

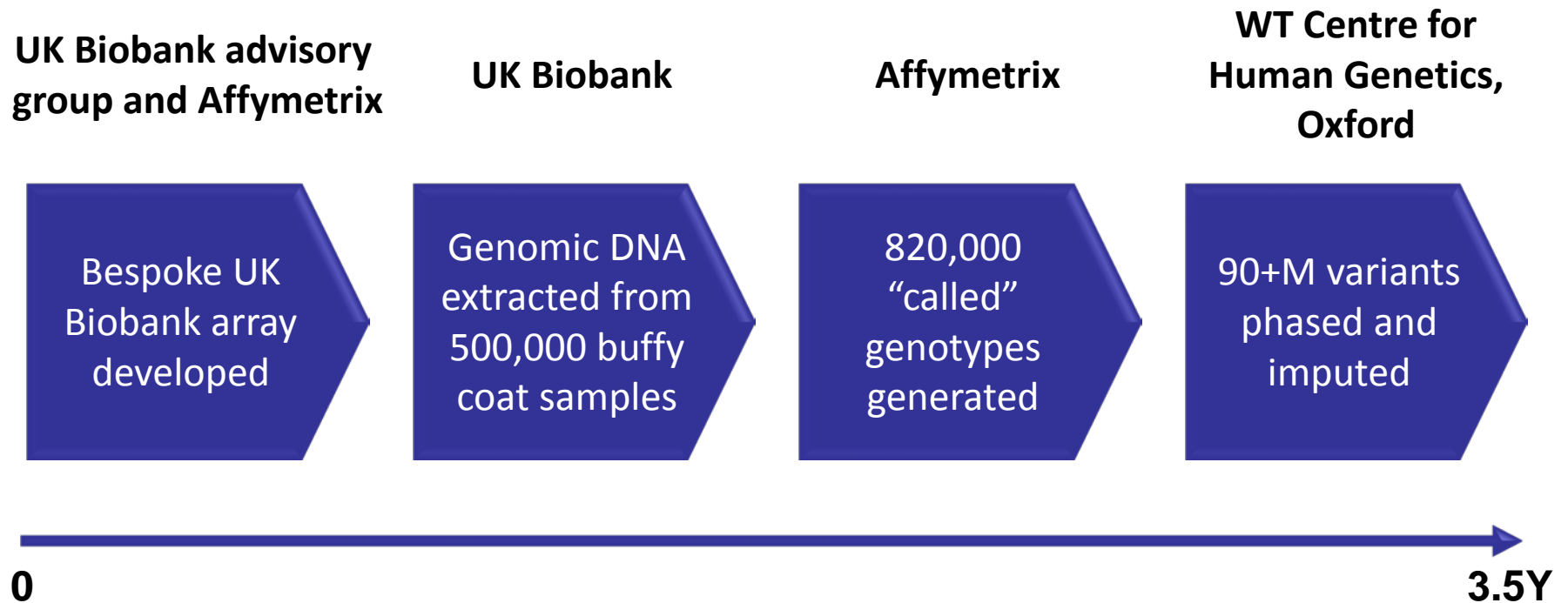
UK Biobank Prospective Cohort

- 500,000 UK men and women aged 40-69 years when recruited and assessed during 2006-2010
- Extensive baseline questions and measurements, with biological samples stored for future assays (but no feedback of their results to participants)
- General consent for all types of health research by both academic and commercial researchers, and for follow-up through all health-related records
- No preferential or exclusive access to the resource (except short period for researcher-derived data)

Main achievements in previous funding period (5+2 years: 2010-2015 plus 2016-2017)

- **Data access:** Resource opened for researchers in Q1 2012, and is now being used by an increasing number and range of researchers, and is starting to yield a rapidly increasing number of publications
- **Genotyping:** DNA extracted and genotyped for 500,000 participants in Q2 2013 to Q3 2015; genotype/imputed data released for 150,000 participants in Q3 2015, and data for full cohort now being released

Generation of UK Biobank genotype data



Release of UK Biobank genotype data

- Interim data release for 150,000 participants: May 2015
- Full data release in progress (with no preferential access)

European Bioinformatics Institute – European Genome Archive



Aspera system – ensures bandwidth and mitigates geographic impact on download speeds



- Total transfer of ~2PB of encrypted genotype data from EBI to 300 research groups with approved research projects
- Simultaneous provision of de-encryption keys and bridging files to all of these groups; current estimate: mid-July 2017

Main achievements in previous funding period (5+2 years: 2010-2015 plus 2016-2017)

- **Data access:** Resource opened for researchers in Q1 2012, and is now being used by an increasing number and range of researchers, and is starting to yield a rapidly increasing number of publications
- **Genotyping:** DNA extracted and genotyped for 500,000 participants in Q2 2013 to Q3 2015; genotype/imputed data released for 150,000 participants in Q3 2015, and data for full cohort now being released
- **Biomarker assays:** Systems established for high quality assays of about 40 biomarkers; completion and release of urine assay data in Q4 2016, and of HbA1c and serum assays in Q4 2017/Q1 2018

First phase of biomarkers being measured (in blood unless indicated otherwise)

Cardio-metabolic	Bones and joints	Cancer	Renal	Liver
Cholesterol	Vitamin D	SHBG	Cystatin C	Albumin
Direct LDL-c	Rh. factor	Testosterone	Total Protein	Direct Bilirubin
HDL-c	ALP	Oestradiol	Urea	Total Bilirubin
Triglyceride	Calcium	IGF-1	Phosphate	GGT
ApoA			Urate	ALT
ApoB			Urinary:	AST
CRP			Albumin	
Lp(a)			Creatinine	
HbA1c			Potassium	
Glucose			Sodium	

Note: Haematological assays were conducted during the recruitment phase, and these data have already been made available to researchers

Status of biomarker assays, quality control analyses and data release timelines

Markers	Assays started	Assays/QC completed	Data release
Urinary	Aug 2014	Apr 2016	Nov 2016
HbA1c	Nov 2014	Jun 2017	Q4 2017
Serum	Nov 2015	Oct 2017	Q1 2018

- Samples retrieved from archives in quasi-random order to avoid systematic differences in true values
- Standard laboratory QC/QA methods plus cohort-wide statistical analyses for shifts in assay results

Advantages of “whole cohort” sample assays: reduce depletion and increase accessibility

- Uncontrolled assays may deplete the available sample rapidly, preventing subsequent studies
- Assays conducted in separate subsets of a cohort (e.g. “nested” case-control strategies) may yield assay data that are not comparable
- By contrast, assays conducted in the whole cohort support many different comparisons, as well as minimising depletion, improving quality control and being cost-effective

But, such “whole cohort” strategies are costly and typically restricted to “standard” assays (although commercial funders can help)

Main achievements in previous funding period (5+2 years: 2010-2015 plus 2016-2017)

- **Data access:** Resource opened for researchers in Q1 2012, and is now being used by an increasing number and range of researchers, and is starting to yield a rapidly increasing number of publications
- **Genotyping:** DNA extracted and genotyped for 500,000 participants in Q2 2013 to Q3 2015; genotype/imputed data released for 150,000 participants in Q3 2015, and data for full cohort now being released
- **Biomarker assays:** Systems established for high quality assays of about 40 biomarkers; completion and release of urine assay data in Q4 2016, and of HbA1c and serum assays in Q4 2017/Q1 2018
- **Imaging study:** Development and piloting of imaging protocol, and initiation of the main phase of imaging for 100,000 participants (as well as assessing the approach to dealing with incidental findings)

Main achievements in previous funding period (5+2 years: 2010-2015 plus 2016-2017)

- **Data access:** Resource opened for researchers in Q1 2012, and is now being used by an increasing number and range of researchers, and is starting to yield a rapidly increasing number of publications
- **Genotyping:** DNA extracted and genotyped for 500,000 participants in Q2 2013 to Q3 2015; genotype/imputed data released for 150,000 participants in Q3 2015, and data for full cohort now being released
- **Biomarker assays:** Systems established for high quality assays of about 40 biomarkers; completion and release of urine assay data in Q4 2016, and of HbA1c and serum assays in Q4 2017/Q1 2018
- **Imaging study:** Development and piloting of imaging protocol, and initiation of the main phase of imaging for 100,000 participants (as well as assessing the approach to dealing with incidental findings)
- **Health outcome data:** Death, cancer and hospitalisation data linkage for all participants (with systems now in place for primary care data), and now starting to implement the outcome adjudication process

National cohort-wide linked healthcare datasets

Data type	Diagnostic coding systems	Country	Data providers	Data available for research with complete follow-up
Certified deaths by cause	ICD-10	England	NHS Digital (ONS)	2006 to Jan 2016
		Wales		
		Scotland	ISD, Scotland	2006 to Nov 2015
Registered cancers	ICD-9 & 10	England	NHS Digital (ONS)	1971 to Nov 2014
		Wales		
		Scotland	ISD, Scotland	1957 to Dec 2014
Hospital admissions	ICD-9 & 10 and OPCS4	England	NHS Digital (HES)	1996 to March 2015
		Wales	SAIL (PEDW)	1998 to Feb 2016
		Scotland	ISD, NSS (SMR)	1981 to Oct 2015

Focus of next 5 years of UK Biobank

- Further development of linkage to health-related datasets (i.e. for primary care as well as for death, cancer and hospitalisation)

Focus of next 5 years of UK Biobank

- Further development of linkage to health-related datasets (i.e. for primary care as well as for death, cancer and hospitalisation)
- Further development of outcome adjudication processes (i.e. scalable approaches for large number and range of outcomes)

Focus of next 5 years of UK Biobank

- Further development of linkage to health-related datasets (i.e. for primary care as well as for death, cancer and hospitalisation)
- Further development of outcome adjudication processes (i.e. scalable approaches for large number and range of outcomes)
- Further development of data storage and access systems, and maintenance/development of the sample archives

Focus of next 5 years of UK Biobank

- Further development of linkage to health-related datasets (i.e. for primary care as well as for death, cancer and hospitalisation)
- Further development of outcome adjudication processes (i.e. scalable approaches for large number and range of outcomes)
- Further development of data storage and access systems, and maintenance/development of the sample archives
- **Complete ongoing funded enhancements:**
 - genotyping (and telomere assays)
 - biochemical assays
 - imaging assessments

Focus of next 5 years of UK Biobank

- Further development of linkage to health-related datasets (i.e. for primary care as well as for death, cancer and hospitalisation)
- Further development of outcome adjudication processes (i.e. scalable approaches for large number and range of outcomes)
- Further development of data storage and access systems, and maintenance/development of the sample archives
- Complete ongoing funded enhancements:
 - genotyping (and telomere assays)
 - biochemical assays
 - imaging assessments
- Additional enhancements under consideration:
 - cardiac rhythm monitoring (£2.5M; funding review)
 - stool collection for microbiome (~£2M; subject to piloting)
 - biomarkers of infectious disease (~£10M; subject to piloting)