Data Analyst
UK Biobank
Candidate Brief
November 2022

Introduction from UK Biobank's Principal Investigator/Chief Executive Professor Sir Rory Collins

UK Biobank is scientifically unparalleled and arguably the UK's most significant scientific asset. With its unique combination of scale, depth, maturity and accessibility, more than 1,700 peer reviewed publications based on UK Biobank in 2021 alone, and an exponential increase in citations and patents, UK Biobank has become the world's most important health resource.

Its great success is testament to a visionary collaboration between successive UK Governments and scientific research funders – most notably the MRC and Wellcome – who understood the significance of advances in genomics and that information from such a database could ultimately lead to improved diagnosis, treatment and prevention strategies for the most devastating diseases, benefiting millions of people in the UK and around the world.

Since inception, UK Biobank has also leveraged £227m of industry funding with increasing levels of interest in, and use of, UK Biobank by industry. UK Biobank is a major contributor to the advancement of modern medicine and treatment and has enabled many scientific discoveries that improve human health.

UK Biobank has a unique role to play in the Government's UK Life Sciences Vision, in both its Preconditions for Success and its Healthcare Missions, set out in June 2021. UK Biobank's role can be best understood both in terms of its ability to attract vital industry investment in UK Life Sciences and in enabling major scientific discoveries that improve public health.

The key focus of UK Biobank over the last five years has been on generating increasingly detailed genetic information on all 500,000 participants and making those data available to the international research community. The unprecedented scale, detail and quality of these data has enabled researchers to make unique discoveries about the relevance of genetics for disease.

For the next five years, our focus will shift to measuring the products of the genes. In particular, the levels of different proteins and metabolites in the blood, urine and saliva collected from participants more than ten years earlier. Completion of the initial whole body imaging of 100,000 participants, followed by repeat imaging of 60,000 of them, will add to the richness of the data available for studying biological changes that lead to the development of disease, or provide protection against it.

None of this would be possible without the altruism of the 500,000 British citizens who agreed to join the study some decades ago and the many tens of thousands of scientists around the world ensuring that their contributions deliver on the long-term vision of the funders to make improvements in human health and wellbeing that are of global significance.

Company Overview

UK Biobank is a uniquely detailed large-scale prospective study, containing genomic, lifestyle, imaging and health information from 500,000 UK participants. The resource is globally accessible to approved researchers to enable novel and important scientific discoveries that improve human health.

UK Biobank's database is unique, combining several powerful differentiators, including:

- scale (500,000 people);
- depth (detailed lifestyle information, genetic and other assay data, imaging data);
- maturity (large numbers of different health outcomes available for study, 15 years ahead of any comparable cohort study); and
- accessibility (with over 28,000 researchers worldwide already using the data).

Increasing awareness of UK Biobank is more important than ever as the database grows in value.
With over 75% of all new researcher registrations coming from outside the UK, it truly is a global resource accessible to approved researchers undertaking health-related research.

UK Biobank is a distributed organisation, currently headquartered in Stockport, Greater Manchester, with four imaging centres delivering the world's largest multi-modal imaging study located in Stockport, Reading, Bristol and Newcastle. UK Biobank's epidemiological, health data teams and many of our systems are hosted at the Oxford Population Health (the Nuffield Department of Population Health, University of Oxford).

UK Research and Innovation have announced that subject to final approval, UK Biobank will receive an additional £128m of funding to support the next phase of our development. This transformative award will replace and enhance UK Biobank's infrastructure, creating a state-of-the-art facility in Manchester that will become the UK Biobank's headquarters from 2026.

UK Biobank's impact on research

With the addition of genetic sequencing data on a scale never seen before in health research, UK Biobank is one of the most significant resources for identifying target genes for drug discovery.

Large scale genetic sequencing on the whole exome and whole genome of 500,000 participants is accelerating scientific discovery for human health. For example, scientists have identified rare protein-coding variants that have a large impact on complex traits, such as obesity. Individuals with a variant that suppresses protein production by the GPR75 gene expressed in the brain were observed to have significantly lower rates of BMI. The finding may lead to development of a drug that can inhibit GPR75 to offer a therapeutic strategy for treating obesity.

UK Biobank is also becoming increasingly important for the NHS and the public purse. Implementation of outputs derived from UK Biobank will result in significant savings of public expenditure, by improving prevention, early diagnosis and treatment, thereby improving population health and lowering NHS costs. For instance, the UK Biobank-derived polygenic risk score concept, allows a low-cost genotyping test to identify a substantial proportion of the population at high risk of one or more common conditions, such as heart disease, who could benefit from more effectively targeted prevention and screening strategies.

The unprecedented scale, detail and quality of UK Biobank's imaging project has enabled researchers to make unique discoveries about the development of many diseases in older age. These images are allowing scientists to study the impact of dementia risk factors on the brain and define early brain "signatures" of dementia that are providing clues to how cognitive decline and dementia start. When combined with the wealth of health data in UK Biobank, such as cognitive function tests, and genetics, this is allowing researchers the opportunity to spot dementia before symptoms appear.

A high degree of participant engagement within the cohort has enabled UK Biobank to play a vital role in helping to manage the global coronavirus pandemic. Surveillance studies during the COVID19 pandemic have increased researchers' understanding of the longer-term health impacts of coronavirus infection with over 237 COVID- related publications produced based on the UK Biobank resource.

Given the current pace of UK Biobank's development, coupled with the large increase in the number of researchers accessing the resource, there is likely to be an exponential growth in the impact of the discoveries that emerge that will be of global significance.

Overview of the Role

- Location: Cheadle
- Salary: £33,000-£39,000 per annum depending upon experience
- Working hours: 35 hours per week
- Contract type: Permanent

UK Biobank are delighted to be hiring for Data Analysts to join our team in Cheadle owing to the increasing volume and complexity of data contained within the UK Biobank study, the ongoing work to generate whole-genome and whole-exome sequencing for our large cohort, and recent development of our own Research Access Platform.

The Data Analyst will process, manage and curate the large and diverse amounts of data acquired by UK Biobank and to maintain, document, interrogate and resolve issues and challenges associated with the data sets and the associated analytic approaches and algorithm.

A Basic DBS check is required for this role.

Principal Duties & Responsibilities

1. Working with UK Biobank IT and Epidemiology teams to clean, validate and refine large amounts of data from a variety of sources for inclusion in the UK Biobank resource to ensure data are error-free, consistent and well-documented.
2. Writing queries to interrogate the study resource and performing descriptive analyses for internal use.
3. Checking, extracting and manipulating the data for external researchers.
4. Documenting and maintaining records of all changes to the UK Biobank resource and associated processes in a well-structured manner.
5. Responding to data queries from researchers wishing to use the resource to address a wide range of research questions.
6. Producing technical reports on data management and procedures.
7. Providing clear annotation of data-fields and supporting documentation for the data in the UK Biobank resource.
8. Managing the transfer of data to and from researchers, where appropriate.
9. Exploring IT and data management tools for handling large and complex datasets.
10. Provision of statistical analysis support to the UK Biobank Epidemiology team, where appropriate.
11. Delivering training sessions and providing guidance to external researchers on the origins and interpretation of UK Biobank data.
12. Carry out other appropriate tasks as required.

Person Specification

Essential

1. An MSc or equivalent in a scientific or numerate subject, or ability to demonstrate equivalent experience.
2. Proficiency in using relational databases and writing SQL queries.
3. Proficiency in using R statistical software package, and associated libraries such as tidyverse.
4. Excellent written and oral communication skills in English, with the ability to present to technical and non-technical audiences.
5. Ability to prioritise workload and to work under pressure.
6. Excellent problem solving and analytical skills, and ability to prepare and maintain detailed documentation.
7. Highly motivated with excellent attention to detail, able to work on own initiative and as part of a multi-disciplinary team.

Desirable

1. Understanding of the application of statistical methods to epidemiological studies.
2. Understanding and/or experience of the use of health records for medical research.
3. Experience in managing large and complex information resources – extraction, conversion, integration and data transfer.
4. Experience of programming and/or scripting in a Linux environment.
5. Experience of bioinformatics and genetics-related toolsets.
6. Experience of working in a multidisciplinary team (e.g. a team involving clinicians, scientists, etc.) and/or in wider collaborative groups.

Salary and Employee Benefits

- Salary: £33,000-£39,000
- 26 days annual leave (which increases with length of service) + bank holidays (including Christmas closure)
- Generous 'hybrid' defined benefit/defined contribution Pension Scheme
  (9.8 % Employee Contribution, 21.6% Employer Contribution)
- Life Assurance Cover
- Enhanced maternity/paternity/adoption/shared parental leave
- Enhanced compassionate leave
- Enhanced Reserve Forces leave
- Enhanced company sick pay scheme
- Cycle to work scheme
- Health and wellbeing initiatives, including Employee Assistance Programme (EAP)
- Eye-care scheme
- Annual flu vaccination
- Employee discount platform scheme
- Active social committee
- Flexible work life balance policy
- Free onsite car parking
- Onsite café
- Free fruit and refreshments

How to Apply

Applications are welcome by sending CV to jobs@ukbiobank.ac.uk (detailing how you meet the role specifications) by 5pm on 31 December (subject to early closure upon receipt of suitable applications).

If you require assistance in applying for the role, or attending interviews, please contact our HR department via hr@ukbiobank.ac.uk so that we can make suitable arrangements.

Please click on these links for further information on Working for UK Biobank and Recruitment guidance for candidates

Our passion for diversity and equality means creating a work environment for all employees that is welcoming, respectful, engaging, and enriched with opportunities for personal and professional development.

For detailed information on how we process your personal data, please review our privacy policy.

In line with GDPR, we ask that you do NOT send us any information that can identify children or any of your Sensitive Personal Data (racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, data concerning health or sex life and sexual orientation, genetic and/or biometric data) in your CV and application documentation. Following this notice, any inclusion of your Sensitive Personal Data in your CV/application documentation will be understood by us as your express consent to process this information going forward. Please also remember to not mention anyone's information or details (e.g. referees) who have not previously agreed to their inclusion.