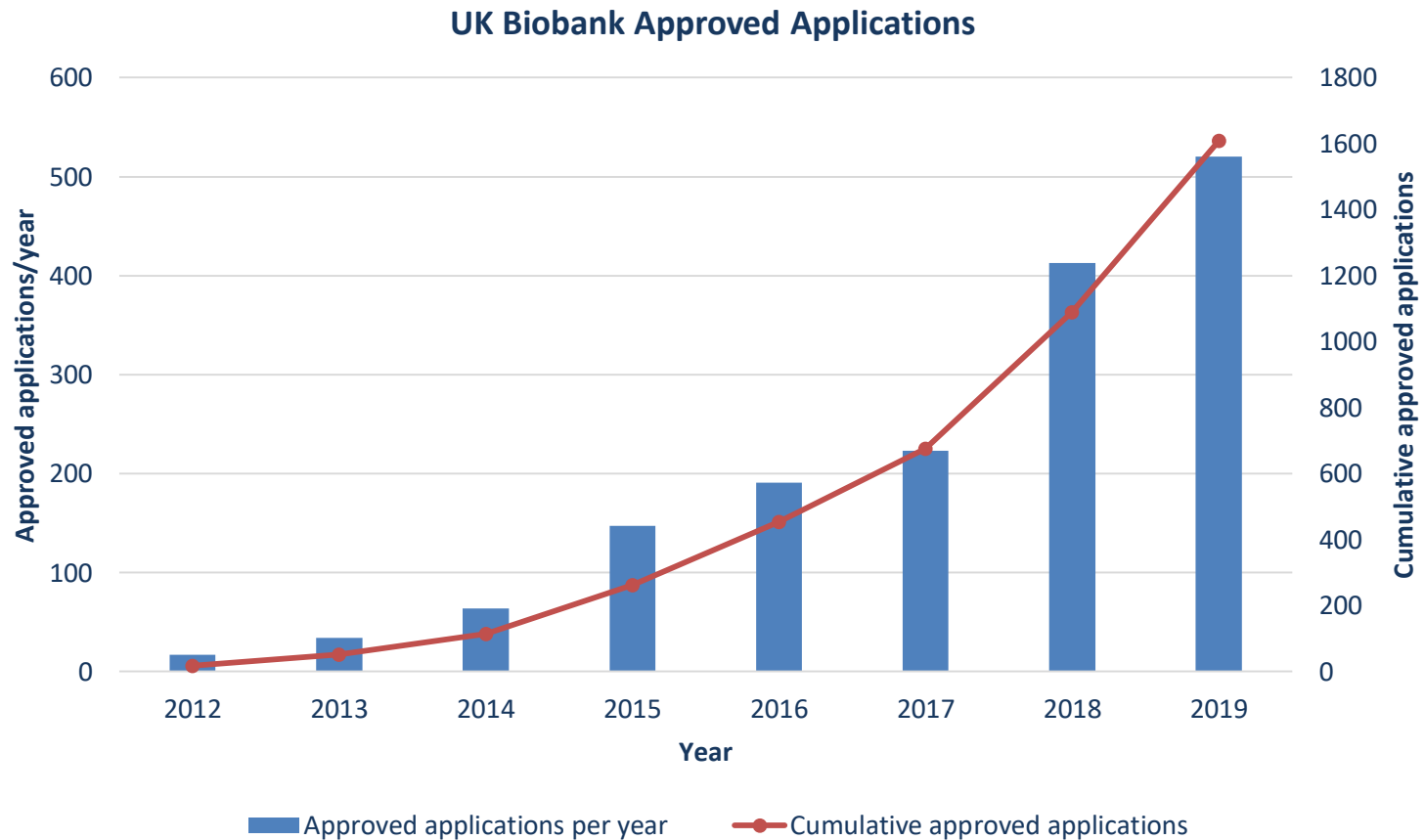
International use of the UK Biobank Resource

Locations of researchers worldwide



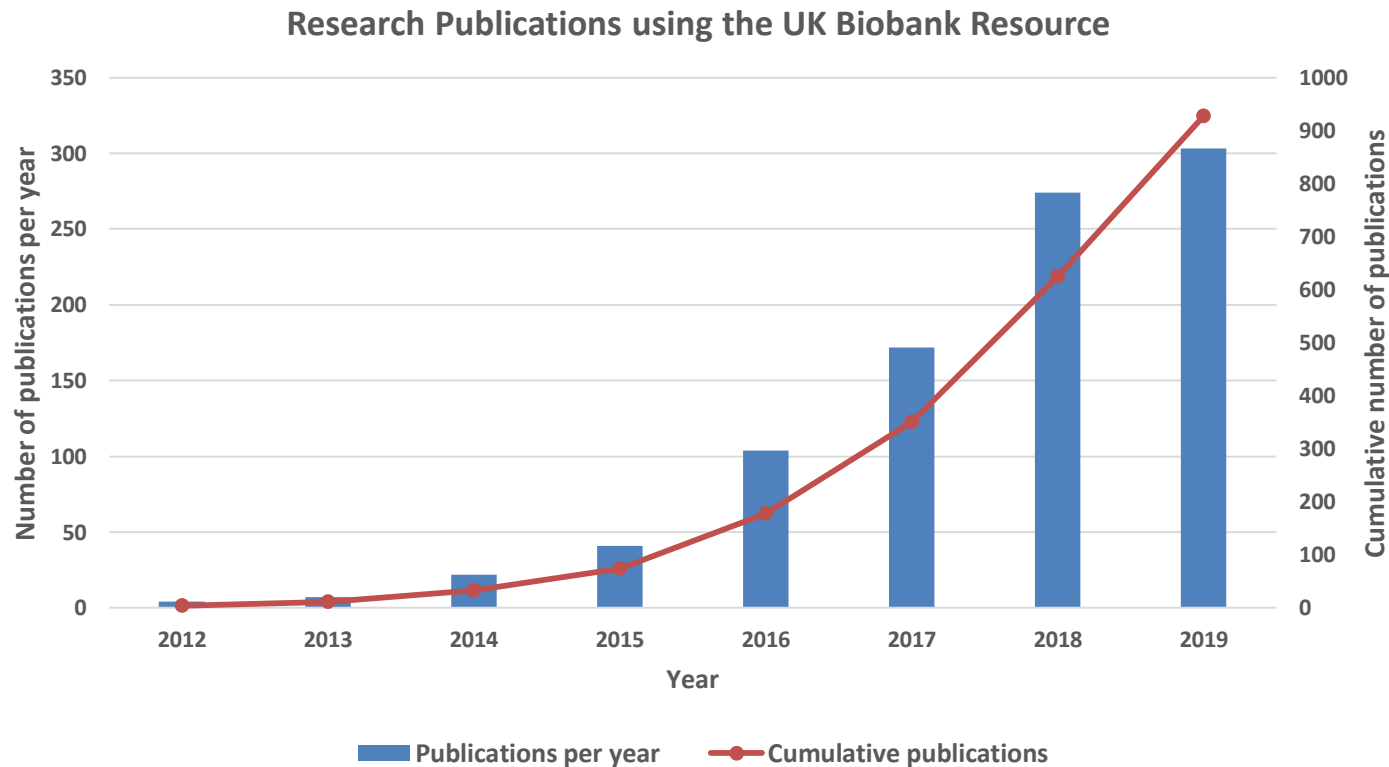
Use of the UK Biobank Resource

Applications



What impact is UK Biobank making?

- There are now over 930 published research papers using the UK Biobank Resource



UK Biobank Resource

Record Linkage

Linkage to medical and other health-related records

Adjudication of health outcomes

Repeat measures

Repeat of baseline assessment
(with enhancements)

Baseline data from assessment centres

Questionnaire /
Interview data

Physical
measures

Biological Samples

Enhancements (Further)

Assays

WBQs

Other

Enhancements (In Progress)

Imaging

Biomarkers

AMSR

Enhancements (Complete)

Genotyping

WBQ

Activity

Returned
data from
use of the
resource

UK Biobank – Enhancement/Access Projects



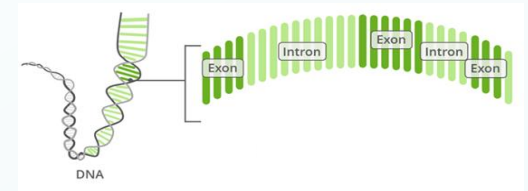
Imaging

- 50,000 participants imaged by January 2020
- Additional centre opening in Bristol 1Q 2020
- Further 10,000 repeats to be included (starting early 2019)



Cardiac Monitoring

- Non-invasive stick-on patch applied in clinic or at home
- Beat-to-beat ECG investigating silent-AF

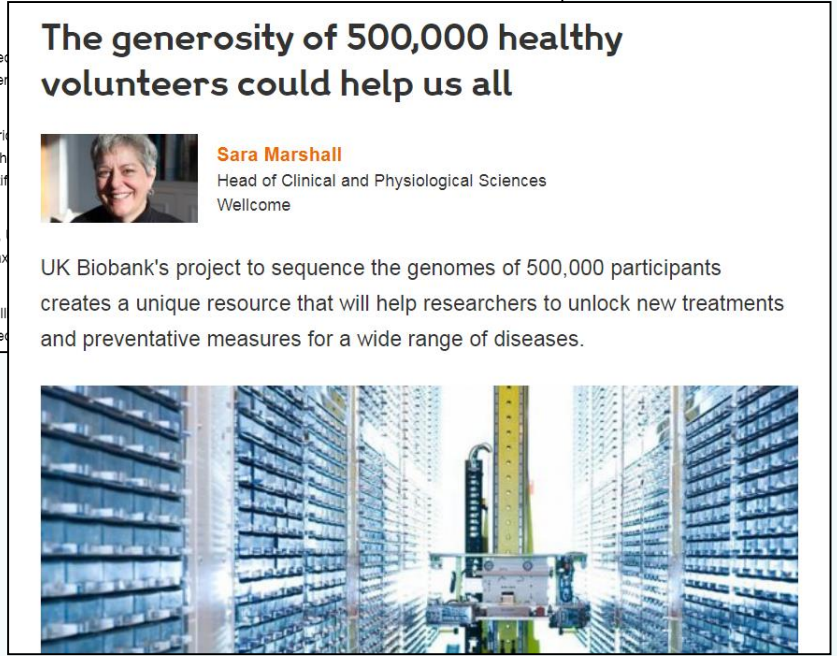
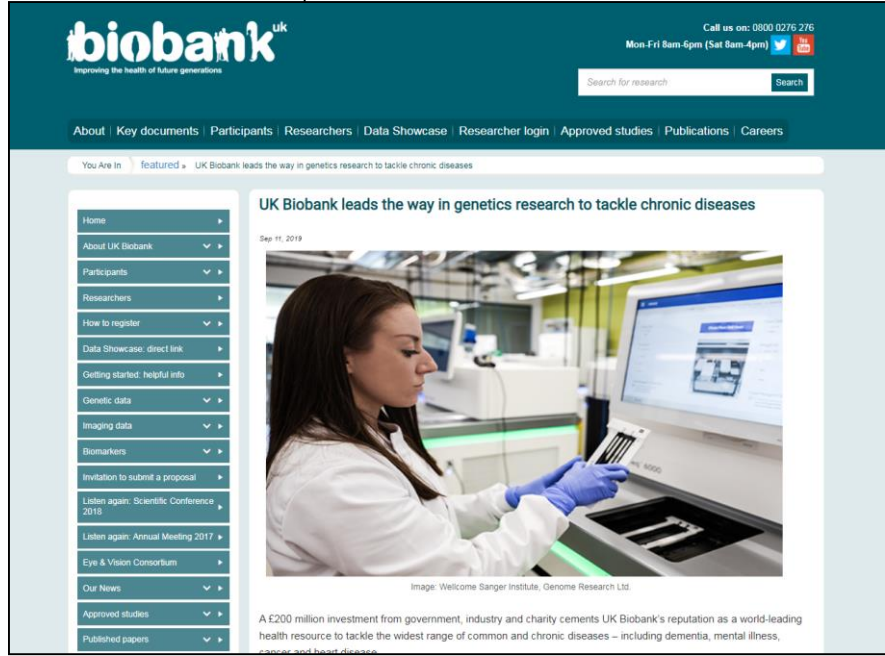
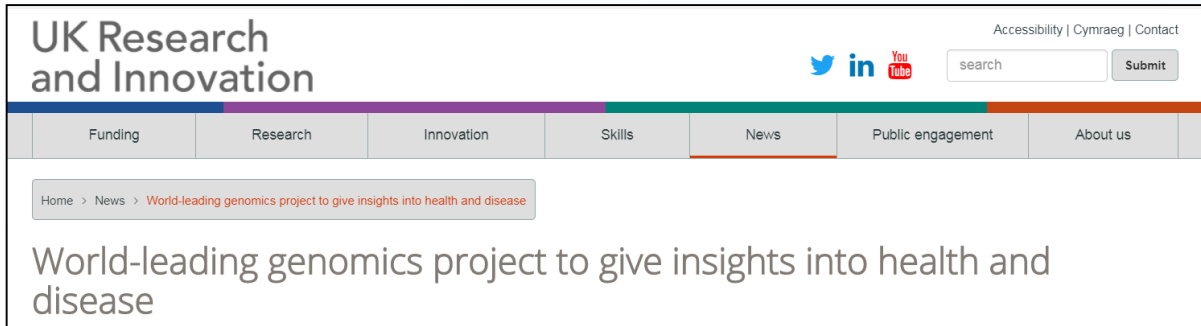


Exome Sequencing

- Regeneron/GSK sequencing 50,000 ppts
- Broader consortium sequencing remaining 450,000
- First 50,000 WES released in earlier this

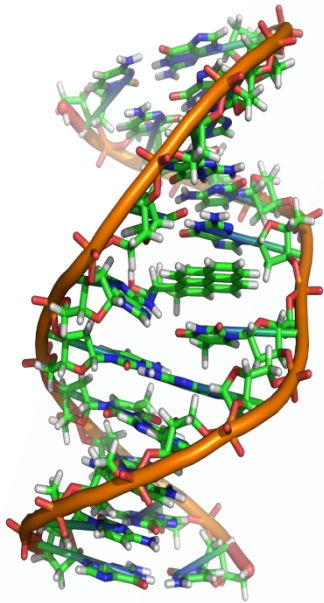
Whole Genome Sequencing of the UK Biobank cohort

31st August 2019: a consortium of 4 industry parties (Amgen, AstraZeneca, GSK and J&J) with additional support from UK government and Wellcome Trust committed the necessary funds to complete whole genome sequencing of the entire UK Biobank cohort



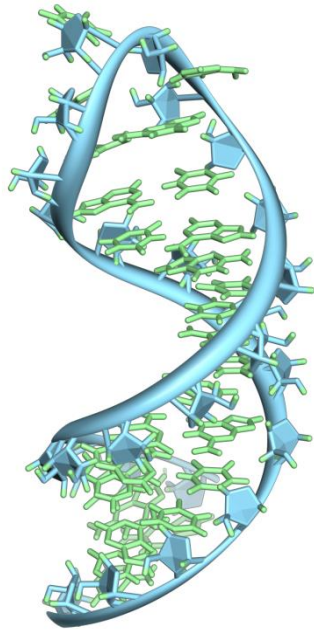
“...the most ambitious sequencing effort of whole human genomes ever undertaken”

Futures: Higher Content Analyses



DNA

Genomics



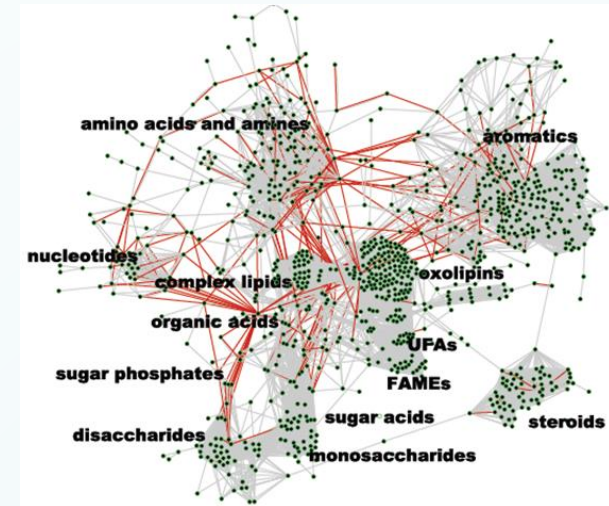
RNA

Transcriptomics



Proteins

Proteomics



Metabolites

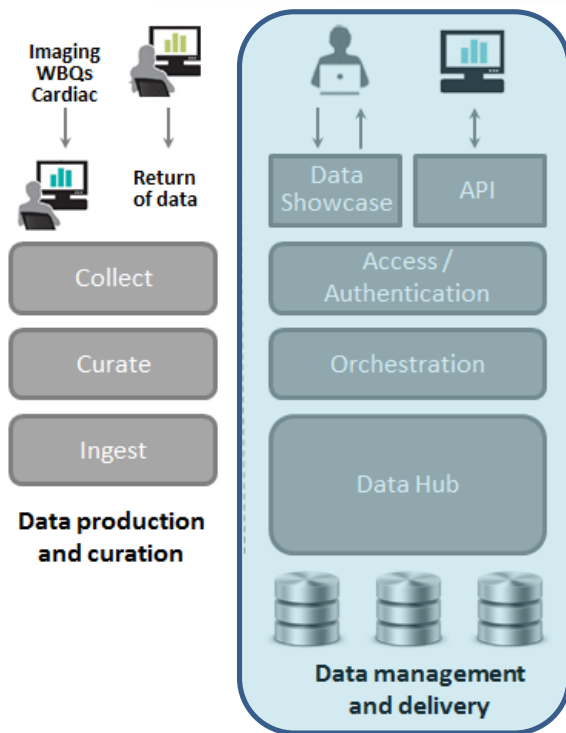
Metabolomics

- If UK Biobank performed Whole Genome Sequencing on the whole cohort....
 - 1 Sequence (CRAM File at 30x read) = 50 GB
 - 500,000 participants = circa 30 PBs of data

Improving the health of future generations

Platform Capability

All UKB data must currently be downloaded by the researcher before being used. The data generated by in flight or proposed projects will dramatically increase the volume of data UKB holds (e.g. 30PB for WGS), rendering such a download-only approach increasingly unsustainable as exome and genomic data are returned to the resource.



There is an emerging ecosystem of platform providers for genomics, and we are already gaining some experience of platforms such as DNAnexus (via the Regeneron project) and Seven Bridges (via the WGS Vanguard interim informatics).

There are a number of key areas any such platform must address, including: data security and access controls; support for multiple pipelines and tools, with appropriate versioning; the use of standard data formats and APIs.